Report by the Commission on the Measurement of Economic Performance and Social Progress

Professor Joseph E. STIGLITZ, Chair, Columbia University

Professor Amartya SEN, Chair Adviser, Harvard University

Professor Jean-Paul FITOUSSI, Coordinator of the Commission, IEP

www.stiglitz-sen-fitoussi.fr

Other Members

Bina AGARWAL	*University of Delhi*
Kenneth J. ARROW	*StanfordUniversity*
Anthony B. ATKINSON	*Warden of Nuffield College*
François BOURGUIGNON	*School of Economics,*
Jean-Philippe COTIS	*Insee,*
Angus S. DEATON	*Princeton University*
Kemal DERVIS	*UNPD*
Marc FLEURBAEY	*Université Paris 5*
Nancy FOLBRE	*University of Massachussets*
Jean GADREY	*Université Lille*
Enrico GIOVANNINI	*OECD*
Roger GUESNERIE	*Collège de France*
James J. HECKMAN	*Chicago University*
Geoffrey HEAL	*Columbia University*
Claude HENRY	*Sciences-Po/Columbia University*
Daniel KAHNEMAN	*Princeton University*
Alan B. KRUEGER	*Princeton University*
Andrew J. OSWALD	*University of Warwick*
Robert D. PUTNAM	*Harvard University*
Nick STERN	*London School of Economics*
Cass SUNSTEIN	*University of Chicago*
Philippe WEIL	*Sciences Po*

Rapporteurs

Jean-Etienne CHAPRON *General Rapporteur*	*INSEE*
Didier BLANCHET	*INSEE*
Jacques LE CACHEUX	*OFCE*
Marco MIRA D'ERCOLE	*OCDE*
Pierre-Alain PIONNIER	*INSEE*
Laurence RIOUX	*INSEE/CREST*
Paul SCHREYER	*OCDE*
Xavier TIMBEAU	*OFCE*
Vincent MARCUS	*INSEE*

Table of contents

EXECUTIVE SUMMARY

I. SHORT NARRATIVE ON THE CONTENT OF THE REPORT

Chapter 1: Classical GDP Issues 21

Chapter 2: Quality of Life 41

Chapter 3: Sustainable Development and Environment 61

II. SUBSTANTIAL ARGUMENTS PRESENTED IN THE REPORT

Chapter 1: Classical GDP Issues 85

Chapter 2: Quality of Life 143

Chapter 3: Sustainable Development and Environment 233

EXECUTIVE SUMMARY

Why has this report been written?

1) In February 2008, the President of the French Republic, Nicholas Sarkozy, unsatisfied with the present state of statistical information about the economy and the society, asked, Joseph Stiglitz (President of the Commission), Amartya Sen (Advisor) and Jean Paul Fitoussi (Coordinator) to create a Commission, subsequently called "The Commission on the Measurement of Economic Performance and Social Progress" (CMEPSP). The Commission's aim has been to identify the limits of GDP as an indicator of economic performance and social progress, including the problems with its measurement; to consider what additional information might be required for the production of more relevant indicators of social progress; to assess the feasibility of alternative measurement tools, and to discuss how to present the statistical information in an appropriate way.

2) In effect, statistical indicators are important for designing and assessing policies aiming at advancing the progress of society, as well as for assessing and influencing the functioning of economic markets. Their role has increased significantly over the last two decades. This reflects improvements in the level of education in the population, increases in the complexity of modern economies and the widespread use of information technology. In the "information society", access to data, including statistical data, is much easier. More and more people look at statistics to be better informed or to make decisions. To respond to the growing demand for information, the supply of statistics has also increased considerably, covering new domains and phenomena.

3) What we measure affects what we do; and if our measurements are flawed, decisions may be distorted. Choices between promoting GDP and protecting the environment may be false choices, once environmental degradation is appropriately included in our measurement of economic performance. So too, we often draw inferences about what are good policies by looking at what policies have promoted economic growth; but if our metrics of performance are flawed, so too may be the inferences that we draw.

4) However, there often seems to be a marked distance between standard measures of important socio economic variables like economic growth, inflation, unemployment, etc. and widespread perceptions. The standard measures may suggest, for instance that there is less inflation or more growth than individuals perceive to be the case, and the gap is so large and so universal that it cannot be explained by reference to money illusion or to human psychology. In some countries, this gap has undermined confidence in official statistics (for example, in France and in the United Kingdom. only one third of citizens trust official figures, and these countries are not exceptions), with a clear impact on the way in which public discourse about the conditions of the economy and necessary policies takes place.

5) There may be several explanations for the gap between the statistical measurement of socio-economic phenomena and citizen perception of the same phenomena:

– The statistical *concepts* may be correct, but the measurement process may be imperfect.

– In many cases, there are debates about what are the right concepts, and the appropriate use of different concepts.

– When there are large changes in inequality (more generally a change in income distribution) gross domestic product (GDP) or any other aggregate computed *per capita* may not provide an accurate assessment of the situation in which most people find themselves. If inequality increases enough relative to the increase in average per capital GDP, most people can be worse off even though average income is increasing

– The commonly used statistics may not be capturing some phenomena, which have an increasing impact on the well-being of citizens. For example, traffic jams may increase GDP as a result of the increased use of gasoline, but obviously not the quality of life. Moreover, if citizens are concerned about the quality of air, and air pollution is increasing, then statistical measures which ignore air pollution will provide an inaccurate estimate of what is happening to citizens' well-being. Or a tendency to measure gradual change may be inadequate to capture risks of abrupt alterations in the environment such as climate change.

– The way in which statistical figures are reported or used may provide a distorted view of the trends of economic phenomena. For example, much emphasis is usually put on GDP although net national product (which takes into account the effect of depreciation), or real household income (which focuses on the real income of households within the economy) may be more relevant. These numbers may differ markedly. Then, GDP is not wrong *as such*, but wrongly used. What is needed is a better understanding of the appropriate use of each measure.

6) Indeed, for a long time there have been concerns about the adequacy of current measures of economic performance, in particular those solely based on GDP. Besides, there are even broader concerns about the relevance of these figures as measures of societal well-being. To focus specifically on the enhancement of inanimate objects of convenience (for example in the GNP or GDP which have been the focus of a myriad of economic studies of progress), could be ultimately justified – to the extent it could be – only through what these objects do to the human lives they can directly or indirectly influence. Moreover, it has long been clear that GDP is an inadequate metric to gauge well-being over time particularly in its economic, environmental, and social dimensions, some aspects of which are often referred to as *sustainability.*

Why is this report important?

7) Between the time that the Commission began working on this report and the completion of this Report, the economic context has radically changed. We are now living one of the worst financial, economic and social crises in post-war history. The reforms in measurement recommended by the Commission would be highly desirable, even if we had not had the crisis. But some members of the Commission believe that the crisis provides heightened urgency to these reforms. They believe that one of the reasons why the crisis took many by surprise is that our measurement system failed us and/or market participants and government officials were not focusing on the right set of statistical

indicators. In their view, neither the private nor the public accounting systems were able to deliver an early warning, and did not alert us that the *seemingly* bright growth performance of the world economy between 2004 and 2007 may have been achieved at the expense of future growth. It is also clear that some of the performance was a "mirage", profits that were based on prices that had been inflated by a bubble. It is perhaps going too far to hope that had we had a better measurement system, one that would have signalled problems ahead, so governments might have taken early measures to avoid or at least to mitigate the present turmoil. But perhaps had there been more awareness of the limitations of standard metrics, like GDP, there would have been less euphoria over economic performance in the years prior to the crisis; metrics which incorporated assessments of sustainability (e.g. increasing indebtedness) would have provided a more cautious view of economic performance. But many countries lack a timely and complete set of wealth accounts – the 'balance sheets' of the economy – that could give a comprehensive picture of assets, debts and liabilities of the main actors in the economy.

8) We are also facing a looming environmental crisis, especially associated with global warming. Market prices are distorted by the fact that there is no charge imposed on carbon emissions; and no account is made of the cost of these emissions in standard national income accounts. Clearly, measures of economic performance that reflected these environmental costs might look markedly different from standard measures.

9) If the view expressed in the preceding paragraphs is not necessarily shared by all members of the Commission, the whole Commission is convinced that the crisis is teaching us a very important lesson: those attempting to guide the economy and our societies are like pilots trying to steering a course without a reliable compass. The decisions they (and we as individual citizens) make depend on what we measure, how good our measurements are and how well our measures are understood. We are almost blind when the metrics on which action is based are ill-designed or when they are not well understood. For many purposes, we need better metrics. Fortunately, research in recent years has enabled us to improve our metrics, and it is time to incorporate in our measurement systems some of these advances. There is also consensus among the Commission members that better measures may enable us to steer our economies better through and out of crises. Many of the indicators put forward by the report will lend themselves to this purpose.

10) The report is about measurement rather than policies, thus it does not discuss how best our societies could advance through collective actions in the pursuit of various goals. However, as what we measure shapes what we collectively strive to pursue – and what we pursue determines what we measure - the report and its implementation may have a significant impact on the way in which our societies looks at themselves and, therefore, on the way in which policies are designed, implemented and assessed.

11) The Commission notes the important progress in statistical measurement that has occurred in recent years, and urges continued efforts to improve our statistical data base and the indicators that are constructed from this data base. The report indicates avenues for more or different measurement efforts in various domains, and we hope that it will influence future statistical policies in both developed and developing countries, as well as the work of international organisations that play a key role in the development of statistical standards worldwide.

By whom has the report been written?

12) This is a report written by economists and social scientists. The members of the Commission represent a broad range of specialisations, from national accounting to the economics of climate change. The members have conducted research on social capital, happiness, and health and mental well-being. They share the belief that it is important to build bridges between different communities – between the producers and users of statistical information, whatever their discipline – that have become increasingly distant in recent years. Commission members see their expertise as a complement to reports on similar topics that were written from a different perspective, for instance by scientists on climate change or by psychologists on mental health. Although the core of the report is rather technical, the summaries of the technical chapter have been written using, as much as possible, non-technical language.

To whom is the report addressed?

13) The Commission hopes that the Report will find a receptive audience among four distinct groups, and it has been written with that in mind. The Report is addressed, first of all, to political leaders. In this time of crises, when new political narratives are necessary to identify where our societies should go, the report advocates a shift of emphasis from a "production-oriented" measurement system to one focused on the well-being of current and future generations, i.e. toward broader measures of social progress.

14) Second, the report is aimed at reaching policy-makers who wish to get a better sense of which indicators are available and useful to design, implement and assess policies aimed at improving well-being and foster social progress. Policy-makers are reminded both of the richness and of the shortcomings of existing data but also of the fact that reliable quantitative information 'does not grow on trees' and significant investments need to be made to develop statistics and indicators that provide policymakers with the information they need to make the decisions confronting them.

15) Third, the report has been written for the academic community, statisticians, and intensive users of statistics. They are reminded of how difficult it can be to produce reliable data and of the numerous assumptions that underlay all statistical series. Academics will, hopefully, become more cautious in the confidence they place in certain statistics. Those in national statistical offices will, hopefully find helpful suggestions about areas where further developments might be particularly valuable.

16) Lastly, the report has been written for civil society organisations that are both users and producers of statistics. More generally, it is addressed to the public at large, whether from richer or poorer countries and whether rich or poor within societies. We hope that through a better understanding of the statistical data and indicators that are available (their strengths and limits), they can make a better assessment of the problems facing their societies. We hope the report will also serve journalists and the media who have a responsibility in enabling citizens to get a sense of what is happening in the society in which they are living. Information is a public good; the more we are informed about what is happening in our society, the better will our democracies be able to function.

What are the main messages and recommendations?

17) The report distinguishes between an assessment of *current well-being* and an assessment of *sustainability*, whether this can last over time. Current well-being has to do with both economic resources, such as income, and with non-economic aspects of peoples' life (what they do and what they can do, how they feel, and the natural environment they live in). Whether these levels of well-being can be sustained over time depends on whether stocks of capital that matter for our lives (natural, physical, human, social) are passed on to future generations.

    To organise its work, the Commission organized itself into three working groups, focusing respectively on: Classical GDP issues, Quality of life and Sustainability. The following main messages and recommendations arise from the report

Towards better measures of economic performance in a complex economy

18) Before **g**oing beyond GDP and tackling the more difficult task of measuring well-being, it is worth asking where existing measures of economic performance need improving. Measuring production – a variable which among other things determines the level of employment – is essential for the monitoring of economic activity. The first main message of our report is that time has come to adapt our system of measurement of economic activity to better reflect the structural changes which have characterized the evolution of modern economies. In effect, the growing share of services and the production of increasingly complex products make the measurement of output and economic performance more difficult than in the past. There are now many products whose quality is complex, multi-dimensional and subject to rapid change. This is obvious for goods, like cars, computers, washing machines and the like, but is even truer for services, such as medical services, educational services, information and communication technologies, research activities and financial services. In some countries and some sectors, increasing "output" is more a matter of an increase in the quality of goods produced and consumed than in the quantity. Capturing quality change is a tremendous challenge, yet this is vital to measuring real income and real consumption, some of the key determinants of people's material well-being. Under-estimating quality improvements is equivalent to over-estimating the rate of inflation, and therefore to under-estimating real income. The opposite is true when quality improvements are over-stated.

19) Governments play an important part in today's economies. They provide services of a "collective" nature, such as security, and of a more "individual" nature, such as medical services and education. The mix between private and public provision of individual services varies significantly across countries and over time. Beyond the contribution of collective services to citizens' living standards, individual services, particularly education, medical services, public housing or public sports facilities, are almost certainly valued positively by citizens. These services tend to be large in scale, and have increased considerably since World War II, but, in many cases, they remain badly measured. Traditionally, measures have been based on the inputs used to produce these services (such as the number of doctors) rather than on the actual outputs produced (such as the number of particular medical treatments). Making adjustments for quality changes is even more difficult. Because outputs are taken to move in tandem with inputs

productivity change in the provision of these services is ignored. It follows that if there is positive (negative) productivity change in the public sector, our measures under (over)-estimate economic growth and real income. For a satisfactory measure of economic performance and living standards it is thus important to come to grips with measuring government output. (In our present, admittedly flawed, system of measurement based on expenditures, government output represents around 20% of GDP in many OECD countries and total government expenditure more than 40% for the OECD countries.)

20) While there are methodological disagreements about how to make the adjustments to quality or how to go about measuring government output, there is a broad consensus that adjustments should be made, and even about the principles which should guide such adjustments. The disagreements arise in the practical implementation of these principles. The Commission has addressed both the principles and the difficulties in implementations, in its Report.

From production to well-being

21) Another key message, and unifying theme of the report, is that the time is ripe for our measurement system to *shift emphasis from measuring economic production to measuring people's well-being*. And measures of well-being should be put in a context of sustainability. Despite deficiencies in our measures of production, we know much more about them than about well-being. Changing emphasis does not mean dismissing GDP and production measures. They emerged from concerns about market production and employment; they continue to provide answers to many important questions such as monitoring economic activity. But emphasising well-being is important because there appears to be an increasing gap between the information contained in aggregate GDP data and what counts for common people's well-being. This means working towards the development of a statistical system that complements measures of market activity by measures centred on people's well-being and by measures that capture sustainability. Such a system must, of necessity, be plural – because no single measure can summarize something as complex as the well-being of the members of society, our system of measurement must encompass a range of different measures. The issue of aggregation across dimensions (that is to say, how we add up, for example, a measure of health with a measure of consumption of conventional goods), while important, is subordinate to the establishment of a broad statistical system that captures as many of the relevant dimensions as possible. Such a system should not just measure *average* levels of well-being within a given community, and how they change over time, but also document the diversity of peoples' experiences and the linkages across various dimensions of people's life. There are several dimensions to well-being but a good place to start is the measurement of material well-being or living standards.

Recommendation 1: When evaluating material well-being, look at income and consumption rather than production

22) 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. Earlier paragraphs have emphasized some of the important areas where more progress is needed in its computation. As statisticians and economists know very well, GDP mainly measures market production – expressed in money units – and as such it is

useful. However, it has often been treated as if it were a measure of economic well-being. Conflating the two can lead to misleading indications about how well-off people are and entail the wrong policy decisions. Material living standards are more closely associated with measures of net national income, real household income and consumption – production can expand while income decreases or vice versa when account is taken of depreciation, income flows into and out of a country, and differences between the prices of output and the prices of consumer products.

Recommendation 2: Emphasise the household perspective

23) While it is informative to track the performance of economies as a whole, trends in citizens' material living standards are better followed through measures of household income and consumption. Indeed, the available national accounts data shows that in a number of OECD countries real household income has grown quite differently from real GDP per capita, and typically at a lower rate. The household perspective entails taking account of payments between sectors, such as taxes going to government, social benefits coming from government, and interest payments on household loans going to financial corporations. Properly defined, household income and consumption should also reflect in-kind services provided by government, such as subsidized health care and educational services. A major effort of statistical reconciliation will also be required to understand why certain measures such as household income can move differently depending on the underlying statistical source.

Recommendation 3: Consider income and consumption jointly with wealth

24) Income and consumption are crucial for assessing living standards, but in the end they can only be gauged in conjunction with information on wealth. A household that spends its wealth on consumption goods increases its current well-being but at the expense of its future well-being. The consequences of such behavior would be captured in a household's balance sheet, and the same holds for other sectors of the economy, and for the economy as a whole. To construct balance sheets, we need comprehensive accounts of assets and liabilities. Balance sheets for countries are not novel in concept, but their availability is still limited and their construction should be promoted. Measures of wealth are central to measuring sustainability. What is carried over into the future necessarily has to be expressed as stocks – of physical, natural, human and social capital. The right valuation of these stocks plays a crucial role, and is often problematic. There is also a need to "stress test" balance sheets with alternative valuations when market prices for assets are not available or are subject to bubbles and bursts. Some more direct non-monetary indicators may be preferable when the monetary valuation is very uncertain or difficult to derive.

Recommendation 4: Give more prominence to the distribution of income, consumption and wealth

25) Average income, consumption and wealth are meaningful statistics, but they do not tell the whole story about living standards. For example, a rise in average income could be unequally shared across groups, leaving some households relatively worse-off than others. Thus, average measures of income, consumption and wealth should be accompanied by indicators that reflect their distribution. Median consumption (income,

wealth) provides a better measure of what is happening to the "typical" individual or household than average consumption (income or wealth). But for many purposes, it is also important to know what is happening at the bottom of the income/wealth distribution (captured in poverty statistics), or at the top. Ideally, such information should not come in isolation but be linked, i.e. one would like information about how well-off households are with regard to different dimensions of material living standards: income, consumption and wealth. After all, a low-income household with above-average wealth is not necessarily worse-off than a medium-income household with no wealth. (The desirability of providing information on the "joint distribution" of the dimensions of people's well-being will be raised once again in the recommendations below on how to measure quality of life.)

Recommendation 5: Broaden income measures to non-market activities

26) There have been major changes in how households and society function. For example, many of the services people received from other family members in the past are now purchased on the market. This shift translates into a rise in income as measured in the national accounts and may give a false impression of a change in living standards, while it merely reflects a shift from non-market to market provision of services. Many services that households produce for themselves are not recognized in official income and production measures, yet they constitute an important aspect of economic activity. While their exclusion from official measures reflects uncertainty about data more than conceptual difficulties, there has been progress in this arena; still, more and more systematic work in this area should be undertaken. This should start with information on how people spend their time that is comparable both over the years and across countries. Comprehensive and periodic accounts of household activity as satellites to the core national accounts should complement the picture. In developing countries, the production of goods (for instance food or shelter) by households plays an important role. Tracking the production of such home-produced goods is important to assess consumption levels of households in these countries.

27) Once one starts focusing on non-market activities, the question of leisure arises. Consuming the same bundle of goods and services but working for 1500 hours a year instead of 2000 hours a year implies an increase in one's standard of living. Although valuation of leisure is fraught with difficulties, comparisons of living standards over time or across countries needs to take into account the amount of leisure that people enjoy.

Well-being is multi-dimensional

28) To define what well-being means a multidimensional definition has to be used. Based on academic research and a number of concrete initiatives developed around the world, the Commission has identified the following key dimension that should be taken into account. At least in principle, these dimensions should be considered simultaneously:

    i. Material living standards (income, consumption and wealth);
    ii. Health;
    iii. Education;
    iv. Personal activities including work
    v. Political voice and governance;

vi. Social connections and relationships;
vii. Environment (present and future conditions);
viii. Insecurity, of an economic as well as a physical nature.

All these dimensions shape people's well-being, and yet many of them are missed by conventional income measures.

Objective and subjective dimensions of well-being are both important

Recommendation 6: Quality of life depends on people's objective conditions and capabilities. Steps should be taken to improve measures of people's health, education, personal activities and environmental conditions. In particular, substantial effort should be devoted to developing and implementing robust, reliable measures of social connections, political voice, and insecurity that can be shown to predict life satisfaction.

29) The information relevant to valuing quality of life goes beyond people's self-reports and perceptions to include measures of their "functionings" and freedoms. In effect, what really matters are the capabilities of people, that is, the extent of their opportunity set and of their freedom to choose among this set, the life they value. The choice of relevant functionings and capabilities for any quality of life measure is a value judgment, rather than a technical exercise. But while the precise list of the features affecting quality of life inevitably rests on value judgments, there is a consensus that quality of life depends on people's health and education, their everyday activities (which include the right to a decent job and housing), their participation in the political process, the social and natural environment in which they live, and the factors shaping their personal and economic security. Measuring all these features requires both objective and subjective data. The challenge in all these fields is to improve upon what has already been achieved, to identify gaps in available information, and to invest in statistical capacity in areas (such as time-use) where available indicators remain deficient.

Recommendation 7: Quality-of-life indicators in all the dimensions covered should assess inequalities in a comprehensive way

30) Inequalities in human conditions are integral to any assessment of quality of life across countries and the way that it is developing over time. Most dimensions of quality-of-life require appropriate separate measures of inequality, but, as noted in par. 25, taking into account linkages and correlations. Inequalities in quality of life should be assessed across people, socio-economic groups, gender and generations, with special attention to inequalities that have arisen more recently, such as those linked to immigration.

Recommendation 8: Surveys should be designed to assess the links between various quality-of-life domains for each person, and this information should be used when designing policies in various fields

31) It is critical to address questions about how developments in one domain of quality of life affect other domains, and how developments in all the various fields are related to income. This is important because the consequences for quality of life of having multiple disadvantages far exceed the sum of their individual effects. Developing measures of

these cumulative effects requires information on the "joint distribution" of the most salient features of quality of life across everyone in a country through dedicated surveys. Steps in this direction could also be taken by including in all surveys some standard questions that allow classifying respondents based on a limited set of characteristics. When designing policies in specific fields, impacts on indicators pertaining to different quality-of-life dimensions should be considered jointly, to address the interactions between dimensions and the needs of people who are disadvantaged in several domains

Recommendation 9: Statistical offices should provide the information needed to aggregate across quality-of-life dimensions, allowing the construction of different indexes.

32) While assessing quality-of-life requires a plurality of indicators, there are strong demands to develop a single summary measure. Several summary measures of quality of life are possible, depending on the question addressed and the approach taken. Some of these measures are already being used, such as average levels of life-satisfaction for a country as a whole, or composite indices that aggregate averages across objective domains, such as the Human Development Index. Others could be implemented if national statistical systems made the necessary investment to provide the data required for their computation. These include measures of the proportion of one's time in which the strongest reported feeling is a negative one, measures based on counting the occurrence and severity of various objective features of people's lives, and (equivalent-income) measures based on people's states and preferences.

33) The Commission believes that in addition to objective indicators of well-being, subjective measures of the quality-of-life should be considered.

Recommendation 10: Measures of both objective and subjective well-being provide key information about people's quality of life. Statistical offices should incorporate questions to capture people's life evaluations, hedonic experiences and priorities in their own survey.

34) Research has shown that it is possible to collect meaningful and reliable data on subjective as well as objective well-being. Subjective well-being encompasses different aspects (cognitive evaluations of one's life, happiness, satisfaction, positive emotions such as joy and pride, and negative emotions such as pain and worry): each of them should be measured separately to derive a more comprehensive appreciation of people's lives. Quantitative measures of these subjective aspects hold the promise of delivering not just a good measure of quality of life per se, but also a better understanding of its determinants, reaching beyond people's income and material conditions. Despite the persistence of many unresolved issues, these subjective measures provide important information about quality of life. Because of this, the types of question that have proved their value within small-scale and unofficial surveys should be included in larger-scale surveys undertaken by official statistical offices.

Use a pragmatic approach towards measuring sustainability

35) Measuring and assessing sustainability has been a central concern of the Commission. Sustainability poses the challenge of determining if at least the current level of well-being can be maintained for future generations. By its very nature, sustainability involves the future and its assessment involves many assumptions and normative choices. This is

further complicated by the fact that at least some aspects of environmental sustainability (notably climate change) is affected by interactions between the socio-economic and environmental models followed by different countries. The issue is indeed complex, more complex than the already complicated issue of measuring current well-being or performance.

Recommendation 11: Sustainability assessment requires a well-identified dashboard of indicators. The distinctive feature of the components of this dashboard should be that they are interpretable as variations of some underlying "stocks". A monetary index of sustainability has its place in such a dashboard but, under the current state of the art, it should remain essentially focused on economic aspects of sustainability.

36) The assessment of sustainability is complementary to the question of current well-being or economic performance, and must be examined separately. This may sound trivial and yet it deserves emphasis, because some existing approaches fail to adopt this principle, leading to potentially confusing messages. For instance, confusion may arise when one tries to combine current well-being and sustainability into a single indicator. To take an analogy, when driving a car, a meter that added up in one single number the current speed of the vehicle and the remaining level of gasoline would not be of any help to the driver. Both pieces of information are critical and need to be displayed in distinct, clearly visible areas of the dashboard.

37) At a minimum, in order to measure sustainability, what we need are indicators that inform us about the change in the quantities of the different factors that matter for future well-being. Put differently, sustainability requires the simultaneous preservation or increase in several "stocks": quantities and qualities of natural resources, and of human, social and physical capital.

38) There are two versions to the stock approach to sustainability. One version just looks at variations in each stock separately, assessing whether the stock is increase or decreasing, with a view particularly to doing whatever is necessary to keep each above some critical threshold. The second version converts all these assets into a monetary equivalent, thereby implicitly assuming substitutability between different types of capital, so that a decrease in, say, natural capital might be offset by a sufficient increase in physical capital (appropriately weighted). Such an approach has significant potential, but also several limitations, the most important being the absence of many markets on which valuation of assets could be based. Even when there are market values, there is no guarantee that they adequately reflect how the different assets matter for future well-being. The monetary approach requires imputations and modelling which raise informational difficulties. All this suggests starting with a more modest approach, i.e. focusing the monetary aggregation on items for which reasonable valuation techniques exist, such as physical capital, human capital and certain natural resources. In so doing, it should be possible to assess the "economic" component of sustainability, that is, whether or not countries are over-consuming their economic wealth.

Physical indicators for environmental pressures

Recommendation 12: The environmental aspects of sustainability deserve a separate follow-up based on a well-chosen set of physical indicators. In particular there is a need for a clear

indicator of our proximity to dangerous levels of environmental damage (such as associated with climate change or the depletion of fishing stocks.)

39) For the reasons mentioned above, placing a monetary value on the natural environment is often difficult and separate sets of physical indicators will be needed to monitor the state of the environment. This is in particular the case when it comes to irreversible and/or discontinuous alterations to the environment. For that reason members of the Commission believe in particular that there is a need for a clear indicator of increases in atmospheric concentrations of greenhouse gases associated with proximity to dangerous levels of climate change (or levels of emissions that might reasonably be expected to lead to such concentrations in the future. Climate change (due to increases in atmospheric concentrations of greenhouse gases) is also special in that it constitutes a truly global issue that cannot be measured with regard to national boundaries. Physical indicators of this kind can only be identified with the help of the scientific community. Fortunately, a good deal of work has already been undertaken in this field.

What is next?

40) The Commission regards its report as opening a discussion rather than closing it. The report hints at issues that ought to be addressed in the context of more comprehensive research efforts. Other bodies, at the national and international level, should discuss the recommendations in this report, identify their limits, and see how best they can contribute to this broad agenda, each from its own perspective.

41) The Commission believes that a global debate around the issues and recommendations raised in this report provides an important venue for a discussion of societal values, for what we, as a society, care about, and whether we are really striving for what is important

42) At the national level, round-tables should be established, with the involvement of stakeholders, to identify and prioritise those indicators that carry to potential for a shared view of how social progress is happening and how it can be sustained over time.

43) The Commission hopes that this Report will provide the impetus not only for this broader discussion, but for on-going research into the development of better metrics that will enable us to assess better economic performance and social progress.

I. Short Narrative on the Content of the Report

CHAPTER 1 : CLASSICAL GDP ISSUES

CHAPTER 2 : QUALITY OF LIFE

CHAPTER 3 : SUSTAINABLE DEVELOPMENT AND ENVIRONMENT

CHAPTER 1 - CLASSICAL GDP ISSUES [1]

1 - Introduction

1. 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 mainly measures *market* production, though it has often been treated as if it were a measure of economic well-being. Conflating the two can lead to misleading indications about how well-off people are and entail the wrong policy decisions.

2. One reason why money measures of economic performance and living standards have come to play such an important role in our societies is that the 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 DVD players, we can add up their values 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 properly, the ratio of one market price to another is reflective of the relative appreciation of the two products by those who purchase them. Moreover, 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. Furthermore, 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 are evolving in real terms.

3. 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), raising the question of how these services should be valued. Second, even where there are market prices, they may deviate from society's underlying valuation. In particular, when the consumption or production of particular products affects society as a whole, the price that individuals pay for those products will differ from their 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.

4. There is yet another problem. While talking about the concepts of "prices" and "quantities" might be straightforward, defining and measuring how they change in practice is an altogether different matter. As it happens, many products change over time – they disappear entirely or new features are added to them. *Quality change* can be very rapid in areas like information and communication technologies. There are also products whose quality is complex, multi-dimensional and hard to measure, such as medical services, educational services, research activities and financial services. Difficulties also arise in

1. Evidence and references in support of the claims presented in this Summary are presented in a companion technical report.

collecting data in an era when an increasing fraction of sales take place over the internet and at sales as well as discount stores. As a consequence, capturing quality change correctly is a tremendous challenge for statisticians, yet this 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 to under-estimating real income. For instance, in the mid-1990s, a report reviewing the measurement of inflation in the United States (Boskin Commission Report) estimated that insufficient accounting for quality improvements in goods and services had led to an annual over-estimation of inflation by 0.6%. This led to a series of changes to the US consumer price index.

5. The debate in Europe has tended to go the opposite way: official price statistics have been criticized for *under-estimating* inflation. This has been partly because people's perception of inflation differs from the national averages presented in the consumer price index, and also because it is felt that statisticians over-adjust for quality improvements in products, thereby painting too rosy a picture of citizens' real income.

6. 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 that they dispose of the relevant 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. The complex and ever-changing bundles of services offered by telecommunications companies are another case in point where it is difficult to ensure the transparency and comparability of price signals.

7. All the above considerations imply that price signals have to be interpreted with care in temporal and spatial comparisons**.** For a number of purposes, they do not provide a useful vehicle for the aggregation of quantities. This does not imply that the use of market prices in constructing measures of economic performance is generally flawed. But it does suggest prudence, in particular with regard to the often over-emphasized measure, GDP.

8. This Chapter suggests five ways of dealing with some of the deficiencies of GDP as an indicator of living standards. First, emphasize well-established indicators other than GDP in the national accounts. 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 the distribution of income, consumption and wealth to data on the average evolution of these elements. Finally, widen the scope of what is being measured. In particular, a significant part of economic activity takes place 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

9. Imputations exist for two related reasons. The first is 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 if two persons receive the same money income but one of them lives in his/her own house while the other rents, they are not equally well-off – hence the imputation in order to better compare incomes over time or between countries. This brings us to the second reason for imputations, the *invariance principle:* the value of the 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.

10. The imputations can be smaller or larger, depending on the country and on the national accounts aggregate considered. In France and Finland, for example, the main imputations account for about one-third of adjusted household disposable income and for just over 20% in the United States. Thus, in the absence of imputations the living standards of French and Finnish households would be understated relative to the United States.

11. But imputations come at a price. One is data quality: imputed values tend to be less reliable than observed values. Another 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 changes in perceived income and changes in measured income. This problem is exacerbated 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. 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.

12. There is no easy way out of the tension between comprehensiveness and comprehensibility except to keep both elements of information available for users and to maintain 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 would represent a significant improvement.

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

3.1 - Emphasize national accounts aggregates other than GDP

13. 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, by accounting for depreciation so as to deal with *net rather than gross measures* of economic activity.

14. Gross measures take no account of the 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

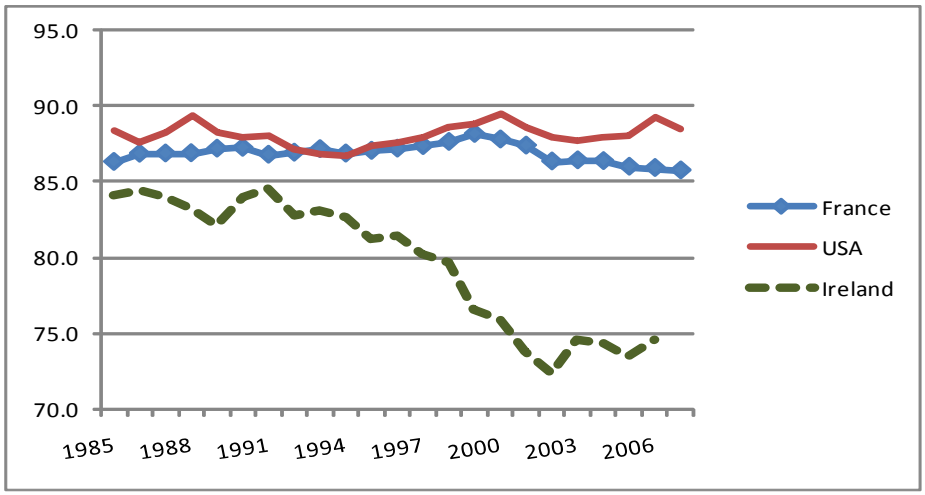
needed. The reason that economists have relied more heavily on GDP than on net domestic product (NDP) is, in part, that depreciation is hard to estimate. When the structure of production remains the same, GDP and NDP move closely together. But in recent years, the structure of production has changed. Information technology (IT) assets have gained importance as capital goods. Computers and software have a shorter life expectancy than do 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 grew at a somewhat slower rate than GDP.

15. Of greater concern for some countries is that standard depreciation measures have not taken into account the degradation in quality of the natural environment. There have been various attempts to widen the scope of depreciation to reflect environmental degradation (or improvements, if such is the case), but without much success. The hurdle is the reliable measurement and monetary valuation of changes in environmental quality.

16. The case of natural resource depletion is slightly different – there is at least a market price, even if it does not reflect environmental damage attributable to the use of the natural resource. Depletion could be captured by excluding the value of the natural resources harvested from the production value of sectors like mining and timber. Their production would then consist only in a pure extraction or logging activity, with a corresponding decrease in GDP. A second possibility would be to take resource depletion into account in depreciation measures. In this case, GDP would be unchanged, but NDP would be lower.

17. In a world of globalization, there may be large differences between the *income* of a country's citizens and measures of domestic *production*, but 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 national accounts. Figure 1.1 below shows how Ireland's income declines relative to its GDP – a reflection of an increasing share of profits that are repatriated by foreign investors. While the profits are included in GDP, they do not enhance the spending power of the country's citizens. For a poor developing country to be told that its GDP has gone up may be of little relevance. It wants to know whether its citizens are better-off, and national income measures are more relevant to this question than GDP.

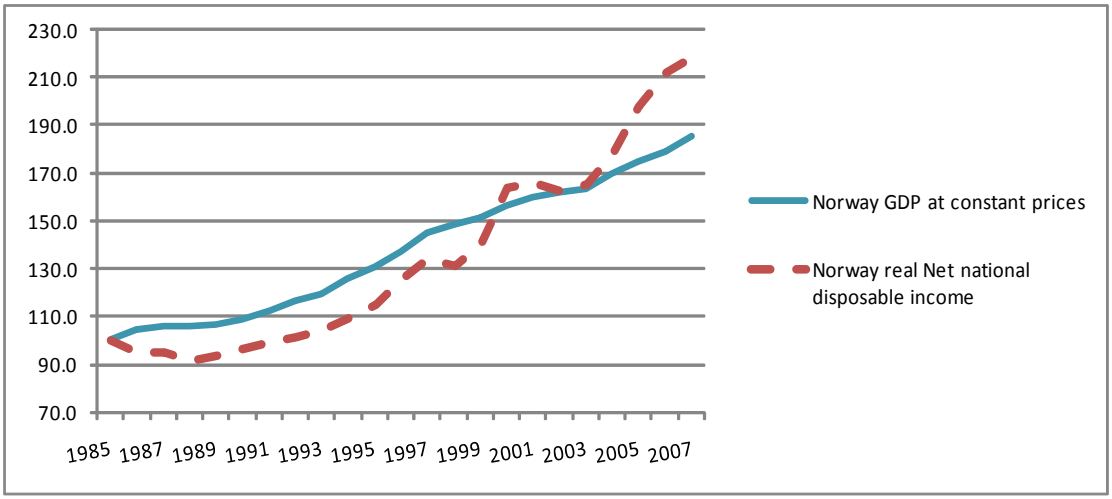
Figure 1.1. Net national disposable income as percentage of gross domestic product



Source: OECD Annual National Accounts.

18. Moreover, the prices of imports evolve very differently from the prices of exports, and these changes in relative prices have to be taken into account in assessing living standards. The figure below shows the divergence between real income and production in 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 have tended to fall relative to import prices, the opposite will be true.

Figure 1.2. GDP and disposable income in Norway



Source: OECD Annual National Accounts.

3.2 - Improving the measurement of services in general

19. In today's economies, services account for up to two-thirds of total production and employment, yet measuring the prices and volumes of services is more difficult than for goods. Retail services are a case in point. In principle, numerous aspects should be taken into account in measuring the services provided: the volume of goods transacted but also the quality of service (accessibility of the shop, general service level of the staff, choice and presentation of products and so forth). It is difficult even to define these services, let alone to measure them. Statistical offices generally use data on the volume of sales as indicators for the volume of trade services. This method leaves aside most quality change in the trade services provided. What is true for retail holds for many other service industries, including those that are often publicly provided, such as health and education. A greater effort will be needed to come to grips with tracking the quantity and quality of services in modern economies.

3.3 - Improving the measurement of government-provided services in particular

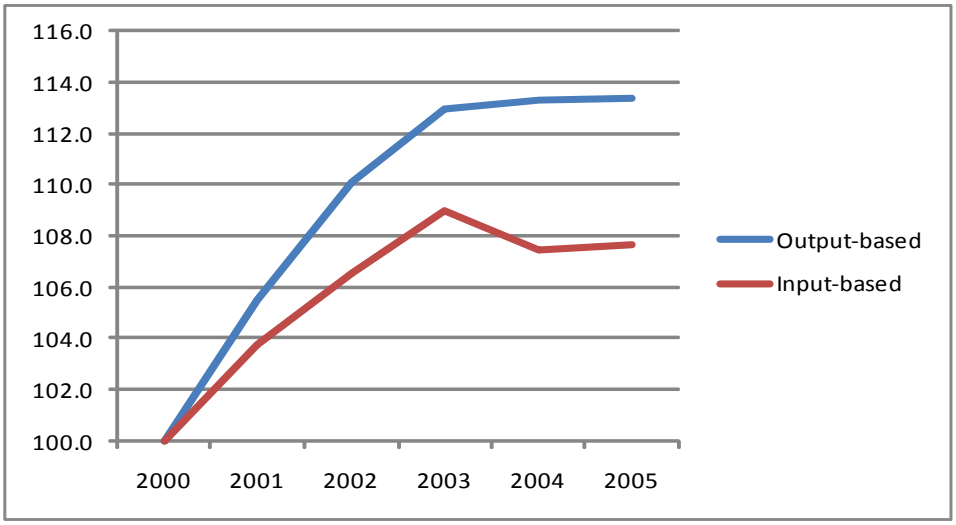
20. Governments play an important part in today's economies. Broadly speaking, they provide two types of services – those of a "collective" nature, such as security, and those of an "individual" nature, such as medical services and education. This does not imply that government is necessarily the only 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 contribution of collective services to citizens' living standards, individual services, particularly education, medical services and public sports facilities, are almost certainly valued positively by citizens. These services tend to be large in scale but 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 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 positive productivity growth in the public sector, our measures under-estimate growth.

21. Work has started in many countries to develop output measures for these government-provided services that are independent of inputs, but the task is formidable. Take the following example: the United States spends more per capita on health care than many European countries, yet in terms of standard health indicators, outcomes are worse. Does this mean that Americans receive 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 American society other than health expenditure? We need to break down the change in health expenditure 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 population's state of health. The problem is that the link between health care expenditure and health status is tenuous at best: expenditures relate to the resources that go into the institutions providing health services, whereas the health status of the population is driven by many factors – and the situation is much the same for education. For example, people's lifestyles will affect health outcomes, and the time parents spend with their children will affect exam scores. Attributing changes in health or education status solely to hospitals or schools and the money spent on them neglects all these factors and can be misleading.

22. 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. One major challenge to these efforts is, once again, to capture quality change. Without a good measure of quality (or equivalently, a good estimate of increases in productivity), it is impossible to ascertain whether the conventional input measures underestimate or overestimate growth. If undifferentiated quantity measures are used, such as a simple number of students or of patients, changes in the composition and quality of the output may be missed. But one has to start somewhere and, because the numbers involved are important, the issue cannot be ignored, For example, with output-based measures, the U.K. economy grew at a 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 the measurement of health output points the other way: output-based production of health services grew more rapidly than input-based production (Figure 1.3. Volume output of health services in Denmark).

23. An important criterion for the reliability of output-based measures is that they are based on observations that are detailed enough to avoid mixing up true volume changes 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 educational services has increased. This may be misleading, however, if costs have gone up because students are taught in smaller classes or if there is a larger share of students that take up engineering studies, which are more costly. The measurement mistake arises because the simple number of students is too undifferentiated an output measure to be meaningful, so a more detailed structure is needed. It would help, for instance, to treat one hour taught to a graduate engineering student as a different product from one hour taught to a first-year arts student, and thus to account for some quality and compositional change. A similar reasoning applies for health care: the treatments of different diseases have to be considered as different medical services. As it turns out, the health-care systems of some countries do provide the administrative data needed to obtain this detailed information. We conclude that despite this being a daunting task, the better measurement of government-provided individual services is central to the better assessment of living standards. Exploiting new administrative data sources is one way of making progress in this direction. Ideally, the information would also capture service quality, for instance, the way patients are accommodated in hospitals or the time devoted to them by the medical staff, though such data may be hard to collect. In this case, new primary data sources such as surveys may be necessary.

Figure 1.3. Volume output of health services in Denmark



Source: Deveci, Heurlén and Sørensen (2008) "Non-Market Health Care Service in Denmark – Empirical Studies of A, B and C Methods"; paper presented at the meeting of the International Association for Research on Income and Wealth, Slovenia.

24. Improving the volume measures of outputs does not dispense with the need to improve – and publish – the volume measures of inputs. Only if both the outputs and inputs of service production are well captured will it be possible to estimate productivity change and undertake productivity comparisons across countries.

3.4 - Revisit the concept of "defensive" expenditures

25. Expenditures required to maintain consumption levels or the functioning of society could be viewed as a sort of intermediate input – there is no direct benefit, and in this sense they do not give rise to a final good or service. Nordhaus and Tobin, in their seminal 1973 paper, 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. Many such "defensive expenditures" are incurred by government, while others are incurred by the private sector. By way of example, expenditure on prisons could be considered a government-incurred defensive expenditure and the costs of commuting to work a privately-incurred defensive expenditure. A number of authors have suggested treating these expenditures as intermediate rather than final products. Consequently, they would not be part of GDP.

26. At the same time, difficulties abound when it comes to identifying which expenditures are "defensive" and which are not. For instance, if a new park is opened, does this constitute defensive expenditure against the disamenties of urban life or is it a non-defensive recreational service? What are the possible ways forward? Some options include:

> *First, focus on household consumption rather than total final consumption*. For many purposes, the former is a more meaningful variable. And all of governments' collective consumption expenditures (which would include things like prisons, military

expenditure and the clean-up of oil spills) are automatically excluded from household final consumption.

Second, widen the asset boundary. In many cases, defensive expenditures include elements of *investment* and *capital* goods. In those cases, they should be treated much like maintenance expenditures in the case of conventional production. For example, health expenditures could be seen as investment in human capital instead of as final consumption. If there is an asset that captures environmental quality, expenditures made to improve or maintain it could also be considered an investment. Conversely, 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 gross measures.

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. With the exception of the consumer's purchase of a ticket for a commuter train, which counts as final consumption, none of the above flows enter measures of production and income. This could be remedied by allowing for the household production of transportation services, which would be considered as an unpaid delivery of intermediate inputs to firms, "subsidized" by private households. Although this treatment would not change overall GDP, it would show a larger contribution to production by households and a smaller contribution by firms.

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

3.5 - Income, wealth and consumption have to be considered together

28. Income flows are an important gauge for the standard of living, but in the end it is consumption and consumption possibilities *over time* that matter. The time dimension brings in wealth. A low-income household with above-average wealth is better off than a low-income household without wealth. The existence of wealth is also one reason why income and consumption are not necessarily equal: for a given income, consumption can be raised by running down assets or by increasing debt, and consumption can be reduced by saving and adding to assets. For this reason, wealth is an important indicator of the sustainability of actual consumption.

29. The same holds for the economy as a whole. To construct the balance sheet of an economy, we need to have comprehensive accounts of its assets (physical capital – and probably 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, natural, human and social capital) minus depreciation and depletion in those same assets.

30. Although information about some central aspects of household wealth is in principle available from national accounts balance sheets, it is often incomplete. Furthermore, certain assets are not recognized as such in the standard accounting framework, not least of all human capital. Studies that have computed monetary estimates of human capital stocks found that they account for an overwhelming part of all wealth (80% or more). A systematic measurement of human capital stock 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.

31. 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. But there may be no markets for certain assets or no trading on the markets, as has recently been the case for certain financial assets. This raises the question of how to value them. And even when market prices do exist, transactions correspond only to a small fraction of the existing stock, 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 the various sectors and the financial risks to which they are exposed.

4 - Bringing out the household perspective

32. Income can be computed for private households as well as for the economy as a whole. Some of citizens' income is taken away in the form of taxes, and so 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. A commonly employed measure of household income adds and subtracts these transfer payments. The resulting measure is referred to as a measure of household disposable income. However, disposable income captures only 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

33. 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 on quality or access.* This is where a purely market-based measure of income meets its limits and where a measure that corrects for differences in institutional arrangements 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.

34. The meaning of adjusted disposable income is best explained by way of an example (Table 1.1). Assume that a society's labour income equals 100 and that individuals who 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 (the actuarial value of a payment of 8) and 2 units of consumption of insurance services. At the same time, persons who are sick receive 8 units as reimbursement of their health expenditures. In this case – let us call it Case A – no taxes are paid and insurance claims and premiums offset each other, so that household disposable income 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, household disposable income has fallen, to 90 currency units. Thus, disposable income here 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.

35. The above example leaves aside, however, any consideration about which insurance regime operates more cost effectively and about the profits that might be made by private insurance companies – it was simply assumed that the 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 make a general observation about the relative efficiency of such schemes. If the insurance 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 and decreased insurance prices. Even if profits are distributed to households in the form of dividends, the change in the form of provision (from private to public) can increase the accessibility of the insurance service. Having an opportunity to insure against certain types of risks has a positive impact on the well-being of people who are risk adverse.

Table 1.1. Private and public insurance schemes

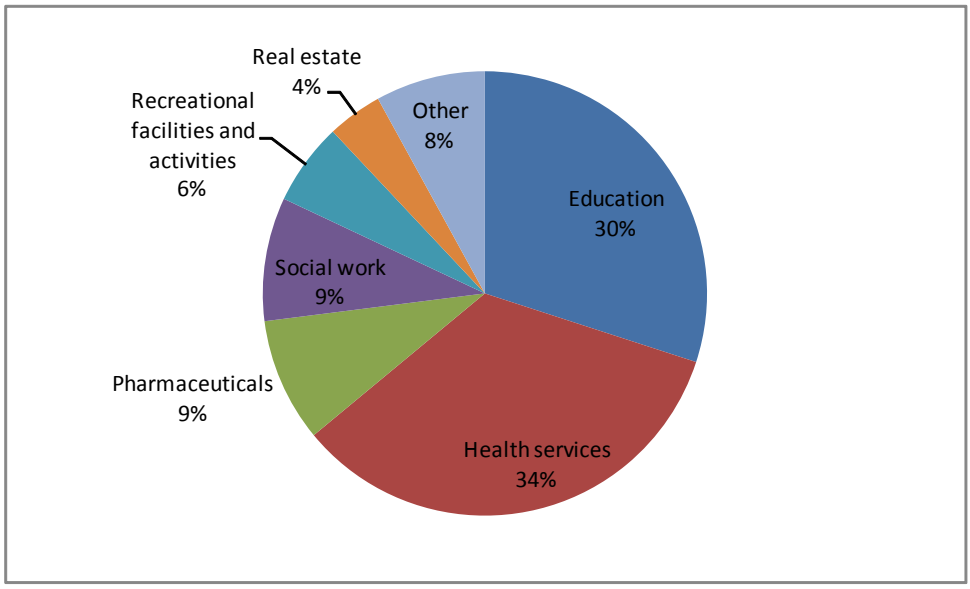
	Private insurance scheme (Case A)	Public insurance scheme (Case B)
Labour income	100	100
Tax	0	- 10
Insurance premiums (excluding insurance services)	- 8	0
Insurance claims	**+ 8**	**0**
Household disposable income	100	90
Social transfers in kind : – reimbursements – running costs of the insurance	0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

36. While the failure to estimate the value of the insurance 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 cost 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. In this situation, the result of using adjusted household income would be a large-scale overestimation of the level of household income and consumption. Some of this can be tackled by using output-based volume measures for the health and education services produced by government. It is also likely that different parts of the population benefit unequally from social transfers in kind provided by government. There is thus an important distributional aspect.

37. Major items included 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, which in 2007 cost about € 290 billion. Education and health services each account for about one-third of total transfers in kind, and housing and recreational and cultural activities (museums, public parks, etc.) account for about 10% (Figure 1.4. Social transfers in kind from general government, France 2007).

Figure 1.4. Social transfers in kind from general government, France 2007



Source: INSEE.

4.2 - Medians and means – distribution of income, consumption and wealth

38. Average measures of per-capita income and wealth give no indication of how the available resources are distributed across persons or households. Similarly, average consumption gives no indication of how people effectively benefit from these resources. 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, consumption and wealth information for different 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, and half below), median consumption and median wealth. The median individual is, in some sense, the "typical" individual. If inequality increases, the differences between medians and averages may well increase, so a focus on averages does not give an accurate

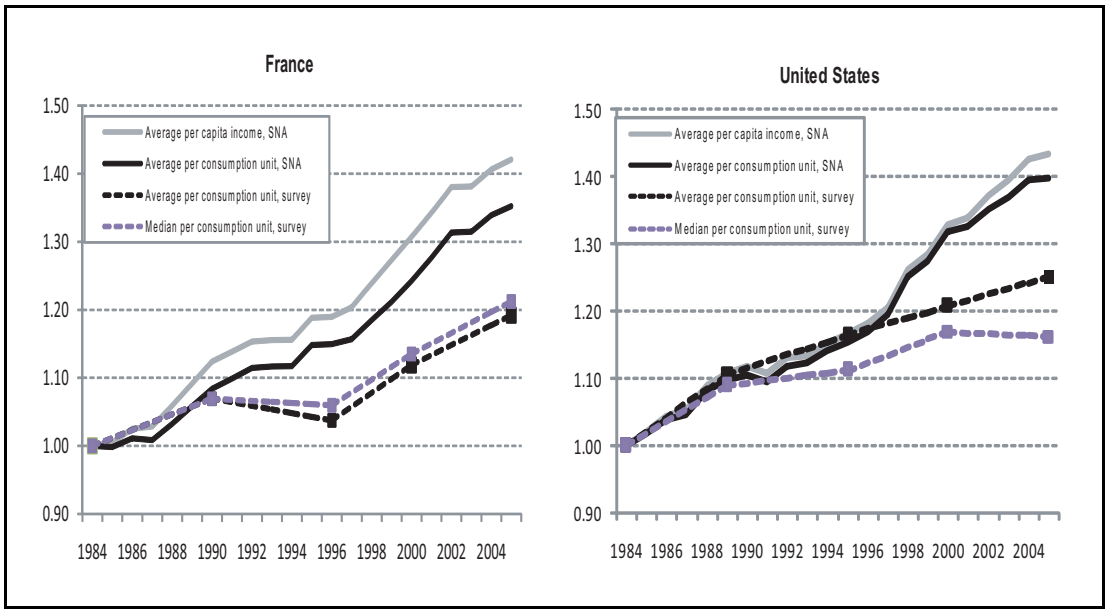
picture of the economic well-being of the "typical" member of society. For example, if all the increases in societal income accrue, say, to those in the top 10%, median income may remain unchanged, while average income increases. Over the past two decades, the dominant pattern in OECD countries is one of a fairly widespread increase in income 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, medians and means would give different pictures of what is happening to societal well-being. Alternatively, changes in the disposable income of different income groups can be tracked. Such an approach would, for instance, look at the numbers of people below a critical income level, or the average income of those in the bottom or top decile. Similar calculations would be useful for consumption and wealth. Empirical research has repeatedly shown that the distribution of consumption can be quite different from the distribution of income. Indeed, the most pertinent measures of the distribution of material living standards are probably based on *jointly* considering the income, consumption and wealth position of households or individuals.

39. In practice, moving from averages to medians is more difficult than meets the eye. Measures of averages are obtained by dividing aggregates by a population figure. To consider distributional elements, micro-economic information is needed that provides information for individual households or groups of households. Micro-economic measures refer to people living in private households and are typically derived from household income surveys, whereas macro-economic measures from the national accounts are based on a range of different sources, and include people living in collective households (such as prisons and institutions for long-term care).

40. An important choice also concerns the unit of measurement. Macro-estimates give totals for a whole country or sector, while micro-data retain the household (or the family) as the 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 rather than per person. Consumption units are households whose size has been adjusted to take account of economies of scale in housing and other costs. This adjustment is of increasing importance as household size shrinks.

41. Against this background, we can consider the evolution of average and median household income in several countries. Figure 1.5. Trends in different measures of household disposable income shows some results for France and the United States. Average income per capita and average income per consumption unit diverge, reflecting a trend towards a 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 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 a widening gap between median and average income, pointing to a more unequal income distribution.

Figure 1.5. Trends in different measures of household disposable income



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

42. There are many measurement issues that can influence the above statements. One 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 move in parallel 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.

43. From the perspective of living standards, what matters is that the distribution of income, consumption and wealth determines who enjoys access to the goods and services produced within a society. Complementing measures of average income by measures with a distributional element is thus a crucial task for official statistics. Ideally, such distributional measures should be compatible in scope with average measures from the national accounts.

44. Similarly, the distribution of the volume of consumption is also important. The same dollar may buy different bundles of products, depending on the income group of the purchaser. Going from nominal to real income and from the value to the volume of consumption means applying a price index, raising the question of *whose price index* are we measuring. Conceptual discussions about price indices are often conducted as there exists 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, e.g. poor people spend more on food, and rich people on entertainment. People 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 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.

45. 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 household purchasing power) in France was to develop consumer price indices for owners of dwellings, for households who rent dwellings and for households who are about to purchase dwellings. A full development of price indices differentiated by socio-economic group requires, however, that different prices be collected for different segments of the population, so that socio-economic aspects are taken into account in the data collection design. This is likely to prove difficult and costly, and should constitute a medium-term research objective – a recommendation that echoes a similar conclusion by the 2002 *Panel on Conceptual, Measurement, and other Statistical Issues in Developing Cost-of-Living Indices* in the United States. Such work would not only enhance the quality of deflation procedures, it would also make it easier for citizens to compare their personal situation with some of the income and price data released by statistical offices.

4.3 - Broader measures of household economic activity

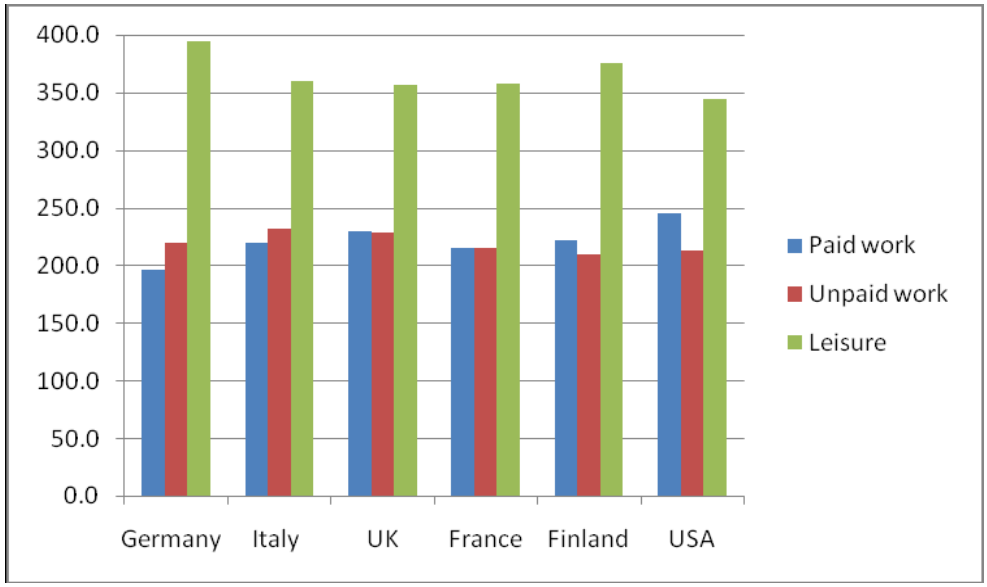
46. There have been major changes in how households and society function. For example, many of the services people received from other family members in the past are now purchased on the market. This shift translates into a rise in income as measured in the national accounts and gives a false impression of a change in living standards, while it merely reflects a shift from non-market to market provision of services. Just as we argued 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 household to market production, or vice-versa, should not affect measured output. We noted earlier that, in practice, current conventions do, however, lead to changes in measured income in both instances.

47. Imagine a two-parent household with two children and an income of 50,000 currency units 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 the shopping, cooks all the meals, does all the cleaning, and performs all the child care. As a result, this household does not need to devote any of its market income to purchasing these services. Now imagine a two-parent household with two children in which *both* parents work full-time for the same global pay (50,000 a year), and neither parent has any time left over for household production or child care. They must pay for all the shopping, cooking, cleaning and child care out of pocket. Their available income is therefore reduced. Conventional measures treat these two households as if they have identical living standards, but obviously they don't. Focusing on market production 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.

48. To get a sense of how important home production is economically, one has to start by examining how people use their time. Figure 1.6. Housework, paid work and leisure1.6 provides a first comparison of time spent per household and per 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, mail and email, and travel time related to all these activities. "Personal care" consists mainly of sleeping, eating and drinking, whereas leisure was defined to include sports, religious and spiritual activities and other leisure activities.

49. Based on these definitions, more time is spent on household production in European countries than in the United States, and more time is spent on leisure in Finland, France, Italy, Germany and the United Kingdom than in the United States (Figure 1.6. Housework, paid work and leisure). Note that some of the classifications are ambiguous, so the results should be read with care. For example, eating and drinking are included in the definition of personal care, whereas, arguably, some eating and drinking is time spent on leisure. The time-use picture would also change if eating time were 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 1.6. Housework, paid work and leisure
Minutes per day and person, latest year available*



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

Source: OECD (2009), *Growing Unequal? Income Distribution and Poverty in OECD Countries*; Paris.

50. If we gloss over these issues, it is possible to come up with an illustrative calculation of the value of household production for France, Finland and the United States. The approach chosen here is simple: the value of the production of household services is measured by its cost. The value of labor is estimated by applying the wage rate of a generalist household worker to the number of hours that people 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 also lack estimates for productivity changes in household production.

51. However, our estimates do 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 sizeable in all countries. Household production amounts to about 35% of conventionally-measured GDP in France (average 1995-2006), about 40% in Finland and 30% in the United States over the same period.

52. Once one starts thinking about non-market income, one also 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 the situation of well-being in these respects. Focusing only on goods and services can therefore bias comparative measures of living standards. This is of particular concern as the world begins to come to terms with environmental constraints. It may not be possible to increase the production, especially of goods, beyond limit, because of the environmental damage that this would entail. Taxes and regulations may be imposed that will discourage production. However, it would be a mistake if, as a result of these measures, we were to conclude that living standards have fallen when leisure time (and environmental quality) has increased. As society progresses, it is not unreasonable to expect people to enjoy some of the fruits of that progress in the form of leisure. Different societies may respond differently to higher living standards, and we do not want to bias our judgments (e.g. of success) *against* societies that choose to enjoy more leisure.

53. Measurement of the value of leisure starts, once again, from time-use data. We multiply the average leisure time per day by the working-age population and then by the average wage rate in the economy. Again, this procedure raises many measurement issues, but the purpose here is to show that estimates are feasible and can produce meaningful cross-country comparisons. For the three countries at hand, the value of leisure roughly doubles net household disposable income in nominal terms. More interesting than nominal income levels is the question of how considering leisure affects the measured growth rates of *real* income and their comparisons across countries. This is captured in Table 1.2. Household income in real terms. 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), the income growth rates of the three countries turn out to be very similar.

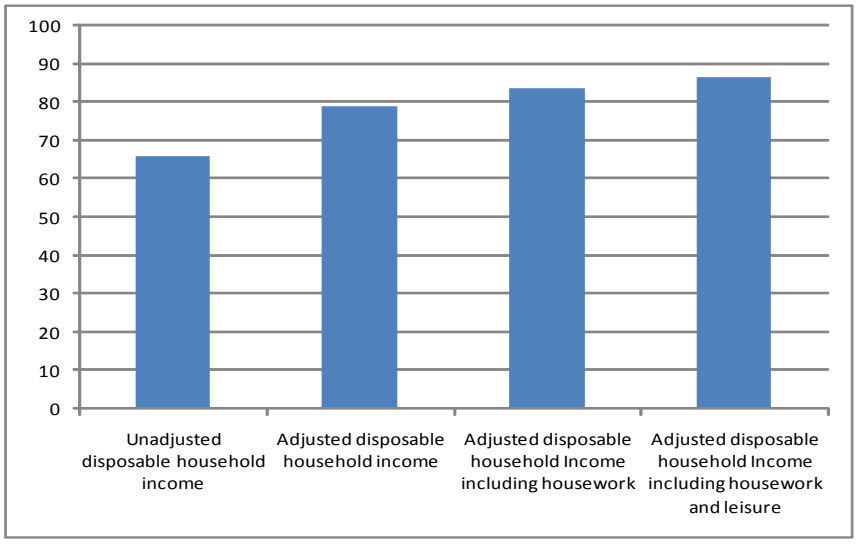
Table 1.2. Household income in real terms
Percentage change at annual rate, 1995-2006

	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%

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

55. More instructive than estimating the rate of change in real income is assessing how household production and leisure bear on the comparison of income *levels* across countries. Income levels should be compared in real terms, so we construct currency converters, so-called *Purchasing Power Parities (PPP)* that permit comparisons of "full" income (including housework and leisure) across countries. Figure 1.7. Real income per capita in France compared to the United States, 2005 compares three income aggregates for France and the United States. The first comparison uses the established disposable income measure. Here, France's per-capita income is about 66% of the comparable United States figure. Adding in government-provided services such as health and education narrows the gap to 79%. If, in addition, housework and leisure are accounted for, one ends up with a relative income level of 87%.

Figure 1.7. Real income per capita in France compared to the United States, 2005
Unites States = 100



4.4 - Distribution of full income

56. It was argued earlier that measures of average income should be accompanied by measures that also provide distributional information. The rationale for examining income distribution holds for market income, but also for broader measures, such as full income. The recognition of the own-account production of household services and leisure affects aggregate measures of income and production, but may also change the established picture of income distribution.

57. Developing distributional measures of full income is, however, a formidable task. The most difficult challenge is to allocate to various 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. Other imputations for own-account services produced by households also fall under this category, as do the distributional effects of government services that are provided in kind.

58. Again, measurement difficulties should not prevent us from developing a more comprehensive picture of the distribution of income and wealth. The distribution of full income should be firmly anchored in the research agenda.

4.5 - Main messages and recommendations

Recommendation 1: Look at income and consumption rather than production.

59. 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 mainly measures market production, though it has often been treated as if it were a measure of economic well-being. Conflating the two can lead to misleading indications about how well-off people are and entail the wrong policy decisions. Material living standards are more closely associated with measures of real income and consumption – production can expand while income decreases or vice versa when account is taken of depreciation, income flows into and out of a country, and differences between the prices of output and the prices of consumer products.

Recommendation 2: Consider income and consumption jointly with wealth.

60. Income and consumption are crucial for assessing living standards, but in the end they can only be gauged in conjunction with information on wealth. A vital 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 need comprehensive accounts of its assets (physical capital – and probably even human, natural and social capital) and its liabilities (what is owed to other countries). Balance sheets for countries are not novel in concept, but their availability is still limited and their construction should be promoted. There is also a need to "stress test" balance sheets with alternative valuations when market prices for assets are not available or are subject to bubbles and bursts. Measures of wealth are also central to measuring sustainability. What is carried over into the future necessarily has to be expressed as stocks – of physical, natural, human or social capital. Here too the right valuation of these stocks plays a crucial role.

Recommendation 3: Emphasise the household perspective.

61. While it is informative to track the performance of economies as a whole, trends in citizens' material living standards are better followed through measures of household income and consumption. Indeed, the available national accounts data shows that in a number of OECD countries real household income has grown quite differently from real GDP, and typically at a lower rate. The household perspective entails taking account of payments between sectors, such as taxes going to government, social benefits coming from government, and interest payments on household loans going to financial corporations. Properly defined,

household income and consumption should also reflect the value of in-kind services provided by government, such as subsidized health care and educational services.

Recommendation 4: Give more prominence to the distribution of income, consumption and wealth.

62. Average income, consumption and wealth are meaningful statistics, but they do not tell the whole story about living standards. For example, a rise in average income could be unequally shared across groups, leaving some households relatively worse-off than others. Thus, average measures of income, consumption and wealth should be accompanied by indicators that reflect their distribution. Ideally, such information should not come in isolation but be linked, i.e. one would like information about how well-off households are with regard to all three dimensions of material living standards: income, consumption and wealth. After all, a low-income household with above-average wealth is not necessarily worse-off than a medium-income household with no wealth. The desirability of disposing of information on the "joint distribution" of dimensions will be encountered once again in Recommendation 3 of the Chapter on the quality of life.

Recommendation 5: Broaden income measures to non-market activities.

63. There have been major changes in how households and society function. For example, many of the services people received from other family members in the past are now purchased on the market. This shift translates into a rise in income as measured in the national accounts and may give a false impression of a change in living standards, while it merely reflects a shift from non-market to market provision of services. Many services that households produce for themselves are not recognized in official income and production measures, yet they constitute an important aspect of economic activity. While their exclusion from official measures reflects uncertainty about data more than it does conceptual dissent, more and more systematic work in this area should be undertaken. This should start with information on how people spend their time that is comparable both over the years and across countries. Comprehensive and periodic accounts of household activity as satellites to the core national accounts should complement the picture.

CHAPTER 2 - QUALITY OF LIFE [2]

1 - Introduction

65. Quality of life is a broader concept than economic production and living standards. It includes the full range of factors that influences what we value in living, reaching beyond its material side. While some extensions of economic accounting (discussed in Chapter 1) allow including some of the elements that shape quality of life in conventional measures of economic well-being, every approach based on resources (or on people's command over commodities) remains limited in important ways. First, resources are means that are transformed into well-being in ways that differ across people: 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. Second, many resources are not marketed, and even when they are, prices will differ across individuals, making it problematic to compare real income across people. Finally, many of the determinants of human well-being are aspects of people's life-circumstances: they cannot be described as resources with imputable prices, even if people do make trade-offs among them. These arguments by themselves are sufficient to suggest that resources are an insufficient metric for quality of life. Which other metric should be used instead for assessing quality of life depends on the philosophical perspective taken.

66. While a long tradition of philosophical thought has addressed the issues of what gives life its quality, recent advances in research have led to measures that are both new and credible. This research suggests that the need to move beyond measures of economic resources is not limited to developing countries (the traditional focus of much work on "human development" in the past) but is even more salient for rich industrialised countries. These measures, while not *replacing* conventional economic indicators, provide an opportunity to *enrich* policy discussions and to inform people's view of the conditions of the communities where they live. More importantly, the new measures now have the potential to move from research to standard statistical practice. While some of them reflect structural conditions that are relatively invariant over time but that differ systematically across countries, others are more responsive to policies and more suitable for monitoring changes over shorter periods of time. Both types of indicator play an important role in evaluating quality of life.

2. Evidence and references in support of the claims presented in this Summary are presented in a companion technical report.

2 - Conceptual approaches to measuring quality of life

67. Three conceptual approaches have retained the attention of the Commission as useful in thinking about how to measure quality of life.

- 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 is a universal goal of human existence.

- 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 participate actively in political life. The foundations of the capability approach, which has strong roots in philosophical notions of social justice, reflect a focus on human ends and on respecting the individual's ability to pursue and realise the goals that he or she values; a rejection of the economic model of individuals acting to maximise their self-interest heedless of relationships and emotions; an emphasis on the complementarities between various capabilities; and a recognition of human diversity, which draws attention to the role played by ethical principles in the design of the "good" society.

- 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, and obtaining information on people's current situations and on their preferences with respect to these points. This approach avoids the pitfall of basing evaluations on an "average" willingness-to-pay that may disproportionately reflect the preferences of those who are better-off in society and focuses instead on equality among all of its members.

68. These approaches have obvious differences, but also certain similarities. For example, subjective well-being is sometimes claimed to encompass all capabilities, in so far as these refer to attributes and freedoms that people *value* (implying that enhancing their capabilities will improve people's subjective states). However, proponents of the capability approach also emphasise that subjective states are not the only things that matter, and that expanding people's opportunities is important in itself, even if this does not show up in greater 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, they all point to the importance of a number of features 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

69. For a long time, economists have assumed that it was sufficient to look at people's 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, and this 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.

70. Subjective measures have always been part of the traditional tool-kit of economists and statisticians, as many features of our economies and societies are measured through people's responses to a standard set of questions (for example, "unemployment" is typically measured based on people's answers as to whether they worked at all in a specific reference week, whether they actively looked for a job, and whether they would be available to start working in the near future). 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 to predict people's behaviour (e.g. workers who report more dissatisfaction in their work are more likely to quit their job), and that they are valid 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).

71. Subjective approaches distinguish between the *dimensions* of quality of life and the objective *factors* shaping these dimensions. In turn, the subjective dimensions of quality of life encompass several aspects. The first is represented by people's evaluations of their life as a whole or of its various domains, such as family, work and financial conditions. These evaluations imply a cognitive exercise by each person and an effort to take stock of and summarise the full range of elements that people value (e.g. their sense of purpose, the fulfilment of their goals and how they are perceived by others). 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 to social pressure 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 characterise the experience of each person.

72. 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 lives. Which of these aspects matters 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 bad choices, and some choices are made unconsciously rather than by weighing the pros and cons of various alternatives.

73. Subjective reports 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 provide information about the determinants of quality of life at the level of *each person*. These determinants include both features of the environment where people live and their individual conditions, and they vary depending on the aspect considered. For example, activities (such as commuting, working, or socialising) may be more important for affects, while conditions (such as being married, or having a rewarding job) may be more important for life evaluations. In both cases, however, these measures provide information beyond that conveyed by income. For example, in most developed countries younger and older people report higher evaluations of their life than prime-age people, a pattern that contrasts sharply with levels of income for the same groups.

74. One area where various subjective measures of people's well-being agree is in pointing to the high 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 a 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.

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

4 - Objective features shaping quality of life

76. Both the capability and the fair allocation approaches give prominence to people's objective conditions and the opportunities available to them, 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 lives.

77. The range of objective features to be considered in any assessment of quality of life will depend on the purpose of the exercise: is the goal to assess changes in conditions within national jurisdictions, or to compare these conditions across countries at different levels of development? Some features may matter as descriptors of people's states (e.g. health), while others may reflect the freedoms that people have to pursue the goals that they value (e.g. political voice). While the question of which elements should belong to a list of objective features inevitably depend on value judgements, in practice most of these themes are shared across countries and constituencies, and there is a large degree of consistency among the various exercises that focus on measuring "well-being" and related concepts.[3] In general, measures for all these objective features highlight that how societies are organised makes a

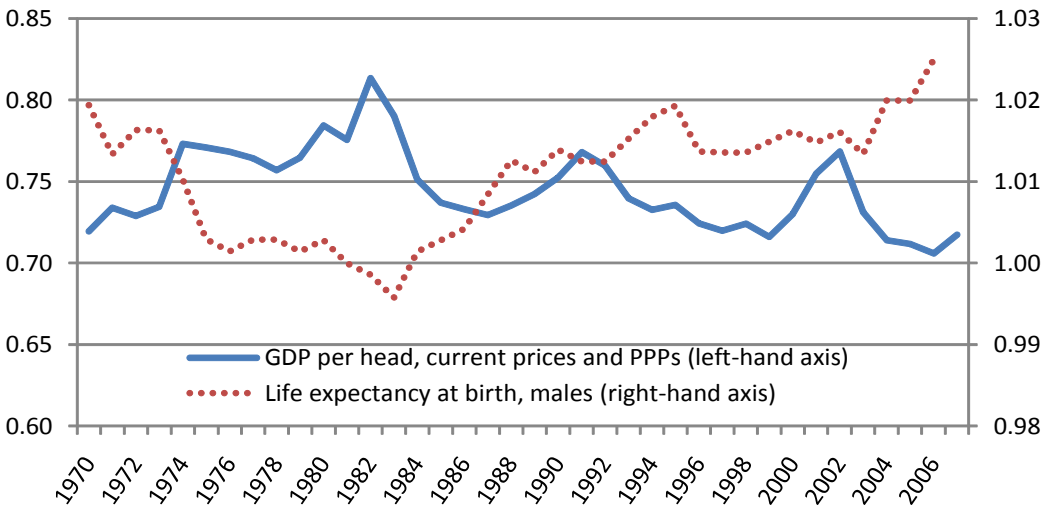
3. See, for example, the Taxonomy developed by the OECD in the context of the "Global Project on Measuring the Progress of Societies' (www.oecd.org/progress/taxonomy).

difference for people's lives, and that their influences are not all captured by conventional measures of economic resources.

4.1 - Health

78. *Health* is a basic feature shaping both the length and the quality of people's lives. 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, which 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 results and rankings when comparing countries with survival curves (by age) that cross each other. This suggests that a variety of mortality measures should be compiled and regularly monitored. Nonetheless, it is significant that non-monetary measures of people's health can diverge significantly from conventional economic measures. For example, although France has a lower GDP per capita than the United States, its life-expectancy at birth is higher, and this advantage has been widening (from less than 6 months in 1960 to almost 2 years in 2006) even while its GDP per capita relative to the US was falling (Figure 2.1).

Figure 2.1. Gaps in GDP per capita and life expectancy at birth between the United States and France



Note: Ratios of French values to US values (values greater than 1 indicate better conditions in France than in the US). For example, in 2006 French GDP per capita was 0.82 of the US level, while the life expectancy of French men was 1.025 times that of men in the United States.

Source: OECD data.

79. The state of progress is far more limited for statistics on morbidity, a situation that has led to long-standing disagreements about whether 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; diagnoses by health professionals; registers for specific diseases; and self-reports drawn for censuses and surveys. Some of these measures relate to the prevalence of diseases or injuries, while others refer to their consequences in terms of the functioning of the person affected (which also depends on the quality of treatment). Variations in the measures and underlying data are inevitable given the many manifestations of poor health, but this also poses a real obstacle to comparing countries and monitoring changes in people's morbidity 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.

80. The variety of dimensions of people's health has led to several attempts to define a summary measure that combines both mortality and morbidity. However, although several combined indices of people's health exist, 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.

81. 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 from lower occupational classes who have less education and income tend to die at younger ages and to suffer, within their shorter lifetimes, a higher prevalence of various health problems. Second, these differences in health conditions do not merely reflect worse outcomes for people at the very bottom of the socio-economic scale but extend to people throughout the socio-economic hierarchy, i.e. they display a "social gradient": for example, life expectancy in the United Kingdom increases when moving from unskilled manual workers to skilled ones, from manual to non-manual workers, from lower-ranked office workers to higher-ranked staff. While these patterns in health inequalities have an obvious relevance for assessing quality of life, existing measures do not allow cross-country comparisons of their magnitude, due to differences in the measures of health outcomes used, in the individual characteristics considered (education, income, ethnicity), and in the reference population and geographic coverage of the various national studies.[4]

4.2 - Education

82. 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, even after controlling for the higher income it brings. Further, better-educated people typically have better health status, lower unemployment, more social connections, and greater engagement in civic and political life. While the available evidence does not always allow conclusions about the directionality

4. Note however that there is ongoing research to measure socio-economic inequalities to health in a standard way, see for example the European Union Working Group in Socio-economic Inequalities to Health.

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

83. Available educational indicators cover a broad range of fields. Some refer to inputs (e.g. school enrolment, educational expenditures and school resources), while others refer to throughputs and outputs (e.g. graduation rates, 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 a country's development and on the goal of the evaluation exercise. The available indicators highlight large differences across countries, with various educational indicators sometimes highlighting contrasting patterns. Some countries, for example, may combine excellence for students that reach university education with widespread under-achievement for a large number of youth, mainly from households at the bottom of the socio-economic ladder. These differences would cancel out in summary measures of education (e.g. mean years of schooling) but have significance for any assessment of quality of life. Within countries, measures of inequality in learning outcomes are especially important for youth at the bottom of the achievement scale who are at risk 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 lives, 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 lives.

84. Some of the most relevant indicators for assessing the effect of education on quality of life are measures of people's competencies. Several tools have been developed in recent years to measure these in standardised ways, though the tools still have significant limitations. First, and most obviously, 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, skills development and improvements in quality of life. Information about the experiences and "soft" competencies acquired by children in their early years remains limited, despite increasing evidence that early-childhood experiences matter for people's learning and quality of life in later years. Measurement tools also remain limited when it comes to comparing the competencies of students in higher education and to 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, 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.

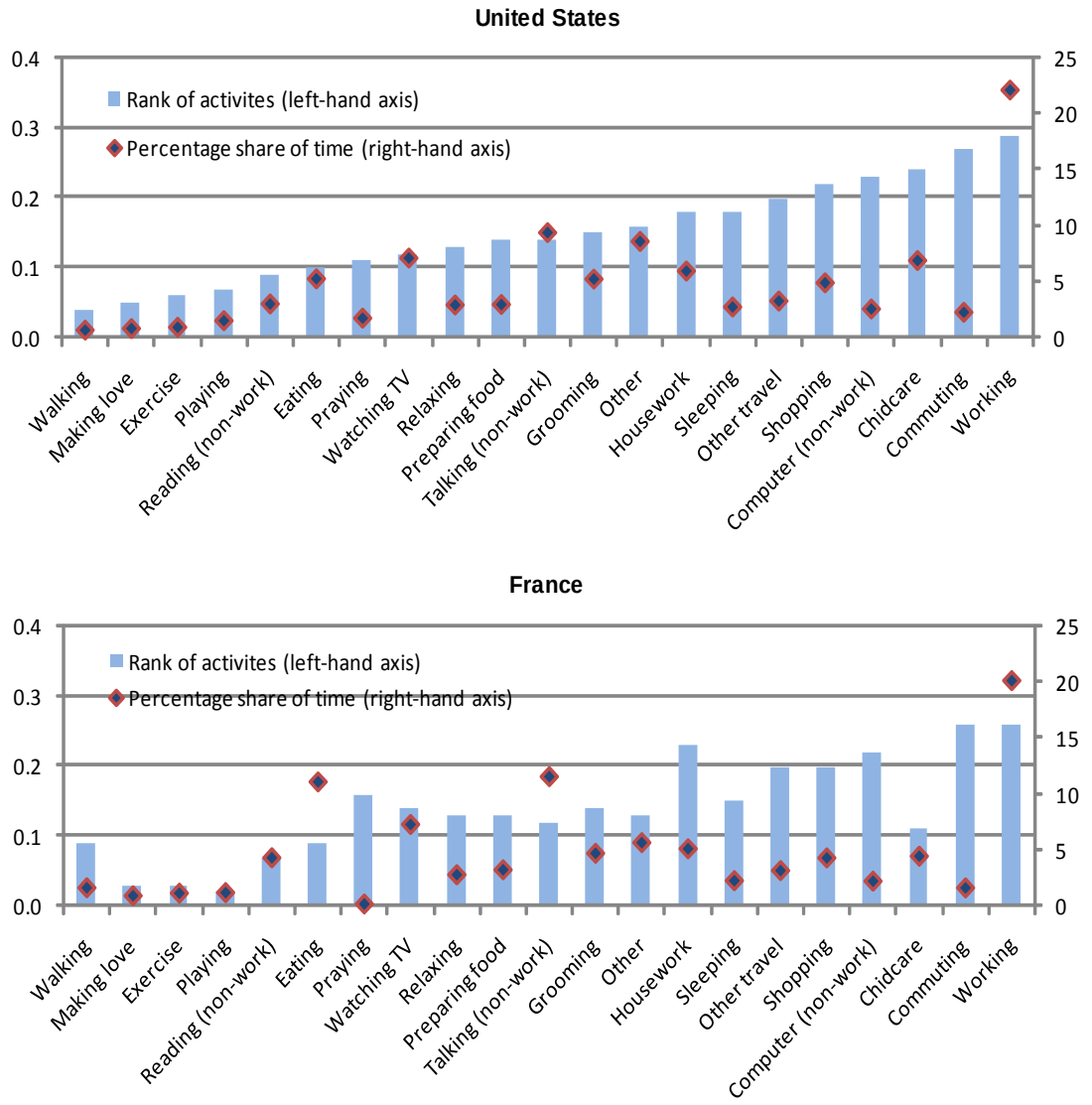
4.3 - Personal activities

85. How people spend their time and the nature of their personal activities matters for quality of life, irrespectively of the income generated. The activities that people engage in have effects on their subjective well-being, in terms of both their hedonic experiences

(Figure 2.2) and their evaluative judgements. More generally, people do not always "choose" among these activities in the same way as they allocate their budget among various goods, due to a lack of effective alternatives. Further, these choices will generally affect other people within the family and community, with some of these personal activities effectively representing indirect costs to production (e.g. commuting) rather than consumption.

Figure 2.2. Ranking of personal activities based on women's hedonic experiences and time devoted to them in selected cities in the United States and France

Activities ranked in decreasing order of enjoyment in the United States



Note: The ranking of activities is based on information on the proportion of 15-minute intervals in which the hedonic experience of "stress", "sadness" or "pain" exceeded that of "happiness". Data refer to a sample of women in Columbus (Ohio, United States) and Rennes (France), interviewed in 2006 with the *Princeton Affect and Time Survey.*

Source: Krueger, A.B., D. Kahneman, D. Schkade, N. Schwarz and A. Stone (2008), "National Time Accounting: The Currency of Life", NBER, forthcoming in A. B. Kruger (ed.), *Measuring the Subjective Well-being of Nations: National Accounts of Time Use and Well-Being*, University of Chicago Press, Chicago.

86. Because of both political demands and the feasibility of providing concrete and comparable measures, the main activities discussed by the Commission have been paid work, unpaid work, commuting and leisure time. Housing, although not representing an activity *per se*, was also discussed because it provides the setting for a number of personal activities.

- Paid work matters for quality of life partly because it provides identity to people and opportunities to socialise with others. However, not all jobs are equally valuable in this respect. This underscores the importance of collecting more systematic information on the *quality* of paid work, as a number of international organisations have been doing in the context of their ongoing studies of "decent work". Some national surveys provide information on many aspects of decent work, such as non-standard employment, gender gaps in employment and wages, discrimination in the workplace, opportunities for lifelong learning, access to employment for disabled persons, working time and "unsocial hours", the work-life balance, work accidents and physical risks, work intensity, social dialogue and workers' autonomy. Their practical use is, however, limited by small sample sizes and survey differences across countries.

- Unpaid domestic work, such as shopping and the 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.

- Commuting time is also key to the quality of work, and monitoring it requires information on the number of hours spent travelling to and from work during a specified period, as well as on the accessibility and affordability of transport.

- A long tradition of research has emphasised the importance of leisure-time for quality of life. This research points to the importance of developing indicators of both leisure quantity (number of hours) and quality (number of episodes, where they took place, presence of other people), as well as of 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).

- Finally, despite the importance of housing for a variety of social outcomes (such as children's 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 shelters and on housing quality (e.g. in terms of the local services available and overcrowding).

87. In several cases, suitable indicators in these various fields already exist, and the challenge is to improve 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 (as argued in Chapter 1) an essential input to the construction of satellite household accounts. One priority should be to develop measurement tools grounded on clear definitions and based on surveys with a consistent design that are representative of patterns over a full year and are undertaken with sufficient regularity – all requirements that are rarely met. Ideally, these surveys should cover both the amount of time spent in various activities and the feelings that they produce. This is important, as the same activity can

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 between different groups in society (e.g. by gender). While these investments in statistical capacity are costly, and compete with other priorities, their pay-off for quality-of-life research is potentially huge.

4.4 - Political voice and governance

88. Political voice is an integral dimension of the quality of life. Intrinsically, 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 what one perceives to be wrong are essential freedoms. Instrumentally, political voice can provide a corrective to public policy: it can ensure the 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 conflicts and enhances the prospect of building consensus on key issues, with pay-offs for economic efficiency, social equity, and inclusiveness in public life.

89. The opportunities for political voice 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. This 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. However, to realise their potential, legal guarantees require effective implementation and substantive justice, which depend 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.

90. Comparisons based on existing indicators of political voice and governance 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 freedoms and rights. Even in the developed world, however, low trust in public institutions and declining political participation point to a growing gap between how citizens and political elites perceive the functioning of democratic institutions. There are also systematic differences in how different groups exercise political voice, and with respect to fundamental rights and opportunities for civic participation in these countries, especially between citizens and the growing numbers of immigrants.

91. Indicators of political voice and governance should help to evaluate the functioning of multiparty democracy and universal suffrage, the level of participation in government decisions at the local level, and 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. 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). Many of these indicators are typically compiled by bodies outside the boundaries of national statistical systems and are based mainly on the opinion of experts. These indicators need to be complemented, and in some cases replaced, by surveys of citizens' own perceptions of how well the political, legal and executive institutions are functioning, 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

92. Social connections improve quality of life in a variety of ways. People with more social connections report higher life-evaluations, as many of 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. the prevalence of crime and 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 with other types of capital, the externalities stemming from social capital can sometimes be negative: for example, belonging to a group may strengthen a sense of unique personal identity that fuels a climate of violence and confrontation with other groups. This, however, underscores the importance of better analysing the *nature* of these social connections and the *breadth* of their effects, rather than underestimating their significance. The available evidence suggests that social connections benefit people in the networks, with effects on non-participants that depend on both the nature of the group and the effects being considered.

93. The drivers of change in people's social connections are not always well understood. Social connections provide services to people (e.g. insurance, security), and the development of both markets and government programmes may have reduced the ties of individuals with their community thanks to the provision of alternative arrangements. What is clear is that a decline in these ties may negatively affect people's lives, even when their functions are taken up by market or government alternatives that increase the level of economic activity (such as when the informal surveillance of neighbours is replaced by salaried security guards). To avoid a biased assessment of human well-being, measures of these social connections are therefore needed.

94. Research on social connections has traditionally relied on proxy measures, such as the number of individual memberships in associations, 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 of 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 that measure 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 voluntary work 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) by building on the experience accumulated by some countries in these fields.

4.6 - Environmental conditions

95. Environmental conditions are important not only for sustainability, but also because of their immediate impact on the quality of people's lives. 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 water cycles, biodiversity loss and natural disasters that affect the health of ecosystems). Secondly, people benefit from environmental services, such as access to clean water and recreation areas, 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 damage both the properties and the lives of the affected populations.

96. Measuring the effects of environmental conditions on people's lives 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.

97. Much progress has been achieved in the last two decades in terms of measuring environmental conditions (through better environmental data, the regular monitoring of indicators, and accounting tools), understanding their impacts (e.g. evaluation of related morbidity and mortality, labour productivity, the economic stakes associated with climate change, biodiversity change, damage from disasters) and establishing a right of access to environmental information. A range of environmental indicators can be used to measure human pressure on the environment, the responses from administrations, firms and households to environmental degradation, and the actual state of environmental quality.

98. However, from a quality-of-life perspective, existing indicators remain limited in important respects. For example, emissions indicators refer mainly to the aggregate quantities of various pollutants, rather than to the share of people exposed to dangerous doses. Existing indicators should hence be supplemented in a number of ways, including the regular monitoring of the number of premature deaths from exposure to air pollution; the number of people who lack access to water services and nature, or who are exposed to dangerous levels of noise and pollution; and the damage inflicted by environmental disasters. Survey measures of people's own feelings and evaluations of the environmental conditions of their neighbourhood are also needed. Because many of the effects of environmental conditions on quality of life differ across people, these indicators should refer to people grouped according to various classification criteria.

4.7 - Personal insecurity

99. Personal insecurity includes external factors that put at risk the physical integrity of each person: crime, accidents, natural disasters, and climate changes are some of the most obvious factors.[5] In extreme cases, 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, one rationale for having specific measures of their frequency is that their effect on people's emotional lives is very different than that of deaths related to medical conditions, as shown by the large impact of bereavement on people's subjective well-being.

100. Less extreme manifestations of personal insecurity such as crime affect quality of life for a significantly larger number of people, with even larger numbers reporting fear of being a victim of a physical aggression. One of the most remarkable feature of reports on subjective fear of crime is how little they are related to experienced victimisation: countries with a higher share of people reporting fear of crime do not experience a higher victimisation while, within countries, older and richer people feel more unsafe than younger and poorer people, despite being less likely to be a victim of crime.

101. These patterns highlight the importance of developing more regular and reliable measures of personal security to orient public discussion. Victimisation surveys are an essential tool to assess the frequency of crime and the fear it generates. Other tools need to be mobilised to assess other threats to personal security, such as domestic violence and violence in countries ravaged by conflict and war.

4.8 - Economic insecurity

102. Uncertainty about 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, depending on the severity of the shock, its duration, the stigma associated with it, the risk aversion of each person, and the financial implications.

103. Job loss can lead to economic insecurity when unemployment is recurrent or persistent, when unemployment benefits are low relative to previous earnings, or when workers have to accept major cuts in pay, hours or both to find a new job. The consequences of job insecurity 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 does find 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 either to evaluate the security of their present job or to rate their expectation of losing their job in the near future. The fear of job loss can have negative consequences for the quality of life of the workers (e.g. physical and mental illness, tensions in family life) as well as for firms (e.g. adverse impacts on workers' motivation and productivity, lower identification with corporate objectives) and society as a whole.

5. Environmental insecurity is not developed here since this issue is already considered above.

104. Illness can cause economic insecurity both directly and indirectly. 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 illness-related economic insecurity 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 has to stop working and the health (or other) insurance does not provide replacement income.

105. Old age is not a risk *per se*, but it can still imply economic insecurity due to uncertainty about needs and resources after withdrawal 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 future pension payments or to greater needs associated with illness or disability. The second is the risk of volatility in pension payments: while all retirement-income systems are exposed to *some* types of risk, the greater role of the private sector in financing old-age pensions (in the form of both occupational pensions and personal savings) has made it possible to extend the coverage of pension systems in many countries but at the cost of shifting risk from governments and firms towards individuals, thereby increasing their insecurity.

106. The many factors shaping economic insecurity are reflected in the variety of approaches used to measure them. Some approaches try to quantify the frequency of specific risks, while others look at the consequences of a risk that materialises and at the means available to people to protect themselves from these risks (especially resources provided by social security programmes). A comprehensive measure of economic insecurity would ideally account for both the frequency of each risk and its 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 final, even more intractable problem is that of accounting for the long-term consequences for quality of life of the various policies used to limit economic insecurity (through their effects on unemployment and labour-force participation).

5 - Cross-cutting issues

107. Most of the measurement challenges described above are specific to each dimension of quality of life, and the Commission has only hinted at some of the work required, leaving it to agencies with expertise in each field to detail concrete action plans. Other challenges, however, are cross-cutting and are unlikely to be picked up through initiatives undertaken separately in each field.[6] Three of these issues deserve special attention.

5.1 - Inequalities in quality of life

108. The first cross-cutting challenge for quality-of-life indicators is to detail the *inequalities* in individual conditions in the various dimensions of life, rather than just the *average*

6. While insecurity is dealt with as an objective factor shaping quality of life, it can also be considered as a cross-cutting issue because of the large variety of risks the individuals are exposed to. The placement of insecurity among the objective factors has been debated at some length and is conventional.

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 about their own conditions.

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

110. The problems, however, go deeper than developing suitable measures. There are many inequalities, and each is significant in itself: this suggests that we should avoid the presumption that one of them (e.g. income) will always encompass all others. At the same time, certain inequalities may be mutually reinforcing. Gender disparities, for example, while pervasive in most countries and groups, are typically much larger for households with lower socio-economic status: in many developing countries, the combined effect of gender and socio-economic status is often to exclude young women in 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 their health at risk. The measurement of some of these inequalities (such as those related to class and socio-economic status) has contributed, over the years, to a wide array of policies and institutions aimed at reducing their intensity and consequences. Other types of inequality, such as between ethnic groups, are more recent (at least in countries that have experienced large waves of immigration) and are set to become more politically salient in the future as immigration continues.

111. 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 lives, 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 inequality and to identify their underlying causes. It is up to the statistical community to regularly feed these analyses with suitable data.

5.2 - Assessing links across quality-of-life dimensions

112. The second cross-cutting challenge, already alluded to above, is to better assess the *relationship* between the various dimensions of quality of life. Some of the most important policy questions involved 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 relationships, in particular at the individual level, are poorly measured and inadequately understood, ignoring the cumulative effects of multiple disadvantages leads to sub-optimal policies. For example, the loss of quality of life due to being both poor and sick far exceeds the sum of the two separate effects, implying that governments may need to target their interventions more specifically at those who cumulate these disadvantages.

113. Assessing these links across the various dimensions of quality of life is not 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 about the "joint distribution" of the most salient features of quality of life (such as hedonic experiences, health status, education, political voice) across all people. While the full development of 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 that could allow both controlling for people's personal characteristics and better analysing the directionality of causation between the different factors shaping life.

5.3 - Aggregating across quality- of-life dimensions

114. The third cross-cutting challenge to quality-of-life research is to *aggregate* the rich array of measures in a 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 for 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 informational 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 should play a role in answering them.

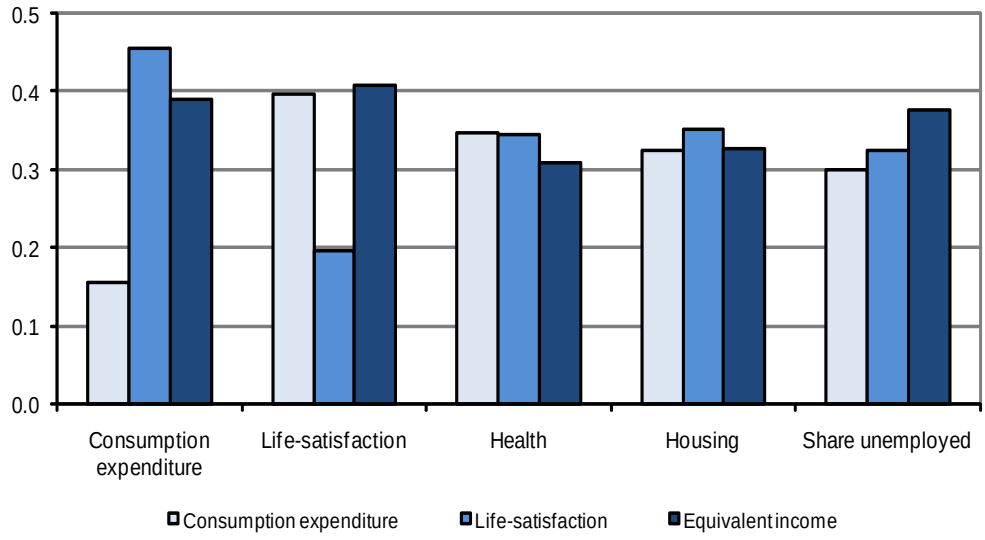
115. Traditionally, the most common response to this demand for parsimony in quality-of-life research has been to aggregate a number of indicators (suitably selected and scaled) of *average* performance in various fields at the country-level. The best known example of this approach is the Human Development Index. This measure has played (and continues to play) 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, the choices on the weights used to construct this (and other similar) indexes 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 in India. More fundamentally, being based on country-averages, these measures ignore the significant correlations between the various features of quality of life across people, and do not say anything about the distribution of these individual conditions within each country. For example, the scalar index will not change if average performance in each domain remains the same while the accumulation of advantages or disadvantages for the same person across various domains of life changes over time.

116. Several aggregate measures of quality of life are possible, depending on the philosophical perspective taken and the question addressed. Some of these measures are already being used sporadically (e.g. average levels of life-satisfaction for a country as a whole, and composite indices such as the Human Development Index, which is mainly focused on developing countries) and could be extended through questionnaire-based measures of people's psychological health, feelings and evaluations and through

consideration of additional dimensions of quality of life. 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 (see Figure 2), requires collecting information on emotional experiences during specific episodes through time-use surveys. Similarly, methods based on counting the occurrences and severity of various objective features for each person (which is linked to the capability approach), before proceeding to construct country-averages, require information on the joint distribution of various objective features. Finally, the notion of "equivalent income" (which is linked to the fair allocations approach) requires information on people's states in various dimensions of quality of life, and on their preferences with respect to these states (for a given reference level in each).

117. In general, different approaches will lead to distinct scalar measures of quality of life for each country, and to different characteristics of the people classified as "worse-off". For example, in a sample of Russian respondents, people in the bottom quintile of the distribution of equivalent income report worse health and a higher incidence of unemployment compared to people identified as "worse-off" based on either their consumption expenditure or their subjective life-evaluations (Figure 2.3). This suggests that, 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.

Figure 2.3. Characteristics of the most deprived people according to different measures of quality of life, Russia in 2000



Note. Data refer to people considered as "worse-off" (i.e. in the bottom quintile of the distribution) according to three different measures of their quality of life: i) household consumption expenditure (adjusted for the number of people in each household); ii) life-satisfaction (based on the question, "*To what extent are you satisfied with your life in general at the present time?"* with answers on a five-point scale); and iii) a measure of equivalent income, based on four "functionings", i.e. self-reported health, employment status, quality of housing, and having incurred wage arrears. For each of these three measures of quality of life, the figure plots the average levels of various factors shaping quality of life among the "worse-off" based on one measure relative to those based on all others.

Source: Fleurbaey M., E. Schokkaert and K. Decancq (2009) "What good is happiness?", CORE Discussion Paper, 2009/17, Université catholique de Louvain, Belgium. Computations based on data from the *Russia Longitudinal Monitoring Survey.*

5.4 - Main messages and recommendations

118. Quality of life includes the full range of factors that make life worth living, including those that are not traded in markets and not captured by monetary measures. While some extensions of economic accounting include some additional elements that shape quality of life in conventional money-based measures of economic well-being, there are limits on how much this approach can achieve. Other indicators have an important role to play in measuring social progress, and recent advances in research have led to new and credible measures for at least some aspects of quality of life. These measures, while not replacing conventional economic indicators, provide an opportunity to enrich policy discussions and to inform people's view of the conditions of the communities in which they live; today, they have the potential to move from research to standard statistical practice. The Commission's recommendations in this field can be summarised as follows:

Recommendation 1: Measures of subjective well-being provide key information about people's quality of life. Statistical offices should incorporate questions to capture people's life evaluations, hedonic experiences and priorities in their own surveys.

119. Research has shown that it is possible to collect meaningful and reliable data on subjective well-being. Subjective well-being encompasses different aspects (cognitive evaluations of one's life, positive emotions such as joy and pride, and negative emotions such as pain and worry): each of them should be measured separately to derive a more comprehensive appreciation of people's lives. Quantitative measures of these subjective aspects hold the promise of delivering not just a good measure of quality of life *per se*, but also a better understanding of its determinants, reaching beyond people's income and material conditions. Despite the persistence of many unresolved issues, these subjective measures provide important information about quality of life. Because of this, the types of question that have proved their value within small-scale, unofficial surveys should be included in larger-scale surveys undertaken by official statistical offices.

Recommendation 2: Quality of life also depends on people's objective conditions and opportunities. Steps should be taken to improve measures of people's health, education, personal activities, political voice, social connections, environmental conditions and insecurity.

120. The information relevant to valuing quality of life goes beyond people's self-reports and perceptions to include measures of their functionings and freedoms. While the precise list of these features inevitably rests on value judgments, there is a consensus that quality of life depends on people's health and education, their everyday activities (which include the right to a decent job and housing), their participation in the political process, the social and natural environment in which they live, and the factors shaping their personal and economic security. Measuring all these features requires both objective and subjective data. The challenge in all these fields is to improve upon what has already been achieved, to identify gaps in available information, and to invest in statistical capacity in areas (such as time-use) where available indicators remain deficient.

Recommendation 3: Quality-of-life indicators in all the dimensions they cover should assess inequalities in a comprehensive way.

121. Inequalities in human conditions are integral to any assessment of quality of life across countries and the way that it is developing over time. Each dimension of quality-of-life requires appropriate measures of inequality, with each of these measures being significant in itself and none claiming absolute priority over others. Inequalities should be assessed across people, socio-economic groups and generations, with special attention to inequalities that have arisen more recently, such as those linked to immigration.

Recommendation 4: Surveys should be designed to assess the links between various quality-of-life domains for each person, and this information should be used when designing policies in various fields.

122. It is critical to address questions about how developments in one domain of quality of life affect other domains, and how developments in all the various fields are related to income. This is important because the consequences for quality of life of having multiple disadvantages far exceed the sum of their individual effects. Developing measures of these cumulative effects requires information on the "joint distribution" of the most salient features of quality of life across everyone in a country through dedicated surveys. Steps in this direction could also be taken by including in all surveys some standard questions that allow classifying respondents based on a limited set of characteristics. When designing policies in specific fields, indicators pertaining to different quality-of-life dimensions should be considered jointly, to address the interactions between dimensions and the needs of people who are disadvantaged in several domains.

Recommendation 5: Statistical offices should provide the information needed to aggregate across quality-of-life dimensions, allowing the construction of different scalar indexes.

123. While assessing quality-of-life requires a plurality of indicators, there are strong demands to develop a single scalar measure. Several scalar measures of quality of life are possible, depending on the question addressed and the approach taken. Some of these measures are already being used, such as average levels of life-satisfaction for a country as a whole, or composite indices that aggregate averages across domains, such as the Human Development Index. Others could be implemented if national statistical systems made the necessary investment to provide the data required for their computation. These include measures of the proportion of one's time in which the strongest reported feeling is a negative one, measures based on counting the occurrence and severity of various objective features of people's lives, and (equivalent-income) measures based on people's states and preferences.

CHAPTER 3 - SUSTAINABLE DEVELOPMENT AND ENVIRONMENT [7]

1. Introduction

124. The first two chapters have dealt extensively with the measurement of current well-being, either along dimensions that can be summed up in monetary units (Chapter 1), or along dimensions that are less amenable to conversion into monetary units (Chapter 2).

125. The sustainability issue that is raised by this last chapter is of a different nature. Sustainability poses the challenge of determining whether we can hope to see the current level of well-being at least maintained for future periods or future generations, or whether the most likely scenario is that it will decline. It is no longer a question of measuring the present, but of predicting the future, and this prospective dimension multiplies the difficulties already encountered in the first two chapters.

126. Despite these difficulties, many proposals have been made for measuring sustainability in quantitative terms, stemming from seminal work such as Nordhaus and Tobin's "Sustainable measure of economic welfare" in the 1970s, or following the strong impulse given by the Brundtland Report in 1987 and the Rio Summit at the turn of the 1990s. The present chapter will start with a short review of these proposals. We shall see that many of them fail to distinguish clearly between the measurement of current well-being and the assessment of its sustainability. To put it very simply, many proposals try to cover all three dimensions examined by the three subgroups of the Commission, and sometimes try to sum them up in a single number. This is not the way the Commission has structured its approach, and with good reason. We firmly believe that sustainability deserves separate measurement, and we shall focus in this chapter on the sustainability issue *stricto sensu*.

127. Such a restriction allows focusing on what the literature calls a "wealth" or "stock-based" approach to sustainability. The idea is the following: the well-being of future generations compared to ours will depend on what resources we pass on to them. Many different forms of resource are involved here. Future well-being will depend upon the magnitude of the stocks of exhaustible resources that we leave to the next generations. It will depend also on how well we maintain the quantity and quality of all the other renewable natural resources that are necessary for life. From a more economic point of view, it will also depend upon how much physical capital – machines and buildings – we pass on, and how much we devote to the constitution of the human capital of future generations, essentially through expenditure on education and research. And it also depends upon the quality of the institutions that we transmit to them, which is still another form of "capital" that is crucial for maintaining a properly functioning human society.

7. Evidence and references in support of the claims presented in this Summary are presented in a companion technical report.

128. How can we measure whether enough of these assets will be left or accumulated for future generations? In other words, when can we say that we are currently living above our means? In particular, is there any reasonable hope of being able to characterize this with one simple number that could play the role for sustainability that GDP has long played for the measurement of economic performance? One reason for such a quest would be to avoid the multiplication of competing numbers. However, if we want to accomplish this, we need to convert all the stocks of resources passed on to future generations into a common metric, be it monetary or not.

129. We shall discuss in some detail why such a goal seems overly ambitious. The aggregation of heterogeneous items seems possible up to a point for physical and human capital or some natural resources that are traded on markets. But the task appears much more complicated for most natural assets, due to the lack of relevant market prices and to the many uncertainties concerning the way these natural assets will interact with other dimensions of sustainability in the future. This will lead us to suggest a pragmatic approach that combines a monetary indicator, which could send us reasonable signals about economic sustainability, and a set of physical indicators devoted to environmental issues. We provide some examples of such physical indicators, yet, in the end, the choice of the most relevant ones must be left to specialists from other fields, before submission to the public debate.

2 - Taking stock

130. Providing a brief summary of the very abundant literature that has been devoted to the measurement of sustainability or durable development is not an easy task. We will use an imperfect but simple typology that distinguishes (1) large and eclectic dashboards, (2) composite indices, (3) indices that consist of correcting GDP in a more or less extensive way, and (4) indices that essentially focus on measuring how far we currently "overconsume" our resources. This last category is itself heterogeneous, since we shall include in it indices as different as the ecological footprint and adjusted net savings, which, as we shall see, convey very different messages.

2.1 - Dashboards or sets of indicators

131. Dashboards or sets of indicators are one widespread approach to the general question of sustainable development. This approach involves gathering and ordering a series of indicators that bear a direct or indirect relationship to socio-economic progress and its durability. In the last couple of decades, international organizations have played a major role in the emergence of sustainability dashboards, with the United Nations playing a prominent role. In particular, the 1992 Rio Summit adopted Agenda 21, whose 40th chapter invites the signatory countries to develop quantitative information about their actions and accomplishments.

132. Other international initiatives to build sustainable development dashboards have been taken by the OECD and Eurostat, following the European Council's adoption of its own Sustainable Development Strategy in 2001. The current version of this dashboard includes 11 indicators for level 1 (Table 3.1), 33 indicators for level 2, and 78 indicators for level 3, with the level 2 and 3 indicators covering 29 sub-themes. Similar national initiatives have accompanied this general movement, albeit in a somewhat scattered way. Local initiatives have also mushroomed over the last decade, some based on the initial impetus from Agenda 21.

133. For the user, the most striking feature of this very abundant literature is the extreme variety of the indicators proposed. Some are very comprehensive – GDP growth retains its place, and is even the first indicator in the European Dashboard – while others are much more specific, such as the percentage of smokers in the population. Some pertain to outcomes, others to instruments. Some can easily be related both to development *and* to sustainability – literacy performance matters for both current well-being and future growth – but others pertain only either to current development or to long-run sustainability. There are even some items whose link with both dimensions is questionable or at least of indeterminate sign: is a high fertility rate a good thing for sustainability? Maybe yes for the sustainability of pensions, but maybe not for environmental sustainability. And is it always a signal of good economic performance? This probably depends on what we consider "high" or "low" in terms of fertility.

134. These dashboards are useful in at least two respects. First, they are an initial step in any analysis of sustainability, which by its nature is highly complex and therefore necessitates an effort at establishing a list of relevant variables and encouraging national and international statistical offices to improve the 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 in some dimensions can compensate for low performance in others. This allows a global assessment of sustainability using mono-dimensional indices. The "strong" approach argues that sustainability requires separately maintaining the quantity or quality of many different environmental items. Following this up therefore requires large sets of separate statistics, each pertaining to one particular subdomain of global sustainability.

135. Dashboards nevertheless suffer because of their heterogeneity, at least in the case of very large and eclectic ones, and most lack indications about causal links, their relationship to sustainability, and/or hierarchies amongst the indicators used. Further, as communications instruments, one frequent criticism is that they lack what has made GDP a success: the powerful attraction of a single headline figure allowing simple comparisons of socio-economic performance over time or across countries.

Table 3.1. Reviewed list of European sustainable development indicators (level 1)

Theme	Level 1 indicators
1: Socio-economic development	Growth rate of GDP per inhabitant
2: Sustainable consumption and production	Resource productivity
3: Social inclusion	At-risk-of-poverty rate after social transfers
4: Demographic changes	Employment rate of older workers
5: Public health	Healthy life years and life expectancy at birth
6: Sustainable development	Total greenhouse gas emissions
	Consumption of renewables
7: Sustainable transport	Energy consumption of transport
8: Natural resources	Common bird index
	Fish catches outside safe biological limits
9: Global partnership	Official Development Assistance (ODA)

Source: Eurostat, 2007.(http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-77-07-115/EN/KS-77-07-115-EN.PDF)

2.2 - Composite indices

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

137. 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 stocks, a proxy for human capital, and the costs of CO2 emissions) are evaluated according to national accounts methodology. Each dimension is normalized through linear scaling (nine OECD countries) and aggregation relies on equal weighting. But at this stage the "green" dimension of this index is still secondary.

138. Other examples focus more specifically on the green dimension, such as 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 the environmental systems), human vulnerability (exposure 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 uses 76 variables to cover these 5 domains. There are, for instance, standard indicators for air and water quality (e.g. SO2 and NOx), health parameters (e.g. infant death rate from respiratory diseases), environmental governance (e.g local Agenda 21 initiatives per million people), etc. The EPI is a reduced form of the ESI, based on 16 indicators (outcomes), and is more policy-oriented.

139. The messages derived from this kind of index are ambiguous. The global ranking of countries has some sense, but it is often considered to present an overly optimistic view of

developed countries' contribution to environmental problems. Problems also arise between developed countries. For instance, the index shows a very narrow gap between the United States and France, despite strong differences in terms of their CO_2 emissions. In fact, the index essentially informs us about a mix of current environmental quality, of pressure on resources and of the intensity of environmental policy, but not about whether a country is actually on a sustainable path: 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.

140. On the whole, these composite indicators are better regarded as invitations to look more closely at the various components that underlie them. This kind of function of composite indicators has often been put forward as one of their main *raisons d'être*. But this is not reason enough to retain them as measures of sustainability *stricto sensu* which could secure the same standing as GDP or other accounting concepts. There are two reasons for this. First, as with large dashboards, there is the lack of a well-defined notion of what sustainability means. The second is a general criticism that is frequently addressed at composite indicators, i.e. the arbitrary character of the procedures used to weight their various components. These aggregation procedures are sometimes presented as superior to the monetary aggregations that are used to build most economic indices, because they are not linked to any form of market valuation. Indeed, and we shall come back to 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 indicators, we have little understanding of the arguments for putting one relative value or another 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 indices, and this is one of the strengths of this literature. The problem is rather that their normative implications are seldom made explicit or justified.

2.3 - Adjusted GDPs

141. 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 using elements that standard GDP does not take into account and that matter for sustainability.

142. 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 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 and legal services) and by adding monetary estimates of activities that do contribute positively to welfare (such as leisure and 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 in education by people belonging to the labor force) and health capital, based on a permanent inventory method with a depreciation rate of 20% per year. But they did not in the end include estimates of environmental damage or natural resource depletion.

143. Two strands have developed from this seminal contribution. The first has tried to enrich Nordhaus and Tobin's approach, sometimes deviating increasingly from the criterion of accounting consistency. Examples include the Index of Sustainable Economic Welfare (ISEW) and the Genuine Progress Indicator (GPI). These indicators deduct some evaluations of the costs of water, air and noise pollution from consumption and also try to account for the loss of wetlands, farmland, and primary forests, and for other natural resource depletion, and for CO2 damage and ozone depletion. Natural resources depletion is valued by measuring the investment necessary to generate a perpetual equivalent stream of renewable substitutes.

144. In all countries for which both ISEW and GPI are available, their values are very similar and at some point in time start diverging from GDP. This 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 further improvement in well-being. In other words, according to such indicators, sustainability is already far behind us, and we have already entered a phase of decline.

145. The other strand is more firmly integrated into the realm of national accounting. It is based on the so-called System of Environmental Economic Accounting (SEEA), a satellite account of the Standard National Accounts (SNA). 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 (UNCEEA), created in 2005, is now looking to mainstream environmental-economic accounting, to elevate the SEEA to an international statistical standard by 2010 and to advance SEEA implementation in countries.

146. The SEEA comprises four categories of accounts. The first considers purely physical data related 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 SNA accounting structure The second category of accounts takes those elements of the existing SNA that are relevant to the good management of the environment and makes 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).

147. These first three categories of the SEEA are vital building blocks for any form of sustainability indicator. But what is at stake here is the fourth and last category of SEEA accounts, which deals with 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 resource depletion, those concerning so-called defensive expenditures (protection expenditures being the most emblematic ones), and those relating to environmental degradation.

148. It is these environmental adjustments to existing SNA aggregates that are better known under the rather loose expression of "Green GDP", which is an extension of the concept of net domestic product. Indeed, just as GDP (Gross) is turned into NDP (Net) by accounting for the consumption of fixed capital (depreciation of produced capital), the idea is that it would be meaningful to compute an "ea-NDP" (environmentally-adjusted) that takes into account the consumption of natural capital. The latter would comprise resource depletion (the over-use of environmental assets as inputs to the production process) and environmental degradation (the value of the decline in the quality of a resource, roughly speaking).

149. Green GDP and eaNDP remain, however, the most controversial outcomes of the SEEA, and as such are less implemented by statistical offices, because of the many problems that are raised by these two concepts. Valuing environmental inputs into the economic system is the (relatively) easier step. Since these inputs are incorporated into products that 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. In contrast, as pollution emissions are outputs, there is no direct way to assign a value to them. All the indirect methods of valuation will depend to some extent 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 among many accountants to this practice.

150. But there is a more fundamental problem with green GDP, which also applies to Nordhaus and Tobin's SMEW and to the ISEW/GNI indices. None of these measures characterize sustainability *per se*. Green GDP just charges GDP for the depletion of or damage to environmental resources. 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 we need are 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 - Indicators focusing on overconsumption or underinvestment

151. Under this heading, we group all kinds of indicators that address the issue of sustainability in terms of overconsumption, underinvestment or excessive pressure on resources. Though such indicators tend to be presented in flow terms, they are built upon the assumption that some stocks that are relevant for sustainability correspond to the measured flows, i.e. stocks that are being transmitted to future generations and determine their opportunity sets. As with GDP and other aggregates, trying to perform this task with a single number requires the choice of a metric and an explicit aggregation procedure for these stocks and their variations.

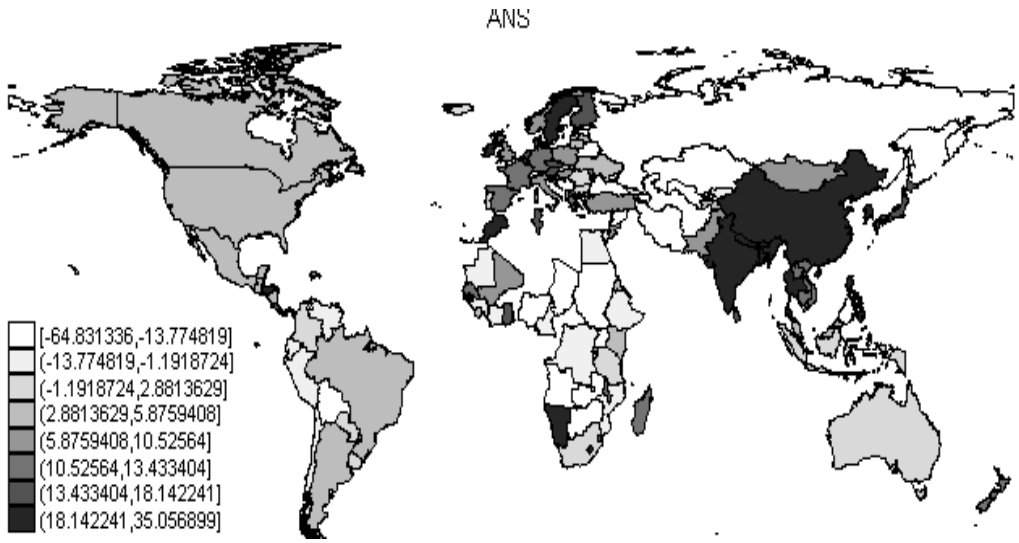
2.4.1. Adjusted net savings (ANS)

152. Adjusted net savings (also known as genuine savings or genuine investment) is a sustainability indicator that builds on the concepts of green national accounts but reformulates these concepts in terms of stock or wealth rather than flows of income or consumption. The theoretical background is the idea that sustainability requires the maintenance of a constant stock of "extended wealth", which is not limited to natural resources but also includes physical, productive capital, as measured in traditional national accounts, and human capital. Net adjusted savings is taken to be the change in this total wealth over a given time period, such as a year. Such a concept clearly appears to be the relevant economic counterpart of the notion of sustainability, in that it includes not only natural resources but also (in principle at least) those other ingredients necessary to provide future generations an opportunity set that is at least as large as what is currently available to living generations.

153. Empirically, adjusted net savings are derived from standard national accounting measures of gross national savings by making four types of adjustment. First, estimates of the capital consumption of produced assets are deducted to obtain net national savings. Second, current expenditures on education are added to net domestic savings as an appropriate value for investment in human capital (in standard national accounting these expenditures are treated as consumption). Finally, 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 cost (including a "normal" return on capital). Finally, global pollution damages from carbon dioxide emissions are deducted[8]. Negative adjusted net savings rates imply that "extended wealth" is in decline, and as such provide a warning of non-sustainability.

154. 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 United States shows that changes over time are almost exclusively driven by gross savings, while the gap in levels between ANS and gross savings is due mostly to capital consumption and human capital accumulation whereas, according to the index, natural capital changes play only a relatively marginal role. Moreover, the ANS figures show that most developed countries are on a sustainable path, while many emerging or developing countries are not. In particular, according to this measure most natural resource-exporting countries are on a non-sustainable path (Figure 3.1).

Figure 3.1. Geographical distribution for Adjusted Net Savings



Source: World Bank, data for 2006.

Reading: Countries are ranked from the most unsustainable (in white) to the most sustainable (in dark). Non-sustainability can be due either to the overextraction of exhaustible resources or to low investment in human and physical capital. The frontiers of countries with missing values are not represented.

8. As for local pollution damages, these 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 damage due to urban air pollution (particulate matter PM10).

155. This kind of approach appeals to to many economists, as it is grounded on an explicit theoretical framework. However, the current methodology underlying empirical calculations has well-known shortcomings: 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 "extended wealth", and on **the price** used to count and aggregate in a context of imperfect or indeed nonexistent valuation by markets – the problem that we already mentioned when discussing the implicit prices used by composite indicators.

156. Indeed, a major shortcoming of ANS estimates is that the adjustment for environmental degradation is only limited to a restricted set of pollutants, the most significant one being carbon dioxide emissions. The authors acknowledge that the calculations do not include other important sources of environmental degradation, such as underground water depletion, unsustainable fisheries, and soil degradation, and *a fortiori* biodiversity loss.

157. 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 warranted only in a context of perfect markets, which is clearly not the case in reality, and especially not for natural resources, where externalities and uncertainties are paramount. Further, 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 and this has very strongly reduced the practical relevance of the ANS for concerned countries.

158. As for pricing environmental degradation, 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, we must evaluate so-called "accounting prices" by modeling the long-term consequences of given changes in environmental capital and how they impact future well-being. But practical implementation raises considerable problems. Under the current state of the art, the prices used to value carbon emissions in existing estimates of 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 policy.

159. Finally, by computing ANS per country we miss the global nature of sustainability. Indeed, one may feel uneasy when faced with the message conveyed by ANS about resource-exporting countries (e.g. oil). In these countries, from the ANS perspective, non-sustainability stems from an insufficient rate of reinvestment of the income generated by the exploitation of the natural resource: "over-consumption" by importing countries is not an issue at all. Developed countries, which are generally less endowed with natural resources but richer in human and physical capital than developing ones, would then appear unduly sustainable. As a consequence, some authors have argued in favor of imputing the consumption of exhaustible resources to their final consumers, i.e. the importing countries.If scarcities were fully reflected in the prices at which exhaustible resources are sold on international markets, it is true that there would be no reason for making such a correction. However, when prices are non-competitive, the importing country pays less for its imports than would be required; it will have a responsibility in global non-sustainability that is not captured by the money-value of its imports. Low prices allow such countries to over-consume and to transfer the long-term costs of this over-consumption to the exporting countries.

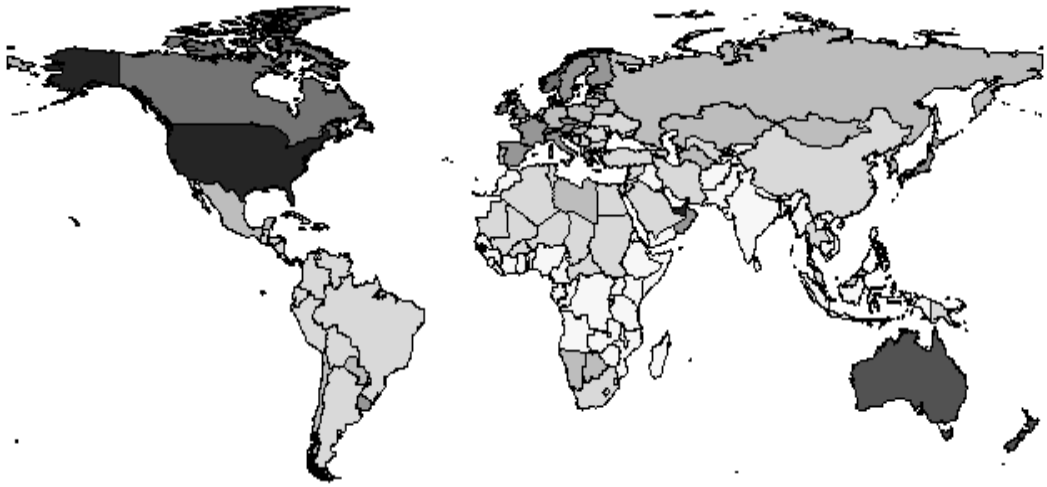
2.4.2. Footprints

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

161. The Ecological Footprint (hereafter EF) measures how much of the regenerative capacity of the biosphere is used up 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. A country’s Footprint (demand side) is the total area required to produce the food, fiber and timber that it consumes, absorb the waste that it generates, and provide space for its infrastructure (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 humankind.

162. The results are well-known and rather striking: since the mid-1980s, humanity’s footprint has been larger than the planet's carrying capacity, and in 2003 humanity’s total Footprint exceeded the Earth’s biocapacity by approximately 25 per cent. 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-capacity of those two geographical zones (Figure 3.2).

Figure 3.2 : Ecological footprint by country



Source : Global Footprint Network, data for year 2005

Reading : Dark areas correspond to countries with the highest values for the ecological footprint, i.e. with the highest contributions to worldwide unsustainability. Countries with missing values are not represented

163. 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 that different forms of natural capital are substitutable and that different natural capital goods are additive in terms of land area, but strongly stands against weak sustainability assumptions. In fact, this indicator gives no role to savings and capital accumulation: any positive ecological surplus (biocapacity that exceeds the ecological footprint) does not entail an increase in some natural capital stock, and hence an improvement in future productive capacity. A fortiori, saving and accumulating manufactured or human capital does not help sustainability. On the other hand, one must observe that the indicator ignores the threat to sustainability resulting from the depletion of non-renewable resources (e.g. oil): the consequences for sustainability are treated only from the waste assimilation (implied CO2 emissions) point of view rather than from an analysis based on depletion dynamics.

164. The results are also problematic for measuring a country's own sustainability, because of the substantial anti-trade bias inherent in the Ecological Footprint methodology. The fact that densely populated (low biocapacity) countries like the Netherlands have ecological deficits, whilst sparsely populated (high biocapacity) countries like Finland enjoy surpluses can be seen as part of a normal situation where trade is mutually beneficial, rather than an indicator of non-sustainability. Indeed, recent reearch has tended to move away from comparing a country's EF with its *own* biocapacity, and to propose instead to divide all countries' EFs by *global* biocapacity. By doing this, one is acknowledging that EFs are not measures of a country's own sustainability but of its contribution to global non-sustainability.

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

166. As a result, less-encompassing but more-rigorously-defined footprints, such as the "Carbon Footprint" (CF), would seem better-suited, insofar as they are more clearly physical measures of stocks that do not rely on specific assumptions about productivity or an equivalence factor. As far as communications is concerned, such an indicator is just as capable of sending strong messages in terms of the over-utilization of the planet's capacity for absorption. The CF also has the interesting feature of being computable at any level of disaggregation. This makes it a powerful instrument for monitoring the behavior of individual actors.

3 - Quantifying sustainability in a consensual way: what are the main stumbling blocks?

167. Let's summarize the main messages so far. The previous section has shown the large number of existing attempts to quantify sustainability. This abundance of measures is a serious drawback insofar as different synthetic indicators convey widely divergent messages. This leads to a great deal of confusion among statisticians and policymakers. It urges a return to the fundamental questions: What do we want to measure exactly? What are the real obstacles to doing so with a single headline measure?

3.1 - What do we want to measure?

168. Since the Brundtland Report, the notion of sustainable development has expanded to become an all-encompassing concept that absorbs every dimension 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 "sustainable" component of "sustainable 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 whether the continuation of present trends does or does not allow it to be maintained.

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

- The extensive dashboards of sustainable development reviewed in section 2.1 effectively conflate the measurement of current well-being and the measurement of its sustainability. This is not to say that dashboards are of no use. Quite the contrary: our final conclusion will be that a unidimensional view of sustainability certainly remains out of reach. But we do want to end up 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.

- Composite indicators raise similar problems, with the additional complication that the way in which various items are weighted is arbitrary, with consequences that are seldom made explicit.

- Measures of a **sustainable standard of living**, such as the Green GDP, are also insufficient for assessing sustainability. 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? What conclusion would we draw from the fact that a given country's green GDP is x% or y% of its GDP defined in standard terms? Does this necessarily imply that this country is on an unsustainable path?

170. In fact, green GDP focuses on only one side of the problem, i.e. the measurement of what can be consumed every year without environmental impoverishment. This does not tell us whether we are on a sustainable path. If we want to measure sustainability, what is required is a **comparison** between this concept of genuine production and current consumption. All this makes the appropriate sustainability index more akin to a concept of net investment or disinvestment, and this is precisely the route that extended wealth or ANS exemplifies, but which is also implicitly followed by footprint indicators that are more specifically focused on the renewal or depletion 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 our passing them sufficient amounts of all the assets that matter for well-being. If we denote by "W" the "extended wealth" index 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".

171. In our view, such a formulation of the sustainability issue has the potential to provide the common language necessary for constructive debates between people from very different perspectives. To take just one example, it fully answers 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 future well-being. This outcome can be avoided only if some action is taken to repair the damage, with these actions being counted as positive investment.

3.2 - Summarizing sustainability in one number: is it realistic?

172. Now, we have seen that both ANS and footprint evaluations are subject to many objections and can be considered, at best, as proxies of what would be genuine indices of changes in extended wealth or its components. Returning to fundamentals means asking precisely what would be required to measure the above-mentioned dW indices 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 be weighted against each other?

173. It is clearly this last question that is more problematic and tends to crystallize opposition between the proponents of monetary indicators and physical indicators. Is there actually some reasonable prospect of evaluating everything in money units, or should we accept that this is possible only up to a certain point?

174. If all assets were traded on perfect markets by perfectly forward-looking agents fully taking into account the welfare of future generations, one could argue 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 it is unlikely that current prices fully reflect this future-oriented dimension, due to market imperfections, myopia and uncertainty. This implies that a true measure of sustainability requires a dW index in which assets are valued not at market prices, but rather using imputed "accounting prices" based on some objective physical or economic modeling of how future damage to the environment will affect well-being, just 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.

175. Recent research has clarified the requisites of such an exercise. One is a full set of economic and physical projection of how initial conditions determine the future joint path of economic, social and environmental variables. Another 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.

176. Equipped with such instruments, it should be possible to derive sustainability indices that have the properties that one would expect, i.e. a capacity to anticipate future declines in well-being below its current level. Some simulations proposed in the technical report illustrate certain aspects of this capacity. First of all, this sustainability index is the best suited

for sending correct forewarnings to countries that are on unsustainable paths because of an insufficient rate of accumulation or of renewal of their produced capital, be it human or physical. And this is of course an important property: even if environmental issues are of considerable importance, we cannot ignore these other dimensions of sustainability.

177. Second, such an indicator is inconsistent with the "strong" view of non-sustainability (i.e. problems arising from the depreciation of environmental assets that are essential to human well-being or even survival) only when it relies on fixed price levels for natural and non-natural assets. But if we were able to derive this index from a physical-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 increases in the relative accounting or "imputed" prices of these critical natural assets.

178. But 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. It can also be used as a tool for emphasizing the many obstacles to the building of a comprehensive index and the need for more pragmatic second-best solutions.

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

179. Measuring sustainability with a single dW index can work only under two strong assumptions: one is that future eco-environmental developments can be predicted perfectly, and the second is that there is perfect knowledge about 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 ignorance and uncertainty about future interactions between the two spheres, and by a lack of consensus about the very definition of the objective function.

180. Let's briefly develop the first point. The future is fundamentally uncertain. Uncertainty takes many forms, some of them amenable to probability computation, while many others are much more radical. This affects not only the parameters of any models that one may try to use to project eco-environmental interactions, but also the structure of the models themselves, the measurement of current stocks, and even the list of the natural assets for which current and future stocks need to be taken into account. Most of the debate concerning long-term environmental change reflects different beliefs about future eco-environmental scenarios. There is no reason why sustainability measurement should escape such difficulties.

181. Some solutions might be considered for this problem. One is to do what all prospectivists do when they want to emphasize the uncertain nature of future trends, i.e. work with scenarios or provide confidence intervals. One could also consider submitting indices to some forms of "stress tests", i.e. recompute them under assumptions of external shocks on asset values. This could include sudden upward adjustments in the value of environmental assets, but also drastic reductions in the value of some other items – such as produced capital and human capital. Such modes of presentation could be explored and eventually adopted.

182. But this could still be insufficient or difficult to present in a convenient way. Questions such as climate change require a specific consideration which drives us back to the distinction between weak and strong sustainability. The point is not that aggregate indices are by nature unable to account for situations of strong non-sustainability. 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 to quantify precisely what these extreme valuations should be. In such cases, and *a fortiori* for items for which we do not even have a single guestimate of a monetary value, a separate physical accounting is unavoidable.

183. The problem then is to present such an index in a compelling way. Monetary indices have the advantage of using units that speak to everyone. In addition, they can be related 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, a tonnage of CO2 emissions is not a very informative number if we do not have some reference for how many tons can be emitted each year without severe consequences for the climate. Other physical indicators have been advocated by climate specialists, including "CO2 radiative forcing", measuring the effect of CO2 on the Earth's energy imbalance, and measuring the regression of permanent ice. But it is difficult for non-experts to take such indicators on board. It is essential to find more suggestive ways to highlight such figures if we want the indicator to have an impact on the debate. One of the major successes of the ecological footprint has been its ability to express pressure on the environment in an easily understandable unit. The EF indicator has limits that make it problematic to many observers. But, given the objective of limiting climate change, the general idea of using the footprint as a generic unit for the different forms of pressure that mankind exerts on Earth's regenerative capacity is an option. A metric like this is used, for instance, with the more focused concept of the Carbon Footprint or the kindred concept of the CO2 budget.

3.4 - Uncertainty is also normative

184. In addition to raising technological issues, measuring sustainability with a single index number would confront us with severe normative questions. The point is that there can be as many indices of sustainability as there are normative definitions of **what** we want to sustain. In standard national accounting practice, the normative issue of defining preferences is generally avoided through the assumption that observed prices reveal the true preferences of people. No explicit normative choice is therefore to be made by the statistician. But as soon as we recognise that market prices cannot be trusted, alternative imputed prices must be computed, whose values will strongly depend upon normative choices.

185. Can we solve this normative problem? One could attempt to solve it empirically by 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 indices of subjective well-being. But can the contingent evaluations and subjective measures established today in our specific eco-environmental setting be used to predict the valuations of future generations in eco-environmental settings that may have become very different? It could be argued 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 this requires that we immediately place a high value on these items just because we think that our descendants may wish to do so.

186. Another example of these normative issues is the question of determining how sustainability indices 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 sustainability indicators should be adapted to such an objective function. This would be in line with one of the other aspects of the Brundtland definition of sustainability that is often overlooked, i.e. its concern for the distribution of resources *within* as well as *between* generations. In a world where inequalities within countries naturally tend to increase, messages concerning sustainability will differ depending on the goal that we set ourselves. 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 offer efficient help in protecting this population from poverty. In principle, the theoretical framework based on extended wealth tells us how we could ideally put some value on this kind of "institutional" investment. But, needless to say, the prospect of actually being able to do this is still more remote than for other assets.

3.5 - An additional source of complexity: the global dimension

187. A global context poses additional problems for sustainability indicators. Advocates of the ANS argue that sustainability problems generally concentrate in poor resource-exporting countries even if it is in developed countries that the 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 the cost of their imports, the developed countries can still maintain 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 also want to be on a sustainable path.

188. Yet this logic holds true only under the assumption of efficient markets. If markets are not efficient and if the natural resource is underpriced, then importing countries benefit from an implicit subsidy while the exporting ones are effectively taxed. This means that the actual sustainability of developed countries is overestimated, while that of the developing countries 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.

189. To illustrate this issue, 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. 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 the resource. Let's push the asymmetry further by assuming that it is only country 2 that is affected by the degradation of the environmental good. Country 1 is completely indifferent to the level of degradation of this environmental good, for instance because its geographical characteristics fully protect it from the consequences.

190. In 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 is to compute 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, the accounting price for the polluter will be zero, because we have assumed that it is not impacted at all by environmental changes, which implies that it

attributes no value at all to the environmental asset. On the other hand, the polluted country will attribute a positive value to the asset. The message conveyed using this extended wealth concept is that the polluter is on a sustainable path, while the polluted is not.

191. From a certain point of view, it is correct to say that the polluter is not confronted by the prospect of a decline in well-being, in contrast to the polluted. But from another viewpoint, the message is clearly misleading. There is nothing the polluted can do to restore its sustainability. It is only a change in the polluter's technology that could help restore the polluted country's sustainability. We are in need of indices that would convey such a message. The popularity of footprint indicators stems precisely from the fact that, whatever their other limitations, they are able to send such messages to policy makers and public opinion. This is one more argument in favor of an eclectic approach that mixes points of views. An approach centred on national sustainabilities may be relevant for some dimensions of sustainability, but not for others. Global warming is a typical example of the latter case, as the prospective consequences of climate change are distributed very unevenly, without necessarily correlating with a country's CO2 emissions.

4 - Conclusion

192. To sum up, what have we learnt, and what can we conclude? This trip through the world of sustainability indicators has been a bit lengthy, and we have not been able to avoid technicalities completely. A wide variety of indicators are already available and we have analyzed the reasons why a comprehensive assessment of sustainability is difficult to establish in a fully consensual way. Assessing sustainability requires many assumptions and normative choices, and it is further complicated by the existence of interactions between the socio-economic and environmental models followed by the different nations. The issue is indeed complex, more complex than the already complicated issue of measuring current well-being or performance. But we shall nevertheless try to articulate a limited set of recommendations, which we shall also try to keep as pragmatic as possible.

Recommendation 1: Sustainability assessment requires a well-identified sub-dashboard of the global dashboard to be recommended by the Commission.

193. The question of sustainability is complementary to the question of current well-being or economic performance, and must be examined separately. This recommendation to separate the two issues might look trivial. Yet it deserves emphasis, because some approaches fail to adopt this principle, leading to confusing messages. The confusion reaches a peak when one tries to combine these two dimensions into a single indicator. This criticism applies not only to composite indices, but also to the notion of green GDP. To take an analogy, when driving a car, a meter that weighed up in one single value the current speed of the vehicle and the remaining level of gasoline would not be of any help to the driver. Both pieces of information are critical and need to be displayed in distinct, clearly visible areas of the dashboard.

Recommendation 2: The distinctive feature of all components of this sub-dashboard should be to inform about variations of those "stocks" that underpin human well-being.

194. In order to measure sustainability, what we need are indicators that tell us the sign of the change in the quantities of the different factors that matter for future well-being. Putting the sustainability issue in these terms compels recognition that sustainability requires the

simultaneous preservation or increase in several "stocks": quantities and qualities not only of natural resources but also of human, social and physical capital. Any approach that focuses on only a part of these items does not offer a comprehensive view of sustainability.

195. Speaking in such terms also avoids many of the misconceptions about the messages sent by traditional national accounts indicators. For instance, a frequent criticism of GDP is that it classifies ecological catastrophes as blessings for the economy, because of the additional economic activity generated by repairs. The stock approach to sustainability clearly avoids this ambiguity. Catastrophes will be recorded as a form of depreciation of natural or physical capital. Any resulting increase in economic activity would have a positive value only insofar as it helps to restore the initial level of the capital stock.

Recommendation 3: A monetary index of sustainability has its place in such a dashboard, but under the current state of the art, it should remain essentially focused on economic aspects of sustainability.

196. The stock approach to sustainability can in turn be broken down into two versions. One version would just look at variations in each stock separately with a view to doing whatever is necessary to keep it from declining or at least to keep it above some critical threshold beyond which further reductions would be highly detrimental to future well-being. Or one could attempt to summarize all stock variations in synthetic figures.

197. This second track is the one followed by so-called extended wealth or adjusted savings approaches, which share the idea of converting all these assets into a monetary equivalent. We have discussed the potential of such an approach, but also its limitations. In certain conditions, it allows to anticipate many forms of non-sustainability, but the requirements for such a capacity are extremely high. This is because the aggregation required by this approach cannot be based on market values: market prices are nonexistent for quite a large number of the assets that matter for future well-being. Even when they are available, there is no guarantee that they adequately reflect how these different assets will matter for future well-being. In the absence of such price messages, we have to turn to imputations, which raises both normative and informational difficulties.

198. All this suggests staying with a more modest approach, i.e. focusing the monetary aggregation on items for which reasonable valuation techniques exist, such as physical capital, human capital and natural resources that are traded in markets. This more or less corresponds to the hard part of "adjusted net savings" as computed by the World Bank and further developed by several authors. "Greening" this index more intensively is of course a relevant objective, and we can keep it on the agenda but we know that the analytical apparatus for doing so is a complex one: large-scale projection models of interactions between the environment and the economy, projecting changes in the relative scarcities of corresponding assets and their impact on relative accounting prices, and allowing also a proper treatment of uncertainties or potential irreversibilities that affect these interactions. Meanwhile, we must focus this indicator essentially on what it does relatively well, i.e. the assessment of the "economic" component of sustainability, that is, the assessment of whether or not countries overconsume their economic wealth.

Recommendation 4: The environmental aspects of sustainability deserve a separate follow-up based on a well-chosen set of physical indicators.

199. As far as environmental sustainability is concerned, the limitations of monetary approaches do not mean that efforts to monetize damages to the environment are no longer needed: it is well known that fully opposing any kind of monetization often leads to policies that act as if environmental goods had no value at all. The point is that we are far from being able to construct monetary values for environmental goods that at the macro level can be reasonably compared to market prices of other capital assets. Given our state of ignorance, the precautionary principle legitimates a separate follow-up of these environmental goods.

200. Another reason for a separate treatment is that these environmental issues often pertain to global public goods, such as the case of the climate. In such cases, the problem with the standard extended wealth approach is that it essentially focuses on country-specific sustainabilities. With global public goods, what is involved is more the contributions by the different countries to global unsustainabilities.

201. The ecological footprint could have been an option for this kind of follow-up. In particular, in contrast to net adjusted savings, it essentially focuses on contributions to global non-sustainability, with the message that the main responsibility lies with the developed countries. Yet the group has taken note of its limitations, and in particular that it is far from being a pure physical indicator of pressure on the environment: it retains some aggregation rules that may be problematic. In fact, much of the information that it conveys about national contributions to non-sustainability is imbedded in a simpler indicator, the carbon footprint, which is therefore one good candidate for monitoring humanity's pressure on the climate, among many indicators proposed by climatologists that are shortly reviewed in the technical report.

202. For other aspects of environmental sustainability, such as air quality, water quality, biodiversity and so on, one can again borrow from these large eclectic dashboards. Just to note a few of the indicators already incorporated in such dashboards, we could mention smog-forming pollutant emissions, nutrient loading to water bodies, the abundance of key specified natural species, rates of conversion of natural habitats to other uses, the proportion of fish catches beyond safe biological limits, and many others. Today, at this stage of the debate, economists do not have any particular qualification for suggesting what the right choices are. This is why we will not propose any closed list of these indicators here.

203. In short, our pragmatic compromise is to suggest a small dashboard, firmly rooted in the logic of the "stock" approach to sustainability, which would combine:

– An indicator more or less derived from the extended wealth approach, "greened" as far as possible on the basis of currently available knowledge, but whose main function, however, would be to send warning messages concerning "economic" non-sustainability. This economic non-sustainability could be due to low savings or low investment in education, or to insufficient reinvestment of income generated by the extraction of fossil resources (for countries that strongly rely on this source of income).

– A set of well-chosen *physical* indicators, which would focus on dimensions of environmental sustainability that are either already important or could become so in the future, and that remain difficult to capture in monetary terms.

204. This scenario has several points of convergence with conclusions reached by other reports recently devoted to the topic, such as the recent OECD/Eurostat/UNECE report on sustainability measurement, whose conclusions were released in 2008, or the more recent report by the French Economic, Social and Environmental Council released in 2009. The first one, in particular, strongly advocates the stock-based approach to sustainability and proposes a small dashboard clearly separating assets that can be monetarized in a reasonable way and other assets for which a separate physical measures are necessary. The second one warns againts limits of the Ecological Footprint and, as far as climatic change is concerned, argues in favor of the Carbon Footprint index. Such points of convergence are reassuring: they suggest that, from the relatively confused situation that was described in section 2, we are steadily moving towards a more consensual framework for the understanding of sustainability issues (See Box 3)[9].

Box 3 : Physical and other non-monetary indicators : which ones to choose ?

The commission's general position has been to avoid formulating definitive turnkey proposals on any the different issues it has raised. All proposals rather intend to stimulate further debate. This is all the more true in the domain of *physical* sustainability indicators where the expertise of specialists from other disciplines is crucial and was only indirectly represented in the commission's composition.

Some suggestions can however be made, in connection with conclusions of some recent related reports.

In 2008, a OECD/UNECE/Eurostat working group has produced a report on measuring sustainable development whose messages have several points in common with ours. It strongly advocates the stock-based approach to sustainability as the relevant way of structuring a micro-dashboard of sustainability indicators gathering both stock and flow variables. It also suggests a line of demarcation between determinants of 'economic' well-being (those that are the most directly amenable to monetary evaluation) and the determinants of 'foundational' well-being, among which four couples of stock/flow environmental indicators devoted respectively to global warming, other forms of atmospheric pollution, quality of water and biodiversity. The details and positions of these indicators in the dashboard can be visualized (in bold) on the table below.

9. Some other points of convergence can be found in reactions by the European Environmental Agency to the first draft of this commission's summary report.

Small set of sustainable development indicators proposed by the UNECE/OECD/Eurostat working group on sustainability measurement

Indicator domain	Stock indicator	Flow indicator
Foundational well-being	Health-adjusted life expectancy	Index of changes in agespecific mortality and morbidity (place holder)
	Percentage of population with post-secondary education	Enrolment in post-secondary education
	Temperature deviations from Normals	Greenhouse gas emissions
	Ground-level ozone and fine particulate concentrations	Smog-forming pollutant emissions
	Quality-adjusted water availability	Nutrient loadings to water bodies
	Fragmentation of natural habitats	Conversion of natural habitats to other uses
Economic well-being	Real *per capita* net foreign financial asset holdings	Real *per capita* investment in foreign financial assets
	Real *per capita* produced capital	Real *per capita* net investment in produced capital
	Real *per capita* human capital	Real *per capita* net investment in human capital
	Real *per capita* natural capita	Real *per capita* net depletion of natural capital
	Reserves of energy resources	Depletion of energy resources
	Reserves of mineral resources	Depletion of mineral resources
	Timber resource stocks	Depletion of timber resources
	Marine resource stocks	Depletion of marine resources

Source : UNECE/OECD/Eurostat (2008)

More recently, the French Economic, Social and environmental Council (CESE) has produced a report whose initial aim was the assessment of the Ecological Footprint but that has more widely explored the different tracks available for quantifying sustainability. It has the same messages as the current report concerning the limits of this EF index, and the fact that most of the relevant information that it conveys is more directly and more neatly reflected in one of its subcomponents, the carbon footprint. As a consequence, it strongly advocates in favor of this index. Compared to Global GHG emissions suggested in the OECD/UNECE/Eurostat Dashboard presented above, the Carbon Footprint has the advantage of being expressed in this 'footprint' unit that is intuitively so appealing and that has made the success of the EF. In addition to this, this CESE report has suggested emphasizing the other physical indicators already present in large international dashboards such as the one elaborated for the European Union Strategy for Sustainable Development. Some of them are those already quoted in the OECD/UNECE/Eurostat dashboard.

As far as climate change is concerned, some other indicators can be considered. Direct observation of mean temperature is one possibility but not the best suited, because it has a tendency to run behind the main components of climate change and because there can always be disagreements about the causes of temperature rises, hence about their permanent or transient character. Consequently, climatologists prefer to make use of a thermodynamic concept, the CO_2 radiative forcing, that measures the earth energy imbalance created by the action of CO_2 as a greenhouse gas.

Alternatively, it is possible to directly use a notion of CO_2 remaining budget: according to climatologists, there is an upper limit of 0.75 trillion tonnes of carbon that might be discharged in the atsmosphere if the risk of temperatures exceeding 2° Celsius above pre-industrial levels is limited to one-in-four, this upperbound at 2°C being largely accepted among climate experts as a "tipping point" opening the door to unstoppable feedback effects (methane from melting permafrost , CO_2 and methane from decaying tropical forests ,all sorts of greenhouse gases released by saturated warming oceans, etc...). Of this 0.75 total budget, emissions to 2008 have already consumed circa 0.5. Hence the importance of monitoring this remaining CO_2 budget. The attractiveness of this indicator is to be strongly consistent with the stock-based approach to sustainability. It can be also rephrased in the very expressive terms of a countdown index, i.e. the time that remains until exhaustion of this stock, under the assumption of emissions remaining on their current trend. This kind of representation is often used for other forms of exhaustible resources.

Still other indirect indicators of global warming are the regression of permanent ice or the oceanic pH. The regression of permanent ice has the advantage of being an advancedone and to be directly related to manifest effects. The oceanic pH increases with the amount of CO_2 that is naturally pumped into the oceans. A consequence of this increase is a decrease in the quantity of phytoplancton, which is itself a carbon sink no less important than the forests. One may therefore say that the physical sink (sea water dissolving atmospheric CO_2) destroys the biological one. This is the reason why the oceanic pH appears to be another good tentative indicator of climate change, pointing to one of the most vicious feedback effects. Among criteria for chosing between all the indicators, two are of particular importance. One is their appropriability by the public, the other is the capacity of declining them at national or even subnational levels: in this respect, the carbon footprint has quite a lot of advantages.

As far as biodiversity is concerned, the issue is currently under review by the TEEB group ("the economics of the environment and biodiversity") working at the initiative of the European Union and it has been also recently adressed by a report by the French *Conseil d'Analyse Stratégique*, in this case with the idea of pushing as far as possible the monetization of this dimension. The reason for this search of monetary equivalent is essentially that it may foster incorporation of this dimension in investment choices : many public decisions such as building a new motorway imply some potential biodiversity loss through fragmentation of natural habitats. But the report provides also a very detailed and technical review of available physical measures of biodiversity, to which the reader is referred for further information.

At last, moving away from environmental preoccupations, but still on the "non-monetary" side, one important issue is the issue of social capital and "institutional assets" that we transmit to future generations. One will have noticed that the UNECE/OECD/Eurostat dashboard presented above did not propose any indicator of this kind, not because the question is not relevant, but mainly because of lack of consensus about the way to measure it. Subgroup 3 was not in a position to explore this question further, but efforts along this direction remain undoubtedly necessary.

205. A subsidiary question concerns a user's guide to such a dashboard. A warning should be given that no limited set of figures can pretend to forecast the sustainable or unsustainable character of a highly complex system with certainty. The purpose is, rather, to have a set of indicators that give an "alert" to situations that pose a high risk of non-sustainability. Whatever we do, however, dashboards and indices are only one part of the story. Most of the efforts involved in assessing sustainability focus on increasing our knowledge about how the economy and the environment interact now and are likely to interact in the future.

II. Substantial Arguments presented in the Report

CHAPTER 1 : GDP RELATED ISSUES

CHAPTER 2 : QUALITY OF LIFE

CHAPTER 3 : SUSTAINABLE DEVELOPMENT AND ENVIRONMENT

CHAPTER 1 - GDP RELATED ISSUES

1 - Introduction

1. 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 (products that are either exchanged through market transactions or produced with inputs purchased on the market), and thus more geared to measure the aggregate supply side of economies than the living standards of its citizens. Although GDP levels are correlated with many indicators of living standards, the correlation is not universal and tends to become weaker when particular sectors of the economy are concerned. For example, real household income – an income measure which is more closely related to living standards – has evolved quite differently from GDP growth in a number of OECD countries. Too much emphasis on GDP as the unique benchmark can lead to misleading indications about how well-off people are and run the risk of leading to the wrong policy decisions. The purpose of this chapter is to go beyond GDP in our quest for better economic measures of living standards. At the same time we shall be looking for indicators that remain within the broad boundaries of a national accounting framework.

2. In the quest for such indicators, we shall encounter a tension between the conceptual soundness of some measures and the ease by which they can be implemented and communicated to users of the data. For example, conceptual considerations might lead us towards inclusive measures of income, taking into account services that households 'deliver to themselves' such as child care, cooking or parents' education services to children. At the same time, such variables are hard to quantify and statisticians have to make more or less reliable imputations. Too many assumptions risk undermining the usefulness of an index. But ignoring these elements may also impair the relevance of these measures.

3. We accommodate the tension between inclusiveness and measurability in several ways. First, the chapter is structured in terms of increasing inclusiveness, starting out with measures based on a relatively narrow market boundary and moving towards broader indicators of living standards. It will be shown that even within the narrow market boundary there are measures that do a better job than GDP in tracking how well off citizens are. We shall also point out where measurement can be improved through additional research and empirical investigation, thus helping to reduce the trade-offs mentioned earlier.

4. Second, we take a pragmatic stance. The simplest approach consists of complementing existing indicators with additional indicators covering neglected dimensions of well-being, without necessarily attempting to rigorously aggregate all these dimensions into a single summary index of living standards. For example, the monetary valuation of non-market activity of households is fraught with uncertainty. But useful insights can be gained by analysing (relatively reliable) data on how people use their time. Such time use data can show how household activities change over time, how they differ between countries and how women's use of time is different from men's. More comprehensive, non-monetary indicators

of the quality of life complement the picture and these are dealt with in a different part of the present report.

5. Third, some issues are cross-cutting and will be taken up in several places of the text. In particular, the question of income (and wealth) distribution is equally pertinent from a perspective that focuses on market income and one that takes a more comprehensive approach.

6. A point of terminology: the present chapter focuses on material standards of living which are based on measures of income, consumption or wealth, typically expressed in monetary terms. These are best understood as one of the determinants of overall well-being of people or of their "human capabilities".[1]

2 - The background: GDP, prices and living standards

2.1. The role of market prices

7. The starting point for measuring economic performance is market transactions. A fundamental merit of market transactions is that they provide 'objective' prices that serve to value quantities of goods and services, making it possible to add up 'apples' and 'oranges'. Macro-economic and fiscal policies require monitoring of market transactions. Indeed, current metrics originated out of an attempt to monitor market production, following the development of Keynesian economics[2].

8. When markets are competitive and in the absence of externalities, relative prices of goods and services mirror the relative values that individuals puts on these commodities. So, in principle, weighting products with their prices implies weighting them with their value for each individual in society. In addition, economic theory (Weitzman 1976) tells us that, in a world where all transactions take place in competitive markets and where economic well-being depends only on consumption of marketed goods, changes in net domestic product (NDP, i.e. GDP adjusted for depreciation) are a good gauge of changes in economic well-being: this holds because an individual's or a country's 'wealth' can be viewed as the present discounted value of consumption. Under these conditions, NDP turns out to be like an interest payment (a 'return') on this wealth. This establishes – albeit under restrictive conditions – a direct link between NDP and economic well-being. It also constitutes the basic framework for considerations about sustainability.

9. In reality, prices may not exist for some goods and services and, even where they exist, they may deviate from society's underlying valuation. In particular, in the presence of externalities, GDP, and indeed pure market-based measures, will not track well-being. Environmental damage caused by production or consumption activities that is not reflected in market prices is a well-known example for a negative externality.

1. The latter are captured, for example, in the work by Krueger et al. (2008) or by Sen (1985).

2. Keynes (1940) undertook national income calculations, using earlier work by Colin Clarke, with a view to analysing the effects of wartime expenditure and inflation. Meade and Stone (1941) present a more complete version of income, expenditure, savings and investment accounts.

10. While it is straightforward to talk about 'prices' and 'quantities' in theory, defining and measuring prices and quantities in practice is altogether a different matter. Two variables are observed by statisticians: the value of transactions (i.e. the product of prices and quantities); and the prices for different types of products, which are used to construct price indices. In many instances, these price indices are used to 'deflate' values and to obtain a measure of 'volume' or 'quantity'. However, constructing price indices is only possible if the products whose prices are observed do not change over time – otherwise like would not be compared with like. As it happens, many products change over time – they disappear entirely or new features are added to them. In other words, there is ***quality change*** and it 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.

11. Capturing quality change correctly is a tremendous challenge for statisticians and yet it is vital to measure 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 under-estimating real income. For instance, 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%. Since then, the U.S Bureau of Labor Statistics – the agency responsible for the CPI – has followed up with a series of measures to address the issues raised by the Boskin Commission. A more recent report by the *Panel on Conceptual, Measurement, and other Statistical Issues in Developing Cost-of-Living Indexes* (Schultze and Mackie 2002) devotes significant space to the quality change and new goods problem in the U.S. consumer price index.

12. The debate in Europe has tended to go the opposite way, in particular following the introduction of the Euro: official measures of inflation (Italy, France, Belgium, Germany and others) have been criticized for under-estimating inflation, thereby painting too rosy a picture of the real income situation of citizens. Capturing quality change in prices and volumes is a recurring issue (see the discussion on government-provided services in section 2.1) without a single methodological solution.[3]

13. A more subtle issue arises in conjunction with market power. When price discrimination by firms with market power increases, these firms' profits will rise. What is not measured by conventional economic statistics is the loss in consumer surplus that households face. The effect of imperfect competition on the reliability of GDP to gauge real living standards of consumers is difficult to determine and can only be assessed in a general equilibrium setting, where both changes in volumes and substitution effects by consumers are taken into account. Similar conceptual issues are posed by profits that are generated by abusive practices, e.g. exploiting consumer ignorance. Most likely, the loss of consumer well-being exceeds the profit gain of firms; while the latter are captured in our GDP measure, the former are not.

14. The proportion of goods and services that come along with externalities may change over time and differ across countries. Thus, in temporal and spatial comparisons, price signals have to be interpreted with care. For a number of purposes, they do not always provide a useful vehicle for aggregation of quantities: a case in point is the use of market prices in

3. For an authoritative overview of quality-adjustment techniques see Triplett (2006).

calculations of sustainability indicators and more is said about this in another chapter of this report.

2.2. The role of imputations

15. While market prices and observed market transactions form the centre piece of measures of economic activity, GDP does not stop at those and, indeed, the *System of National Accounts* includes a range of income, expenditure and production items that do not reflect market transactions. These non-market transactions are generally based on imputations, as leaving them out would provide a biased picture of economic activity and distort comparisons across countries and over time. Some of these imputations bring GDP and income measures closer to an economically meaningful gauge of living standards.

16. Imputations come, however, at a price. Imputations allow attributing a value to production, income or consumption even when there is no or no explicit economic transaction. The main reason[4] for imputing values is comprehensiveness of measures. Some of these imputations (such as imputed rents, see below) have been incorporated into GDP. This is not the case for depreciation, for instance, an imputation to account for the loss of value of capital used in production and needed to move from GDP to Net Domestic Product (NDP). Some imputed transactions take place within the market sphere but are implicit and implicitly-priced, such as financial intermediation services (FISIM – see below). Other transactions are explicit and the imputation consists in attributing them to somewhere else in the economy (individual consumption of goods and services provided by government - see below). Leaving out these imputations would mean accepting a very incomplete picture of economic activity.[5]

17. Another consequence of imputations is that they help preserving the "invariance principle" for national accounts. This suggests that the value of main accounting aggregates should not depend on the institutional arrangements in a country. For example, if the same medical services can be provided by either the public or the private sector, overall production should be unaffected by a switch between the two institutional settings. The invariance principle increases comparability, over time and between countries.

18. In the present system of national accounts, the most important imputations are:

- ***Individual consumption of goods and services provided by government***: these goods and services when provided for free (essentially health and education services) are not part of households' consumption expenditures because they are not directly paid by households. Of course, households pay for these services in the form of taxes or social

4. Note that the national accounts also rely on a number of statistical imputations. For example, enterprise surveys are rarely complete and imputations must be made for units not covered in the sample. Or imputations have to be made to deal with non-response by units that have been surveyed. In these cases, economic transactions do occur. Statistical imputations are only needed to correctly estimate their value.

5. Depending on the institutional context, imputations tend to be more or less intensely debated. The topic is, for instance, important in Europe where national accounts data are used for administrative purposes for example to determine countries' financial contributions to the budget of the European Union or to allocate regional aid flows. All these operations require comparable and reliable figures. Some experts argue that the use of imputations should be minimised in this context so as to maximise reliability. Others support the idea of using imputations to reduce the impact of institutional differences on the comparability of results.

contributions, and measures of disposable income reflect such payments. But if these payments are deducted from household income, it would make sense to also add the value of goods and services to households' income and consumption. The latter adjustment is not included in conventional measures of household income and expenditures, but is included (through imputations) in the "adjusted" measures recommended by SNA (see Section 2.7).

- ***Own-account production of goods and services by households***: the value of all *goods* produced by households for their own consumption is imputed in the national accounts. In rich countries, this tends to be a small share of production and income, whereas the share is larger in developing countries. Own-account goods include, for example, home-grown agricultural products. Own-account services are excluded with one major exception, dwelling services 'produced' by owner-occupiers (i.e. imputed rents). Other economic activities of households (cleaning, cooking, child care etc.) are simply omitted from the SNA definition; they will be discussed further in this report in the section on broader measures of household production.

- ***Financial intermediation services indirectly measured (FISIM)***: only part of the services provided by financial institutions is explicitly priced, for instance through fees for deposit management. Other services, such as liquidity services or accounting services that a customer receives when opening a current account, are often not invoiced directly. Payment takes place indirectly, through interests paid to depositors are lower than market interest rates so that the interest margin constitutes the implicit price of the service rendered. This affects cross-county comparisons, as what may be implicitly priced in one country may be explicitly priced in another. Pricing may also change within a country over time. Hence, by making an imputation for these implicitly priced services, the overall value of financial services is invariant to changes in the pricing system or to differences in the pricing system between countries. A practical difficulty consists in allocating the overall amount of FISIM to the different sectors in the economy – in particular households and corporations.

- ***Consumption of non-life insurance services***: insurance premiums paid by households and firms comprise a payment into the 'insurance pool', which serves to settle claims, and a payment for the service that the insurance company provides by managing premiums and claims (e.g. by advising customers in their choice of policy). Broadly speaking, national accounts separate the two elements by assuming that the value of the insurance service corresponds to whatever is left after claims have been deducted from premiums. This (imputed) insurance service is considered in the SNA as a consumption item for households. Its measurement is fraught with difficulties.

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 two European countries (France and Finland) and for just over 20% in the United States. Thus, the living standards of French and Finnish households would be understated relative to the United States in the absence of imputations. Figure 1 Imputations in household income and consumption in France, 20071 provides a more detailed picture for France, and relates data on imputations to

both income and consumption. As it turns out, about 20 % of the final consumption expenditure figure for France can be attributed to the types of imputations described above[6].

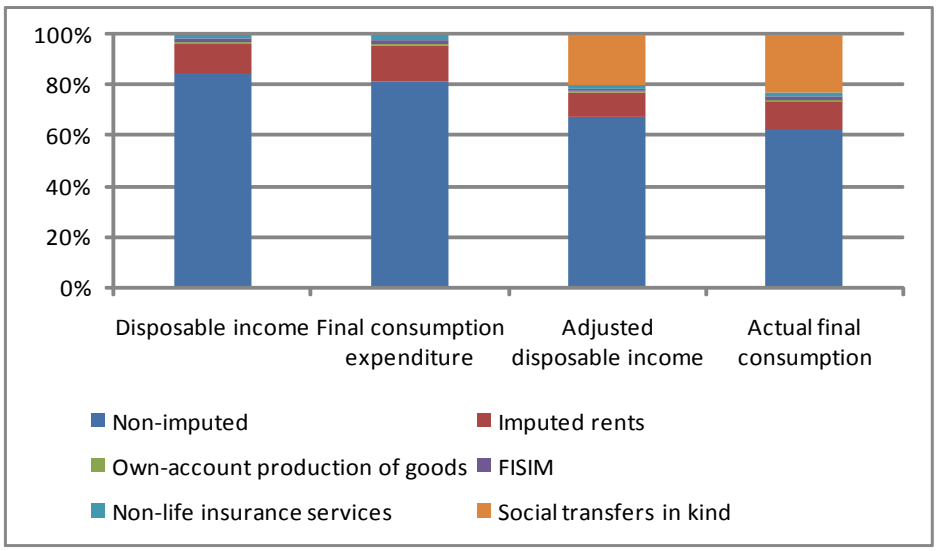
Table 1 Major imputed and non-imputed components of adjusted disposable household income *

	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*.

* Only financial intermediation services indirectly measured (FISIM) are imputed. Data available at the OECD do not allow distinguishing between FISIM and other financial services.

Figure 1 Imputations in household income and consumption in France, 2007



20. But imputations come at a price. One is data accuracy: imputed values tend to be less reliable than observed values because they often require assumptions about implicit transactions. The other is the effect of imputations on the comprehensibility of national accounts. Not all of these imputations will be easily understood. For example, people may not recognize the value of owner-occupied housing services as income, and the result may be a discrepancy between the changes in perceived and the changes in measured income. This problem is reinforced when the scope of economic activity is widened 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.

6. By their very nature, there is some uncertainty associated with imputed values. Part of it reflects differences in methods to derive imputations, for example for owner-occupied rents. See Diewert and Nakamura (2009) for a recent discussion.

21. The usefulness of including certain imputations into income and consumption statistics may depend on the countries considered. Deaton (2005) discusses the main income items that are imputed in the national accounts but absent from household surveys: imputed rents for owner-occupied dwellings and financial intermediation services indirectly measured. In India, the value of FISIM increased from close to zero in 1983/84 to 2.5 percent of consumption in 1993/94, accounting for a quarter of a percentage point per year of the difference in annual growth rates between national accounts and survey consumption data. Deaton questions whether, from the perspective of the poor, much of these financial intermediation services are relevant for their living standards. As average income or consumption as measured through the national accounts tends to grow more rapidly than measures drawn from household surveys, there is a danger that SNA measures paint too rosy a picture of the income and consumption of the poor if they benefit disproportionally from the items captured by imputations[7]. It may thus be that statistical procedures in poor countries "understate the rate of global poverty reduction, and overstate growth in the world." (Deaton 2005).

22. In developed countries, the picture may be different. In France, for instance, the share of FISIM in total disposable income is small (between 1 and 2%) and has hardly changed since the 1960s. The share of imputed rents, on the other hand, moved between 5% and 10% of disposable income since 1960. This reflects fluctuations of the prices of rents more than a steady trend. Similarly, the share of social transfers in kind in adjusted disposable income has risen steadily in France.

23. There is no easy way out of the tension between comprehensiveness and comprehensibility except making available to users both elements of information, and distinguishing between core and satellite accounts. A comprehensive household accounts, for example (see below), may not be well placed within the core national accounts system, but could be developed as a satellite account, providing a comprehensive valuation of comprehensive forms of household production.

3 - A first step: emphasizing national accounts aggregates other than GDP

3.1. Taking account of depreciation and depletion

24. GDP is a measure of the *amount of final goods and services produced within a country in a year (or a quarter)*. Gross output measures take no account of depreciation of capital goods. But 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. Thus, an immediate adjustment to GDP is to account for depreciation; doing so, leads to a measure of *net domestic product (NDP)*. Thus, ***net measures*** should be ***emphasized over gross measures*** of economic activity when the objective is to track standards of living.

7. There are many factors to explain the lower growth of survey-based household income and consumption. One is that richer households are less likely to participate in surveys. A second factor is that the national accounts' measures of household income and consumption contain several important and rapidly growing items that are not consumed by the poor and that are not included in surveys. So it is possible for consumption of the poor to grow less rapidly than national consumption, without any increase in measured inequality.

25. The reason why economists have relied more heavily on GDP than on NDP is, in part, that depreciation is hard to estimate. True economic depreciation is the change in the value of capital goods as a result of its wearing out or obsolescence[8] and the shortening of its future lifespan. But most businesses (and, often, national accountants) use simple rules: if a machine wears out on average in ten years, each year's depreciation is estimated to be one-tenth of the value of the machine. When the structure of production remains the same, GDP and NDP move closely together. Then, although we know that GDP overestimates the level of net output, percentage changes in GDP tended to provide a good measure of changes in NDP. But in recent years, the structure of production has changed. Information technology (IT) assets have become more important, and they account for a higher share of all capital goods. Life expectancy of computers and software is also shorter than of steel mills. On those grounds, the discrepancy between GDP and NDP should be increasing, and by implication, volume NDP should rise less rapidly than GDP. For example, nominal GDP in the United States rose by 6.1% per year during the period 1995-2007. The value of consumption of fixed capital (the national accountants' term for depreciation) rose by 6.9% over the same period. As a consequence, nominal NDP grows at a slower rate (6.0%) than GDP. These differences are small at the level of the entire economy but can be larger for individual industries.[9]

26. Of greater concern for some countries is that the standard measures of NDP have not taken into account the degradation in quality of the natural environment. The System of National Accounts foresees entries for the appearance and disappearance of natural economic assets as well as for the quality change in these assets due to economic uses. Thus, the degradation of land, water resources and other natural assets from economic activity is in theory recorded, as is the depletion of subsoil resources, as long as these natural assets qualify as economic assets (i.e. ownership rights can be exerted). Environmental resources that do not qualify as economic "assets" (such as air or biodiversity) are thus excluded from the measurement of degradation.[10]

27. Even for those natural resources that *are* recognized as economic assets, measures of depletion are rarely available in statistical practice, and where they exist, these entries do not show up as a charge against income: they are rather treated as a reduction in the quantity of assets held, just as one would treat the loss of a building due to an earthquake – a capital loss,

8. Obsolescence can be related either to technological advances, or simply to changes in factor prices. As wages rise, machines that require high inputs of labour become less valuable. Technological obsolescence is obviously of central importance for many high tech investments, which are assuming an increasing large share of GDP.

9. We have expressed the comparison between net and gross product in nominal terms because it is not immediately obvious how to interpret a *volume* or quantity change of depreciation or net *product*. A volume change of GDP can be relatively easily characterized as the volume change of all the final goods and services produced in the economy over a given accounting period. Depreciation is a charge against income and reflects a value loss of capital goods used in production that can be expressed in equivalent units of new capital goods but this is only one possibility. From a perspective of material standards of living, it is better to reason in terms or *real* net domestic *income*: the amount of resources generated during the period and expressed in equivalents of (consumption) goods or services. There are different ways to express real income – consumption product equivalents are most frequently used for household income whereas products of domestic demand are typically used as the equivalent 'unit' in which economy-wide real income is expressed.

10. Environment resources are not the only non-marketed assets which play an important role in determining societal well-being. Health, human capital, and knowledge are other assets that are at least partially non-marketed. See the discussion below.

but not a reduction in income. From an accounting point of view, two solutions can be imagined. First, the depletion of natural resources could be captured by excluding the value of natural resources from the production value of sectors like mining and timber. Their production would then only consist in a pure extraction or logging activity with a corresponding decrease in GDP. A second possibility is to take resource depletion into account in the depreciation measures. In this case, GDP would be unchanged but NDP would be lower. The hurdle is reliable monetary valuation of the natural resource that accounts for changes to environmental quality

28. A related question is how to treat *additions* to the stocks of natural resources. In principle, accounting should be symmetric: just as natural resources are reduced by extraction, they should increase with additions. For biological resources, this addition happens through natural growth. For subsoil reserves of minerals and energy, the *known* stock may rise as the result of mineral exploration and appraisal. Others would maintain that the stock of sub-soil assets is non-renewable on a human scale. If, and how much of such additions to stocks should be recognized as an addition to income, has been a matter of debate.

29. None of the issues arising from the recognition of depletion and environmental degradation in the national accounts are new, and much work on these issues has been undertaken at the national and at the international level, as demonstrated for example by the *International Handbook on Integrated Environmental and Economic Accounting* (SEEA - 2003), or by the ongoing work of the 'London Group', a gathering of statisticians who deal with the environment and economics. There are some conceptual problems (such as those alluded to in the previous paragraph), many measurement problems (especially concerning environmental degradation), and occasional political problems. Taking into account resource depletion would, for instance, suggest a smaller weight for sectors like coal mining and timber in the economy; and, in some instance, there has been lobbying against the development of more comprehensive accounts that would reflect resource depletion and adverse environmental effects. Further issues are posed by internationally harmonized implementation[11].

30. When the use of natural resources is not recognised as a cost of production, there is less incentive to use these resources optimally. If services provided by natural assets are shown, resource productivity – i.e. the efficiency with which natural resources are used – can be tracked and put on the same footing as measures of labour productivity or (produced) capital productivity. Work has progressed in this area[12] but much remains to be done before comparable measures of resource efficiency are available for a broad number of countries.

3.2. Domestic and national income

31. Although we have been referring to net *product*, it is more relevant (from a perspective of economic well-being) to refer to net *income*. 'Product' relates to the supply side of the economy whereas 'income' refers to the ultimate purpose of production, namely use for

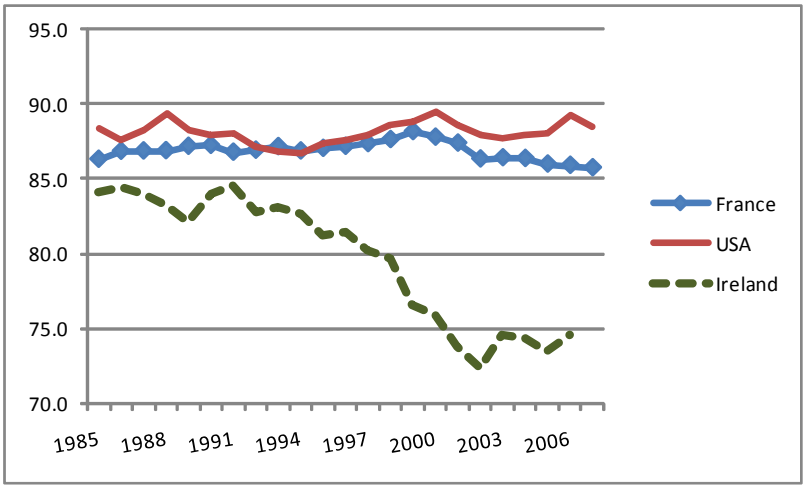
11. The United Nations Committee on Economic and Environmental Accounts (UNCEEA) is supervising the work in this area, with the aim of elevating the SEEA Handbook to a statistical standard.

12. For example, OECD Countries recently signed up to a resolution in this area; see *www.oecd.org/environment/resourceefficiency*

consumption and higher standards of living. In what follows, we shall therefore reason in terms of 'income' rather than 'product'. When turning to real income as opposed to its money value, a question arises on how to deflate nominal values. Whereas 'product' is typically expressed as the quantity or volume of goods and services produced, real income expresses the quantity of products that can be purchased with a given sum of money income. Before turning to the measurement of real income, additional adjustments will be discussed that can be brought to the measure of net (nominal or monetary) income.

32. Globalization may lead to large differences between measures of the income of a country and measures of its production. The former is more relevant to peoples' living standards because some of the income generated in production by residents is sent abroad, and some residents receive income from abroad. Consequently, a more relevant measure than GDP and NDP, in our search for a measure of living standards, is net national disposable income (NNDI, see Box 1). This measure accounts for payments and receipts of income to and from abroad[13]. This too is a standard variable in countries' national accounts.

Figure 2: Net national disposable income as percentage of gross domestic product



33. As production shifts from manufacturing to services, the differences between GDP and NNDI has increased in some countries. This affects the judgments of how well off people are. Assume, for instance, that more and more production occurs inside a country by firms owned abroad. While the profits generated by these firms are included in GDP, they do not enhance the spending power of the citizens of the country. For citizens of a poor developing country, to be told that GDP has gone up may be of little relevance; they want to know about their own living standards. This is especially the case in those countries relying heavily on mining or oil, which may receive a small royalty but where most of the returns accrue to the headquarters of a multinational company. Even among relatively wealthy OECD countries, the gap between NNDI and GDP can be of importance, as Figure 2: Net national disposable income as percentage of gross domestic product2 shows in the case of Ireland. There, the declining share of NNDI in GDP reflects the large foreign direct investments in the economy

13. For the treatment of international capital flows in national see Box 2.

and the large profits that are transferred outside of Ireland. By this measure, Irish income has increased by less than GDP growth would have suggested.

34. Changes in standards of living are determined by changes in both money income and prices of products that can be acquired with a given sum of money. One key determinant of real income is the relative price of foreign products, that is the rate at which exports may be traded against imports from the rest of the world (terms of trade). When a country's export prices rise more quickly than the prices of its imports, the country's citizens are better off and vice versa. Such gains or losses in the terms of trade can be important for small open economies, in particular when exports or imports are clustered around particular product groups. They are particularly important for small countries that export oil or other mineral resources but import a significant part of consumption.

Box 1. National Income and National Disposable Income - Two different concepts

Although national income (NI) and national disposable income (NDI) both refer to the income of the whole economy, NDI is a more comprehensive aggregate than NI. NI takes into account international transfers associated with the compensation of employees, taxes on production and imports, subsidies on products and production, and property income (interest payments, dividends , property income distributed to insurance policy holders, rents). NI is the starting point for computing NDI: to NI are added international transfers concerning current taxes on income and wealth, social contributions, social benefits in cash, and other current transfers from and to the rest of the world (e.g. non-life insurance premiums, non-life insurance claims, current international cooperation or current transfers between households).

The difference between NI and NDI hence reflects an element of income distribution between sectors. This is best explained by applying the concept to a household. A household's (primary) income consists of wages and property income such as dividends received. But households have to pay taxes and social contributions and they may receive social benefits and transfer payments. Accounting for these transactions leads to measures of disposable income. At the level of the whole economy, taxes, social security payments and so on that take place inside the country cancel out; but current transfers from and to other countries do not, and the difference between them mark the difference between NI and NDI. Thus, NDI better measures how well off citizens are.

Both NI and NDI can be computed gross or net of depreciation. As mentioned elsewhere, for the purpose at hand, net measures are conceptually preferable to gross measures.

35. Taking these changes in relative prices into account, along with real international transfers and real depreciation, yields a measure of real net national income for the entire economy. The figures below show that there is little to report home about the difference between constant price GDP and real net disposable income for some countries – the United States and France being cases in point. However, the example from Norway suggests that international price changes can drive a significant wedge between volume GDP and real income.[14,15] Norway's economy and real income benefited enormously from rising oil prices

14. Terms-of-trade effects can be worked into real income comparisons more systematically. The methodology for such measures has been worked out by Diewert and Morrison (1986), Kohli (1991) and Diewert et al. (2005) with the first application of the decomposition method to Australia in Diewert and Lawrence (2006). For a comprehensive treatment of terms-of-trade measurement, see IMF et al. (2009).

15. For each country, the deflator of net domestic demand (i.e.: final consumption + net capital formation) has been used to compute real income.

until 2008, allowing Norwegians to buy more imports for the same amount of oil exported. This is reflected in the more rapid rise of real net disposable income compared to that of GDP at constant prices. The measurement effect comes through because real NNDI is obtained by applying a price index for final domestic demand (final consumption and investment), part of which is imported. Note, however, that the 'net' calculation that underlies the Norwegian income measure does not reflect the depletion of Norwegian subsoil resources.

Figure 3: GDP and disposable income in the USA and in France

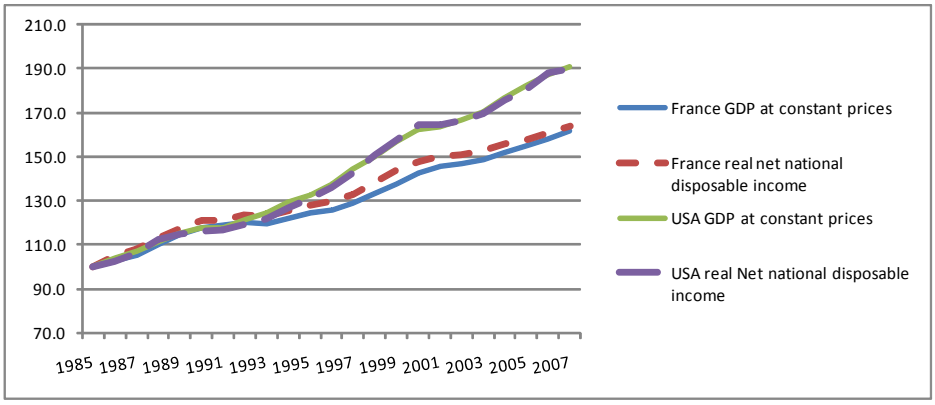
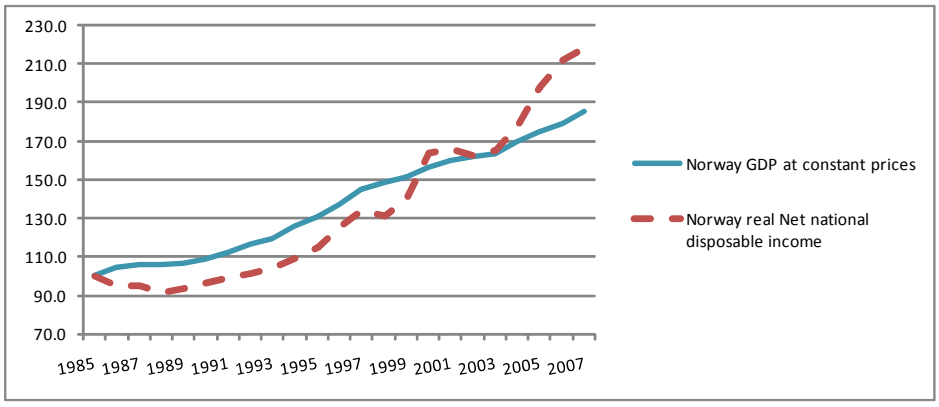


Figure 4: GDP and disposable income in Norway



Source: *OECD Annual National Accounts*.

36. Just as real income and volume GDP can be compared over time for a particular country, real income and volume GDP can be compared across countries at a particular point in time. Feenstra et al. (2009) describe the methodology underlying such spatial comparisons. The authors demonstrate large terms-of-trade effects for several countries, including Ireland, Mexico and Switzerland. Real income comparisons across countries and over time will also be discussed when considering measures for the living standards of private households in SectionSecond step: the household perspective below.

Box 2. National accounting treatment of international capital flows

Globalization entails not only more imports and exports of goods and services but also more international capital flows. Inflows and outflows of foreign capital are financial transactions that do not by themselves affect GDP. The impact on GDP is only indirect, for example through the investment expenditures that the international financial flows allow to finance. The same is true for outflows of foreign capital from a country: they impact on GDP only in the sense that some investment expenditures inside the country may be postponed or cancelled.

However, where foreign financial investments are important, national income may evolve differently from national product, as has for instance been the case in Ireland since the beginning of the 1990s.There is a special treatment of the income flows associated with foreign direct investments (FDI). These constitute investments of a long-term nature as opposed to 'portfolio investments' that are of a more short-term nature. The treatment of FDI is best explained by way of an example. Say there is a firm with its headquarters in country A (firm A) and it opens a subsidiary in country B (firm B). The following accounting treatment occurs:

- When the subsidiary is created, the flow of cash from firm A is recorded as a purchase by A of the shares or equity issued by B. This purchase constitutes a financial transaction. The cash received by B may then be used to finance a production line or a building. Only these transactions are recorded as formation of fixed capital, and they increase investment and GDP of country B.
- Now suppose that firm B earns profits and repatriates part of them back to firm A. The repatriated profits are property income for firm A and so they raise country A's national income. In fact, and by convention, national accounts count *all* of B's profits as property income (hence of national income) for country A. The imputed flow – the profits that are not repatriated – is recorded under the heading "reinvested earnings on foreign direct investment" in the national accounts. What is the overall consequence of this treatment? GDP of countries A and B is unaffected – all of the profits generated by the subsidiary are simply part of GDP in country B. However, the national income of country A will increase by an amount equal to total profits and the national income of country B will decrease by the same amount. This effect is independent of the extent to which profits have actually been repatriated or reinvested.
- But the story does not end here. Another flow, equivalent in amount to 'reinvested earnings', is entered as a financial transaction between the two countries. The reinvested profits are treated as if country A purchased more equity in country B. This is tantamount to saying that investments of firm B are always financed through the emission of new shares or a rise in equity rather than self-financed.

Now assume that firm B just accumulates retained earnings rather than purchasing capital goods with them. After several years of sitting in a bank account in country B, the cumulated profits are sent back to the mother company A. There is now neither an effect on GDP, nor on GNI and GNDI, of any of the countries. The transfer of cumulated profits is recorded as a sale of shares or equity and a transfer of currency, that is, as a financial transaction. Similarly, if firm A decides to sell its participation in firm B, the only direct consequence is for the financial accounts of countries A and B.

The treatment of reinvested earnings is an example of how imputations help reducing the dependence of income measures on specific institutional arrangements and accounting decisions of corporations.

3.3. Services in general and government-provided services in particular

37. In today's economies, services account for up to two thirds of total production and employment. At the same time, measuring the prices and volumes of services is more difficult than for goods. Retail services are a case in point. In principle, many aspects should be taken into account in the measurement of the services provided: the volume of goods transacted but also the quality of service (accessibility of the shop, general service level of the staff, choice and presentation of products and so forth). It is difficult to even define these services, let alone measuring them. Statistical offices generally use data on the volume of sales as indicators for the volume of trade services. This method leaves aside most quality change in the trade services provided. What is true for retail holds for many other service industries, including

those that are often publicly provided such as health and education. More efforts will be needed to get to grips with tracking the quantity and quality of services in modern economies.

38. Services provided by government are of particular interest. Broadly speaking, government provides 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. Education and health services are by far the most important individual services and their availability, accessibility and quality play a tremendous role in people's lives. Health and education are large sectors in most economies but their output tends to be badly measured. Traditionally, the output of government-provided non-market services is measured based on the *inputs* used to produce these services[16]. For clarity of the discussion to follow, two distinctions are helpful.

39. The first one is between the *values* and *volumes* of government-provided services. The second distinction is between the *health and education system* at large and the *institutions* that provide health and education services. The usefulness of both distinctions is best explained by way of an example. The United States spend more per capita on health care, and yet standard indicators of health outcomes are worse than in many European countries. Does this mean that Americans get less health care output? Or does it mean health care in the US 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? The distinction between values and volumes answer precisely this question. Over time, or across countries, we need to break down the change (or the difference across countries) in health expenditures into a price and a volume effect. But what exactly are the volumes that one is looking for? It is tempting to measure them by way of health or education outcomes, i.e., by the state of the health of the population or by the state of knowledge of pupils. The problem is that the link between expenditures and these outcomes is tenuous at best: expenditures relate to the resources that go into the institutions providing health and education services, whereas final outcomes (e.g. the health status of the population or the competencies of students) are driven by many factors. For example, people's lifestyle will affect health outcomes, in the same way as the time parents spend with their children will affect exam scores. Cutler, Deaton and Lleras-Muney (2006) examine causes for changes in mortality rates over time and they identify a host of factors other than medical care that may have had as big an impact on mortality as health care per se. Attributing changes in the health or education status only to hospitals and schools (and to the money spent on them) neglects all these factors.

40. How, then, are the *values* of government services measured in practice? The general practice used by most statistical offices is simply to add up costs linked to their provision. To interpret this from a standard of living perspective means assuming that costs are optimally allocated. However, as Atkinson and Stiglitz point out, this assumption breaks down "…once we recognise that government spending is financed by distortionary taxes. This has led some people to conclude that public provision stops short and that therefore we should value the

16. There are exceptions. For example, final consumption expenditures for pharmaceutical products provided by general government are based on product values. These are directly observed and do not have to be computed as the sum of inputs like for medical services.

output at an amount greater than the value of the inputs. However, we cannot say categorically that the value should be greater (Atkinson and Stiglitz, 1980, Lecture 16)."

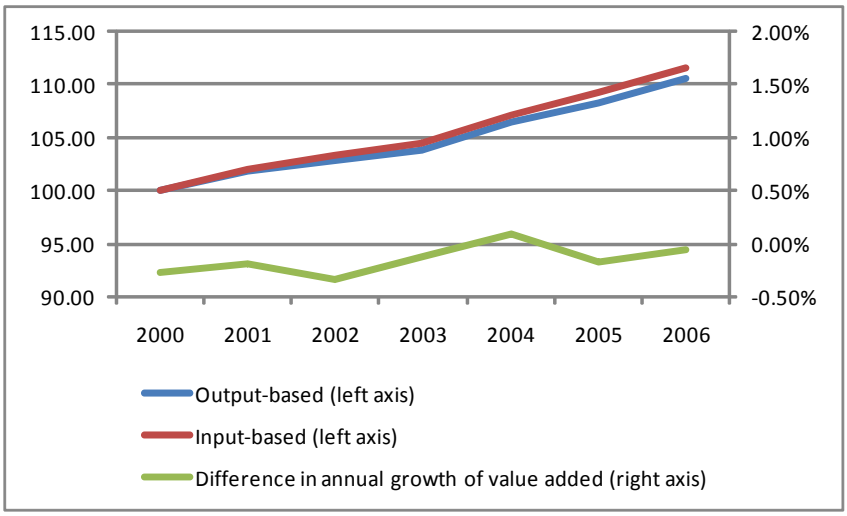
41. Privately-provided services would seem to escape this problem but this is only partially true. Consider privately provided medical services. Under standard conditions, market prices reflect individuals' marginal valuations of a good or service. But most purchases of health services are paid for by third parties, by individuals who are relatively uninformed about the marginal valuation: they are relying on other peoples' judgments. Thus, even in the absence of quality changes, there would be little reason to assume that prices reflect marginal valuations. Finally, in the area of health care, patients are constrained in their choice: while there is some substitution between treatments for a particular disease, there is certainly no substitution between diseases. Therefore, relative prices of treatment of diseases are in no obvious way indicative of people's preferences.

42. Turning to the measurement of *volumes* of government-provided services, these have traditionally been measured by the volumes of inputs and this is still the case in many countries. For example, the number of nurses or physician-hours may have been used to measure the volume of medical care. An immediate consequence of this procedure is that multifactor productivity change is ignored, as outputs are simply assumed to change at the same rhythm as inputs.

43. The quest is for more accurate measures of the volume growth of public services and work has started. Many European countries as well as Australia and New Zealand have developed output-based measures for government-provided services. A major challenge in these efforts is capturing quality change. Without a good measure of quality (or, equivalently, of a good estimate of the increase in productivity) it is impossible to ascertain whether the conventional input measures under- or overestimate growth of these sectors. If undifferentiated quantity measures (such as the global overall number of students or of patients) are used, changes in the composition of output and in its quality may be missed. The ultimate output of the education sector is the increased competencies of the students who are educated as a result in schools and universities; similarly, the ultimate output of the medical sector is the increased health status that results from medical care. In principle, quality change in the provision of these services could be gauged by the marginal contribution of medical care (or education) to the health status (or to human capital), while controlling for all other influences. In practice, this is fraught with problems, because it is difficult to separate the influences of medical care or education services from other factors that impact on outcomes.

44. Using the (admittedly imperfect) output measures focusing on patients treated or students educated has important effects, and these effects go in different directions. 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, input-based, convention had continued to be used, the growth rate would have been 3% (Atkinson 2005). Similarly, total economy value-added for France between 2000 and 2006 grows at an average rate of 2.0% per year when using an output method to compute volumes for non-market education and health, and of 2.15% using an input method (Figure 5: Impact of output and input-based methods on total value-added, France5). The difference is mostly due to non-market education.

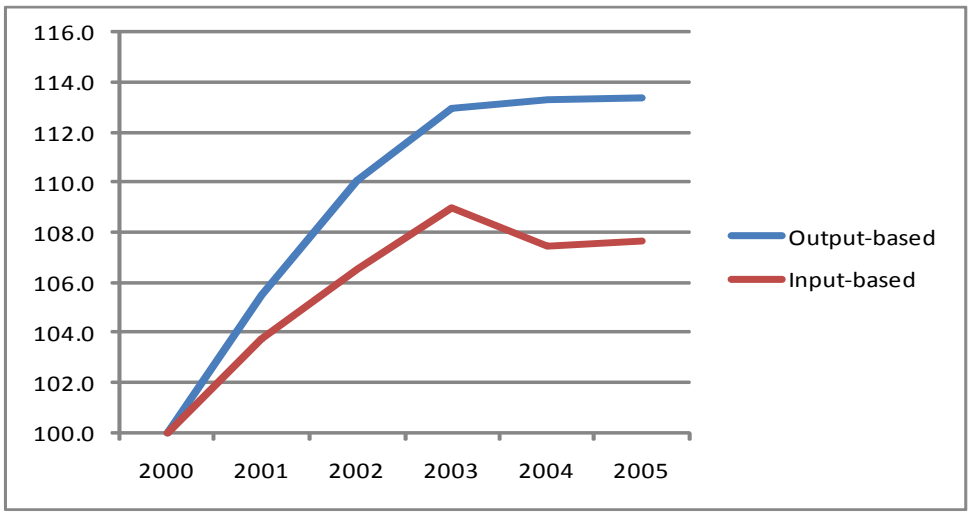
Figure 5: Impact of output and input-based methods on total value-added, France



Source: INSEE.

45. Turning to the effects of output versus input methods on the evolution of households' real disposable incomes, one finds annual differences of between 0.1 and 0.4 percentage points per year. In France, most adjustments when moving from an input to an output method are downwards, i.e., measured household income in real terms over the period 2000-2006 grows more slowly when using an output-based measure for government provided services. The effects, however, can also go in the opposite direction. 46. An important criterion for assessing the reliability of output-based measures is that they are based on observations that are detailed enough to avoid conflating true volume changes with compositional effects. For example, if spending per student increases, one might conclude that the unit cost of education services has increased. This may, however, be misleading if the higher costs result from students being taught in smaller classes, or if a larger share of students take up more costly engineering studies. The problem is that the number of students is a poor measure of output; better proxies of output are needed, for example by considering 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 achieve a certain level of implicit quality adjustment and control for 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, changes in health care administration in some countries have made available the administrative data needed to obtain this detailed information. 6 shows measures of value-added for the Danish health sector. The input-based index rose significantly slower than the output-based index, which captures treatment of diseases. In other words, the Danish health care industry experienced positive multifactor productivity growth. Measurement differences from the health care industry carry over – albeit with reduced strength – to total measures of government production and GDP.

Figure 6: Volume output health care services in Denmark



Source: Deveci, Heurlén and Sørensen (2008).

46. An important criterion for assessing the reliability of output-based measures is that they are based on observations that are detailed enough to avoid conflating true volume changes with compositional effects. For example, if spending per student increases, one might conclude that the unit cost of education services has increased. This may, however, be misleading if the higher costs result from students being taught in smaller classes, or if a larger share of students take up more costly engineering studies. The problem is that the number of students is a poor measure of output; better proxies of output are needed, for example by considering 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 achieve a certain level of *implicit quality adjustment* and control for 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, changes in health care administration in some countries have made available the administrative data needed to obtain this detailed information.

47. In addition to implicit quality adjustment through product differentiation, explicit quality adjustment may be called for and this is an important area of research. Explicit quality adjustments are very difficult to implement and typically micro-econometric studies are necessary. At this juncture it is difficult to draw a general conclusion as to whether measured GDP growth slows or accelerates as a consequence of moving from an input to an output-based method (see also OECD 2008a). While many empirical problems will have to be addressed before output-based volumes can be accurately measured, experience shows that progress can be made. Better measurement of government-provided individual services is central to the assessment of living standards. Exploitation of new administrative sources to deepen the level of information detail is one way of progressing in this direction. Lastly, the development of output-based measures of production does not dispense from improving the accuracy of the measures of inputs. Only if both inputs and outputs are adequately captured can there be reliable productivity measurement.

3.4. Defensive expenditures

48. Expenditures required to maintain consumption levels or the functioning of society could be viewed as intermediate inputs – i.e. these expenditures confer no direct benefit. Many such ‘defensive expenditures’ are incurred by government, others are incurred by the private sector. For example, expenditure on prisons could be considered a government-incurred defensive expenditure, and expenditures on commuting to work are an example of privately-incurred defensive expenditures. Several authors suggested treating these expenditures as intermediate rather than final products, hence excluding them from GDP.

Box 3. ‘Defensive expenditures’ in the literature

In 1939, John Hicks wrote that "the services of police, justice and defence do contribute to production, and may be thought of as used in production in the same way as power and fuel". By 1948, he reached the conclusion that a practical distinction between government activities which are final output and those which intermediate was feasible. Similarly, Kuznets (1951) considered that the expansion of government activities after World War II was so important that national accountants had to distinguish between final and intermediate output. He proposed several criteria to identify "the services by government to individuals that result directly in a flow of goods to ultimate consumers" and, therefore, had to be considered as final consumption:

First, that "the individual recipient of the service from government pays no price or only a token price";

Second, that "the government service be available to the individual only upon his overt initiative, rather than to him as a member of a social group who, as an individual, may be quite unaware of the service".

Third, that "the services by government to individuals have an analogue in the private markets". So, education, health, theaters and recreational activities may be included in final consumption of government services, but not justice, police and defence. Services meeting this criterion have a "widespread use" in private markets. This is to ensure that, for instance, police activities are not considered as service to ultimate consumers just because some people hire bodyguards on private markets.

Eisner (1988) argued along the same vein. He also pointed out an asymmetry: police services are counted as final product when provided by government but not when purchased by a corporation. If a large company agrees with a municipality to provide for its own security services in return for a reduction in local taxes, this would be counted as intermediate consumption by the company. Then he notes: “The point may well be made that police services, however useful or necessary, are in fact intermediate in nature no matter who pays for them. The final output is always the goods and services being produced, and the police necessary to protect them are simply a cost of that final output. Treating police services on this basis would enable us to avoid the anomaly of showing increases in real GNP as both crime and the police efforts to hold it in check rise apace.”

Defensive expenditures are not only limited to 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". As a result, they adjust income downwards for expenditures that arise as a consequence of urbanization and a complex modern life. This approach raises an issue of classifying certain goods and services as intermediate inputs rather than as final products.

Even when defensive expenditures are treated as final, the present report argues that some of the issues can be addressed by treating them as investment products rather than as consumption goods, as will be explained below.

49. At the same time, difficulties abound when it comes to identifying which expenditures are 'defensive', which are not and how to treat them in national accounts. What are possible ways forward? Options include:

- *First, focus on household consumption rather than total consumption*. For many purposes, this can be a meaningful variable. All of government consumption expenditures (such as prisons, military expenditures or the clean-up of oil spills) are excluded from households' final consumption. If one wants to capture 'individual consumption' provided by government, the SNA measure of actual final household consumption is appropriate because it captures these services provided by government (government 'collective consumption' is never imputed to households). This distinction between individual and collective consumption follows directly from the second criteria proposed by Kuznets (see Box) to distinguish between final and intermediate consumption of government services.

- *Second, widening the asset boundary.* An alternative solution would be to treat part of these activities as an investment. In many cases, there are elements of *investment* and *capital* goods in these defensive expenditures (as in the case of spending to remedy a deterioration in environmental quality). In those cases, defensive expenditures could be treated as maintenance expenditures: for example, expenditure on security could be regarded as an investment in social capital, and health expenditures as an investment in human capital. If there is an asset that captures environmental quality, expenditures made to improve or to maintain it could also be considered as an investment. Conversely, the consequences of economic activity that lower the quality or quantity of that asset could be captured in an extended measure of depreciation or depletion so that the *net* measure of income or production is reduced accordingly. Spending on pollution abatement reduces the adverse environmental effects—the benefits are realized in terms of a smaller deterioration in the quality of the environment that otherwise would occur.

- *Third, widening 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 (labor input) and their car (capital input) for this purpose. With the exception of the consumer's purchase of a ticket for a commuter train, none of the above flows currently enter measures of production and income. This could be remedied by allowing for household production of transportation services[17] that are provided as unpaid intermediate inputs to firms. Because firms don't actually pay for this input, a transfer from households to producers would also have to be imputed. Under the new treatment, value-added would have been shifted from firms to households. While this provides more accurate information about each sector's contribution to total value-added, it does not reduce GDP in the presence of commuting.

50. One consideration may help to decide whether collective non-market services provided by general government should be considered as intermediate consumption or as investment. By definition, intermediate consumption is an input into production that is *used up* within the accounting period. Collective services like national defence or security are conditions of

17. In the case where employers provide, say a bus, to ensure transport of employees to work, this would be registered as an intermediate input under present accounting conventions.

economic activity but, clearly, they are not "used up" during the accounting period. Moreover, as non-rival and non-exclusive public goods, they can benefit many processes of production at one time. To this extent, collective services better fit the definition of an (intangible) fixed asset. In order to make clear that this type of assets can be used simultaneously by all agents in the economy, a notion of collective investment, indicating the propriety of society as a whole on these assets, could be introduced.

51. Even individual consumption of non-market services such as education could be treated as an investment undertaken by the individual. Methods to value human capital exist in the economic literature (Jorgenson and Fraumeni 1989). The 1993 SNA treats expenditures on education as final consumption.[18] One reason is that other forms of personal investment (learning, studying…) contribute to the accumulation of the stock of human capital and they are not recorded as productive activities in the national accounts. Integrating them into the national accounts is possible but would require developing a system of household production. While the theory of valuing human capital is well-developed, this is not necessarily the case for other assets. As mentioned earlier, widening the asset boundary in national accounting comes at the price of added imputations and many empirical difficulties.

52. But there is no conceptual reason not to go down this avenue. Treating some non-market 'defensive' activities as an investment would have advantages:

– It would allow taking into account interactions between economic output and the level of these assets. For example, the contribution of health capital to output (health is related to the productivity of workers) could be estimated. Conversely, health capital could be explicitly affected by the conditions of production (for instance, industrial pollution could contribute to the depreciation of health capital).

– It would also help answering the criticism that an increase in defensive expenditures in response to, say, a deterioration of the state of security or the deterioration of the environment should not be recorded as an increase in the living standards. One could better distinguish the initial deterioration (which would appear negatively in the stock account) from the activity of repairing this deterioration (which would appear positively in the flow account). This treatment would be coherent with the fact that, conditional on damage having occurred, defensive expenditures do have a welfare enhancing effect.

– Moreover, this modification would not affect the level of GDP. However, it would affect NDP, and measures of net income as should be the case because they are relevant for living standards. More attention would be paid to balance sheets. Efforts along those lines would help develop better measures of sustainability, discussed later in this report.

53. In practice, changes in the statistical treatment of a given product (i.e. from intermediate input to investment) have happened on a number of occasions. The most prominent is the treatment of research and development expenditure (from intermediate input to a capital good, see Box 4 below).

18. The SNA states that "while knowledge, skills and qualifications are clearly assets in the broad sense of the term, they cannot be equated with fixed assets as understood in the [SNA]. They are not produced because they are acquired through learning, studying and practising - activities that are not themselves processes of production. The education services produced by schools, colleges, universities, etc. are consumed by students in the process of their acquiring knowledge and skills" (1993 SNA, § 1.52).

54. The biggest obstacle to these approaches lies in their implementation. What should be the scope of defensive expenditures? How should the new assets and in-kind flows be valued? By necessity, all accounting decisions would constitute simple conventions and the question remains whether the resulting data hides more than it reveals. That said, the questions around 'defensive expenditure' constitute a vast and important field of more research of a conceptual and of an empirical nature.

Box 4. From intermediate input to investment: the case of intellectual property products

The borderline between intermediate consumption and gross fixed capital formation has been a matter of discussion and has changed over time as the international guidelines concerning intellectual property products. Examples of these products are research and development (R&D), mineral exploration, computer software and entertainment, literary or artistic originals.

Under the 1993 SNA accounting standard, all expenditures in intellectual property products financed by corporations were treated as intermediate consumption. Government-financed R&D was treated as final consumption but neither private nor public R&D was treated as an investment good.

The new 2008 SNA recognizes R&D as part of capital formation. The reason is simple: R&D activity leads to knowledge assets that are of increasing importance for firms and economies and which should be recognized as such in the national accounts. A similar point can be made with regard to mineral exploration: it generates a stock of knowledge about reserves of subsoil assets. This shows that what is considered as an investment good or as an intermediate input is not cast in stone. However, conceptual feasibility does not necessary entail practical feasibility.

3.5. Income, consumption and wealth come together

55. Income flows are an important gauge for the standard of living but in the end, it is consumption and consumption possibilities *over time* that matter. The time dimension brings in wealth. A low income household with above-average wealth is better off than a low-income household without wealth. The existence of wealth is also one reason why income and consumption are not necessarily equal: for a given income, consumption can be raised by running down assets or by increasing debts and consumption can be reduced by saving and adding to assets. For this reason, wealth is an important indicator of the sustainability of actual consumption.

56. For firms as well as for an entire country, the information about wealth is brought together in balance sheets. To construct the balance sheet of an economy, we need 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 a country as a whole, we need to ascertain changes in total wealth – economic, social, and environmental. In some instances, it may be easier to account for *changes* in wealth than to estimate the *level* of wealth. The importance of measuring wealth, in all its main dimensions, is also a core recommendation of recent work on the measurement of sustainability by the UN-ECE, the OECD and Eurostat (2009).

57. Although information about many aspects of economic wealth is in principle available from national accounts' balance sheets, it is often incomplete. For some assets, price indices are incomplete or do not follow an agreed methodology. This concerns, for instance, the

single most important type of asset for private households, dwellings[19]. 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 wealth[20]. A systematic measurement of human capital stocks is, however, of interest from a number of perspectives. It constitutes an integral part of an extended measure of household production (see below), and it is a central input for the construction of sustainability indicators.

58. Changes in wealth entail gross investments (in physical and human capital) minus depreciation and depletion (of physical, human, and natural capital). Wealth also changes through revaluation: in the present economic crisis, plummeting house prices negatively affect many households, and revaluations of pension funds' assets directly influence consumption possibilities of pensioners. There is thus a direct link between stocks and flows and information on both is needed to assess peoples' living standards.

Box 5. National accounting treatment of bank activities

Banks and insurance companies constitute the bulk of what the national accounts call the financial corporations sector. Banks main activity is considered financial intermediation, i.e., "to channel funds from lenders to borrowers by intermediating between them. They collect funds from lenders and transform, or repackage, them in ways which suit the requirements of borrowers" (1993 SNA, § 4.78). Some financial intermediaries raise most of their funds by taking deposits, others do so by issuing bills, bonds or other securities. Conversely, they lend funds by making loans or advances, or by purchasing bills, bonds or other securities. They bear the risk of raising and lending funds at different horizons.

Financial intermediaries also supply means of payment and provide services such as currency exchange or advice about investments or taxation. Along with FISIM (see above), these activities are part of their production but their value is based on explicitly charges on customers. In the French national accounts, about one third of banks' production is imputed and the other two thirds are directly measured.

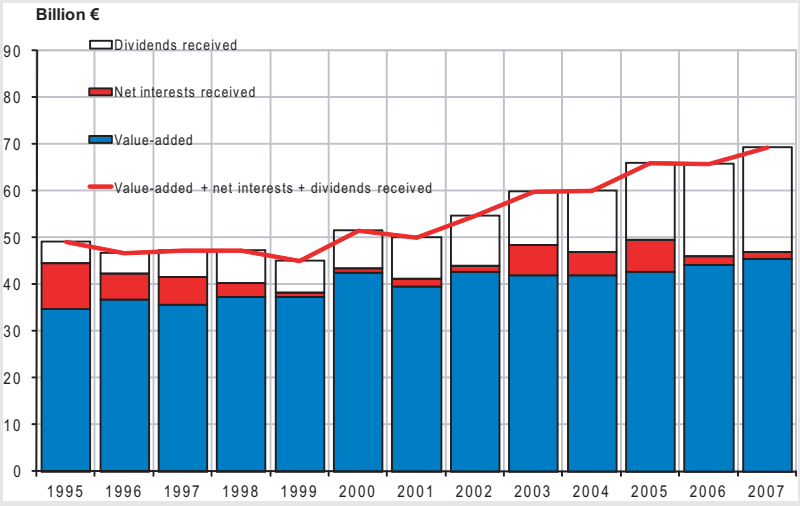
Financial corporations account for about 5% of total value-added in France and in Germany and for about 8% in the United States. This share is much higher in countries specialized in banking services such as Luxemburg (about 27%). Over the past decade, this share has remained quite stable in many OECD countries. This may seem odd given the large profits (as measured in business accounts) that financial institutions have reported in the years before the present financial crisis. It must not be forgotten however, that capital gains and losses are not part of the value-added generated by banks. This is a big difference between SNA and business accounting. For example, a well-known financial indicator for banks, net banking income, includes property income and capital gains and losses induced by own-account financial transactions.

Nor, for that matter, are interest flows (above and beyond FISIM) considered part of the value added generated by banks. Thus, the treatment of property income (interest and dividends) and capital gains or losses is one of the reasons why the share of financial intermediaries remained stable in countries such as France, Germany, the United States and Luxemburg between 1999 and 2007 although banks made record profits.

19. In 2009, work has started at the international level to develop a handbook on the measurement of dwelling prices but it will take some time before this effort leads to time series of dwelling prices with improved international comparability.

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

While neither property income nor capital gains are part of banks' value-added, a measure can be constructed that resembles net banking income as understood in banks' financial statements: net interest flows (received minus paid) and dividends received are added to value-added. The resulting measure differs from the 'balance of primary incomes' in national accounts because dividends paid are not deducted.. Capital gains and losses, on the other hand, are a purely financial occurrence that affects wealth but not income or production. The figure below shows income of French banks when net interests and dividends received are taken into consideration: these income flows have risen much more quickly than banks'value-added, i.e. their contribution to GDP.



Source : INSEE

59. Price changes in assets such as residential dwellings or shares are not part of income in a national accounts' sense [21]. They are revaluations of assets, rather than revenues from production. This does not change the fact that many persons (and not least tax authorities) *do* consider capital gains as a form of income that influences their behavior and economic well-being. This has led several authors to advocate the inclusion of capital gains and losses into measures of economic well-being (Eisner 1988).

60. In all probability, an income measure extended to include capital gains would exhibit greater volatility than the present measure. Consideration of capital gains and losses would also unevenly affect different income groups. For example, large capital gains or losses in shares would be proportionately more important for upper income groups – another field for research and numerical assessment. Here again, there are important conceptual problems. For example, in some countries, assets are increasingly held in individual retirement accounts (to which individuals have only limited access until they reach a certain age). Should the increase in share values be attributed back to individuals? And how should defined benefit schemes be accounted for? There is currently no consensus on how to answer these questions.

21. The International Expert Group on Household Income Statistics (Canberra Group 2001) produced a set of recommendations for how best to measure household income in surveys, which included the recommendation to exclude capital gains from the income measure.

61. The volatility of asset prices, apart from being a reason against including them in income measures, raises a more fundamental problem: we use prices to "add" apples and oranges because we believe that they represent a good measure of marginal valuations (the terms at which individuals are willing to trade off one good against another). But the instability of asset prices is suggestive of market failures. Did the true scarcity value of oil suddenly change from $147 a barrel to $36 a barrel in three months? If prices do reflect future resource scarcities, then the high volatility of asset prices does reflect volatility in societal well-being. But there are good reasons to believe that societal well-being is not so volatile.

62. Furthermore, the prices at which (market) assets are bought and sold are used to value the stock as a whole. However, there may be no markets for some assets, or no trading on markets that exist (as illustrated by recent experience for a range of financial assets): this raises the question on how to value assets in these cases. 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.

63. We conclude that measures of wealth are both important and uncertain. One way forward may indeed be to consider households' economic situation in terms of both income and assets. Another, more ambitious goal, is to consider lifetime measures, although this raises problems of implementation and of interpretation. Despite or because all these complications, basic information on assets and liabilities is key to assessing the economic health of all sectors and the financial risks to which they are exposed.

4 - Second step: the household perspective

64. Much of the public discussion about living standards focuses on indicators for the entire economy, and more often than not GDP. But, in the end, it is individuals whose economic situation should be assessed when talking about the standard of living. A glance at the evolution of the real income of households and the volume change in GDP (Figure 7: Real household disposable income and GDP Percentage growth at annual rate, 1996-20067) confirms that in general, one is not a good substitute for the other. Although in some countries real disposable income by households closely tracks volume GDP growth, there are many examples where this is not the case – Italy, Japan, Korea, Poland, Slovakia, Germany – to name a few.

65. In other words, the household perspective provides different and policy-relevant information in addition to GDP. Household-related figures are, however, not always easily found in accounting tables. It is not obvious for non-economists to establish a link between the aggregates recorded in national accounts publications and the income that he or she record in their tax returns. Some items may be recognisable but, more often than not, the meaning and the connection between different items are hard to follow and evolutions of income over time tend to be at variance with people's personal perception. A greater focus on people's well-being can be achieved by focusing on national accounts measures for private households.

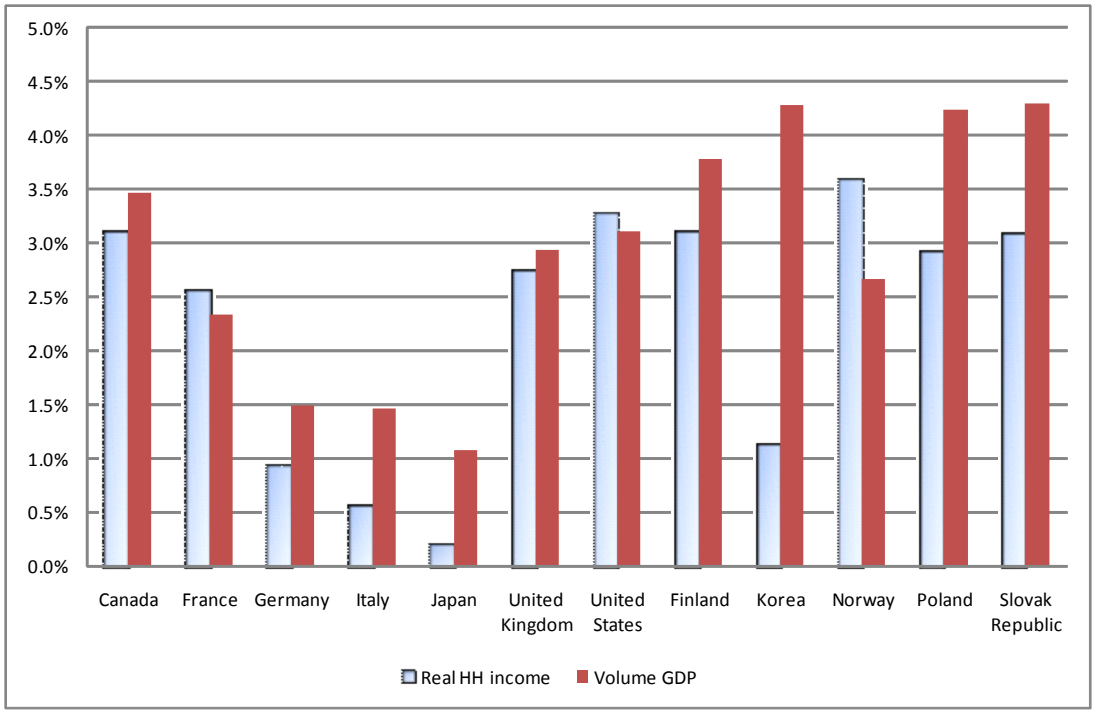
66. Measures of real household income appear to be suitable for this purpose. Earlier on, we discussed some of the income concepts for entire economies (national income, disposable income etc.). These income categories can also be computed for private households. In so doing, the re-distribution of income flows between economic sectors need to be taken into account. For example, some of the income of citizens is taken away in the form of taxes: this

is money that is not at their disposal. Conversely, households receive payments from government and this must be added to their income. Households also receive and pay property income, for instance dividends paid out by corporations and mortgage interest paid to banks. When all these monetary flows are taken into account, one ends up with a measure of household *disposable income*.

67. In a world of perfect, symmetric information and efficient markets, it could be argued that households see through the 'sectoral veil' and account, for instance, for the fact that eventually corporations are owned by households and that government expenditure now may lead to future taxes. But perfect information is an unrealistic assumption and more often than not, households will simply look at their income and wealth in judging their economic situation and their consumption possibilities.

68. While disposable income is a useful statistic, it suffers from an important asymmetry. Some of the money that government collects from citizens via taxes is used to provide public goods and services, and to invest in infrastructure. While disposable income measures add and subtract transfer payments between sectors, no adjustment is made for the value of the goods and services provided by government to households in return for the taxes they paid. When an adjustment is made for the value of goods and services received, one obtains a measure of *adjusted disposable income*. Similar to the adjustment of disposable income, an adjustment for government-provided services can be made to household consumption. This leads to a measure that has been labeled *actual final consumption* by households (Box 6).

Figure 7: Real household disposable income and GDP
Percentage growth at annual rate, 1996-2006



Source: OECD Annual National Accounts.

4.1. Adjusting income and consumption measures for government services provided in kind

69. The invariance principle mentioned earlier implies that a movement of an activity from the public to the private sector, or vice versa, should not change our measure of economic performance, unless the switch between public and private provision affects quality of the service or access to it. This is where a purely market-based measure of income 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 household disposable income* and *actual final consumption* are measures that go some way towards accommodating the invariance principle, at least where 'social transfers in kind' by government are concerned. These measures are computed by adding to household income and to household consumption expenditure the equivalent of the goods and services provided in kind by the government (Box 6).

70. The meaning of adjusted disposable income is best explained by way of an exampleTable 2: Private and public insurance schemes (2). Assume that a society's labour income equals 100 and that individuals who 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 a loss of 8) and 2 units of consumption of insurance services. At the same time, persons who are sick receive 8 units as reimbursement of their health expenditures. In this case – let us call it Case A – no taxes are paid and insurance claims and premiums offset each other, so that household disposable income 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 here 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.

71. The above example leaves aside any consideration about which insurance regime operates more cost effectively and about the effects of the profits made by private insurance companies– it was simply assumed that private and public insurance services are equivalent to 2 currency units. In practice, insurance costs between the two schemes are likely to differ, although it is difficult to establish a general pattern. If the insurance 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 and decreased insurance prices. Even if this has no consequence on household disposable income if profits are distributed to them in the form of dividends, it can make a difference on the accessibility of the insurance service. It is well known that the possibility to insure against certain types of risks has a positive impact on the well-being of people who are risk adverse.

Table 2: Private and public insurance schemes

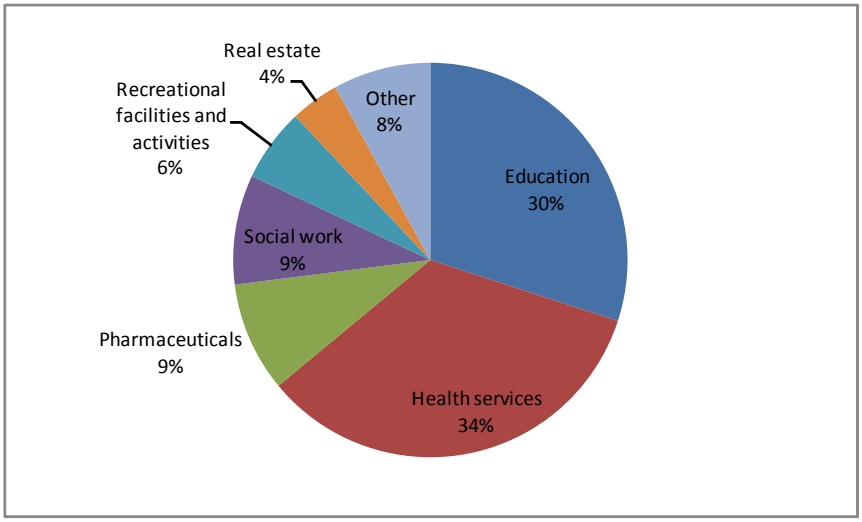
	Private insurance scheme (Case A)	Public insurance scheme (Case B)
Labour income	100	100
Tax	0	- 10
Insurance premiums (excluding insurance services)	- 8	0
Insurance claims	+ 8	0
Household disposable income	100	90
Social transfers in kind : - reimbursements - running costs of the insurance	0	+ 10 + 8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

72. While the failure to estimate the value of security services provided under the two schemes causes one set of biases, other biases may arise from the fact that the value of most 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 most households, who may receive little or nothing in return. The result is a large overestimation of the level of adjusted household income and consumption. Some of this overestimation can be tackled by using output-based volume measures for health and education services produced by government. It is also likely that different parts of the population benefit differently from social transfers in kind provided by government, which introduces an important distributional aspect.

73. It should also be noted that the neutrality of adjusted disposable income with regard to the public or private ownership of the service provider does not apply for collective services (e.g., security, environmental expenditures to reduce greenhouse gas emissions). When there is a switch from private to public provision of the service, conventional accounting lowers households' disposable income by the amount of tax payments although they may be better off. This may distort comparisons either over time or across countries.

74. Major items in social transfers in kind are health and education services, subsidized housing, sport and recreation facilities 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. As can be seen from74. Major items in social transfers in kind are health and education services, subsidized housing, sport and recreation facilities 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. As can be seen from 8, education and health services account for about one third of total transfers in kind, and housing and recreational and cultural activities (museums, public parks…) account for about 10%. The relative magnitude of these transfers varies significantly between countries.. 8, education and health services account for about one third of total transfers in kind, and housing and recreational and cultural activities (museums, public parks…) account for about 10%. The relative magnitude of these transfers varies significantly between countries..

Figure 8: Social transfers in kind, France 2007



Source: INSEE.

75. Annex D discusses in more details the questions concerning the treatment of private or public insurance and security functions in national accounts, focusing in particular on different types of insurances (health insurance, retirement insurance based on personal capitalization or following a pay-as-you-go scheme)

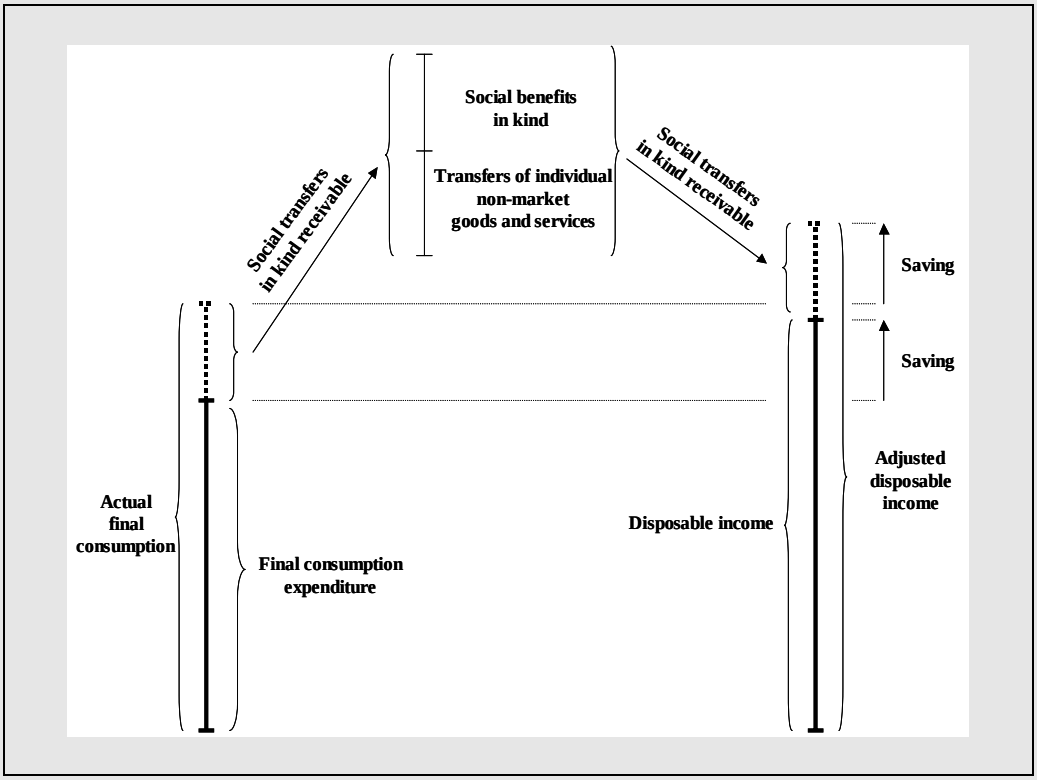
76. Measures of disposable household income are routinely compiled by a large majority of OECD countries. Despite the fact that the System of National Accounts also foresees measurement of adjusted household disposable income, there are gaps in its availability. For example, neither the United States, nor Canada or the United Kingdom regularly release such data. We conclude that giving more prominence to income measures of households, especially indicators of adjusted disposable income and actual individual consumption, are simple and useful ways to enhance the relevance of national accounts statistics to the measurement of material living standards.

Box 6. Two concepts of household consumption and disposable income in national accounts

The System of National Accounts draws a distinction between two concepts of household consumption:

- *Final consumption expenditure* is simply the expenditure incurred by households on consumption products, plus some imputed expenditure items such as the imputed rents that house owners pay to themselves (see above for a discussion).
- *Actual final consumption adds* to final consumption expenditure the value of social transfers in kind that households receive from governments. The value of these transfers is essentially measured by the costs that the government incurs to provide them to citizens. National accountants draw a further distinction between 'social benefits in kind' and 'transfers of individual non-market goods and services'. Social benefits in kind correspond, for instance, to pharmaceutical products that are reimbursed by the government. This consumption is recorded in government final consumption expenditure but in household actual final consumption. Transfers of individual non-market goods and services correspond, for instance, to the running costs of schools and hospitals.

These two concepts of consumption have a direct correspondence in two concepts of household income (*disposable income* and *adjusted disposable income)* discussed in the main text. The latter is derived from the former by adding the value of social transfers in kind. Savings can be computed as the difference between disposable income and final consumption expenditure or as the difference between adjusted disposable income and actual final consumption.



4.2. Medians, mean and the distribution of market income and consumption

Income

77. Average measures of household income per person are helpful but give no indication about how available resources are distributed across people and households. For example, *average* income per capita can remain unchanged while the distribution of income becomes less equal. To better monitor how people's income situation evolves, it is therefore necessary to look at income information referring to different groups. A simple way of capturing distribution aspects is to measure *median* income, i.e. the income such that half of all individuals are above that level and half below. The median individual is, in some sense, the "typical" individual. When inequality rises, median and average income may depart from each other, and a focus on average income will not give an accurate picture of the economic well-being of the 'typical' member of society. For example, if all the increases in societal income accrue to the richest 10% in society, median income may remain unchanged, while average income increases.

78. 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 the total population. To compute a mean, microeconomic information is needed that provides income information for

individuals or households. Microeconomic measures of household income refer to people living in private households and are typically derived from household surveys whereas macroeconomic measures are provided by the national accounts. The two sources are not necessarily compatible making it difficult to compute a median or other distributional information from surveys that is compatible with the average figure in the national accounts (Box 7).

Box 7. Differences between macro and micro-estimates of household income

There are several differences in concept and in statistical practice between income measures emanating from the national accounts ('macro-estimates') and income measures derived from household surveys ('micro-estimates'). Most important differences are:

- Micro-estimates of household income exclude people living in institutions as well as non-profit institutions serving households. Conversely, both of these are covered in macro-estimates.
- Micro-estimates of household income are generally limited to those income flows that are received by households on a regular basis, i.e. excluding those irregular flows that are instead covered in macro-estimates, such as bonuses.
- Household surveys generally refer to income received in cash or quasi-cash. Macro-estimates, on the other hand, include income in kind and several imputed items, such as the agricultural goods produced for own consumption (which are important in countries with a large subsistence-agriculture) and several types of property income. The single most important item in the national accounts is imputed rents from owner-occupied housing

Beyond these differences in definitions, other factors affect the measurement of individual income components in the two sources. For example, the national accounts include social security contributions paid by firms in both the "compensation of employees" received by households and in the taxes they pay, while survey measures of "gross earnings" are reported net of these contributions.

Reconciliation of micro- and macro-estimates is tedious but possible. Of particular import here is how imputed income flows can be distributed among income groups of households. Research in this area has progressed (see Annex A) but is still sporadic.

Some household surveys focus on consumption expenditure rather than income. In principle, many of the distinctions noted above are not relevant. Still, there are large, and possibly increasing, disparities in measures of household consumption from the two sources, due to reporting biases—upper income individuals may not report all of their consumption. Further problems are introduced by the fact that the prices paid for various goods may systematically vary across income groups.

Consumption, income and wealth

79. Income and its distribution are meaningful ways to assess living standards. Another candidate is consumption and its distribution among individuals. While correlated with income, consumption and its distribution are not necessarily identical to income and several reasons account for this. First, consumption tends to be driven by permanent, long-term income more than by short-term changes in income. Many households compensate short-term income fluctuations by increased savings or by borrowing. In this sense, income distributions should be more subject to transitory shifts in persons' income (which may reverse themselves after a while) than distributions of consumption.

80. Second, differences between a household's position in the income and in the consumption distribution are often reflective of differences in the distribution of wealth. Only some wealth effects are picked up by income measures, for example receipts of rents as part of property income. Other wealth effects such as realised capital gains are not normally reflected in income measures but are likely to affect consumption. Thus, consumption

patterns are more reflective of wealth than income patterns. One could also say that both income and wealth determine *consumption possibilities* which will then give rise to *actual consumption*. This explains why some authors have opted for wealth-adjusted income measures to capture consumption possibilities and their distribution in the population.

81. However, there are empirical advantages to using income rather than consumption data – income information at the individual level tends to be more readily available than consumption data.

Distribution measures

82. There is a large literature on how best to reflect the distribution of resources[22]. One of the most intuitive measures was used earlier in the present text, the difference between the mean and the median of a distribution. Other standard measures include the Lorenz curve (indicating the percentage of resources over which a certain percentage of the population has command) and the Gini coefficient (a summary measure of the area between exactly equality of the distribution of resources and the actual distribution, i.e., the Lorenz curve). Another common approach is dividing the population into quantile or decile groups and then following the share of resources of each group over time. Some authors have advocated combined measures, for example Yitzhaki (1979) who shows that the product of the Gini coefficient and average income reflects a concept of relative deprivation where people's living standards not only depend on the absolute income or consumption possibilities but also on where their consumption possibility are with regard to a reference group.

83. Some measure of distribution have been used to adjust aggregate variables, for example Sen (1976) proposed an adjustment of total income by (one minus) the Gini coefficient. Kolm (1969) and Atkinson (1970) developed indices that explicitly introduce distributional objectives into measures of inequality by adding a parameter reflecting the degree of aversion that society attaches to inequality. Jorgenson (1990) showed how information about consumption expenditure and aversion towards inequality can be combined to yield a measure of living standards.

84. Other aspects of the income distribution, beyond median income, are also important. For instance, social problems may be related to developments in the bottom part of the income distribution, which shape the living conditions of the poor. Real incomes of different quintiles of the income distribution can also be followed

85. In France, INSEE is presently working on a breakdown of household disposable income and consumption by household type in the national accounts. First results by quintile have already been published for the year 2003. Households are distinguished by household composition, income, age and social category of the reference person in the household (see Accardo et al., 2009).

86. Table 3: Trend in real household income* by 3 shows trends in survey-based measures of household disposable income deflated with a consumer price index over the past two decades. Thus, no account has been taken of the fact that different income groups may experience different rates of price change. Broadly, the earlier messages about France and the

22. The report of the Canberra Group (2001) provides an overview and Cowell (2000) a more extensive review of the various approaches.

United States are confirmed: over the past ten years, the United States have recorded a markedly slower income growth for people in the bottom quintile relative to the other quintiles whereas in France income in the bottom and middle quintiles grew about at the same rate as average income.

87. An important choice lies 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 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 economies of scale in housing and other costs. This adjustment is important as the size of households changes over time and differs across countries[23]. With average household size declining, increases in income per household will exceed those based on consumption units[24]. On the other hand, there are diseconomies of scale. The benefits of living in smaller units is, presumably, one of the reasons that as incomes have risen, household size has decreased. Thus, if one argues that people choose to live in smaller households and voluntarily foregoing the scale economies from living in larger units, the case for measuring income per consumption unit appears less compelling. Our measure of "consumption units" does not take this extra value from living in smaller units into account, and may accordingly understate increases in living standards. That said, many empirical studies have opted for the consumption unit, including a recent major report in France: Ruiz (2009) reports that real household income per consumption unit may decline even when total real household income per capita goes up. The author concludes also that presenting data in consumption unit brings perceived and measured real income changes better in line.

88. In many countries, there are important issues of intra-family distribution of income, consumption, and more generally and well-being. Unfortunately, data on these intra-family distributions is only available episodically. We know from time use studies (described below), however, that there are large gender differences.

Some results

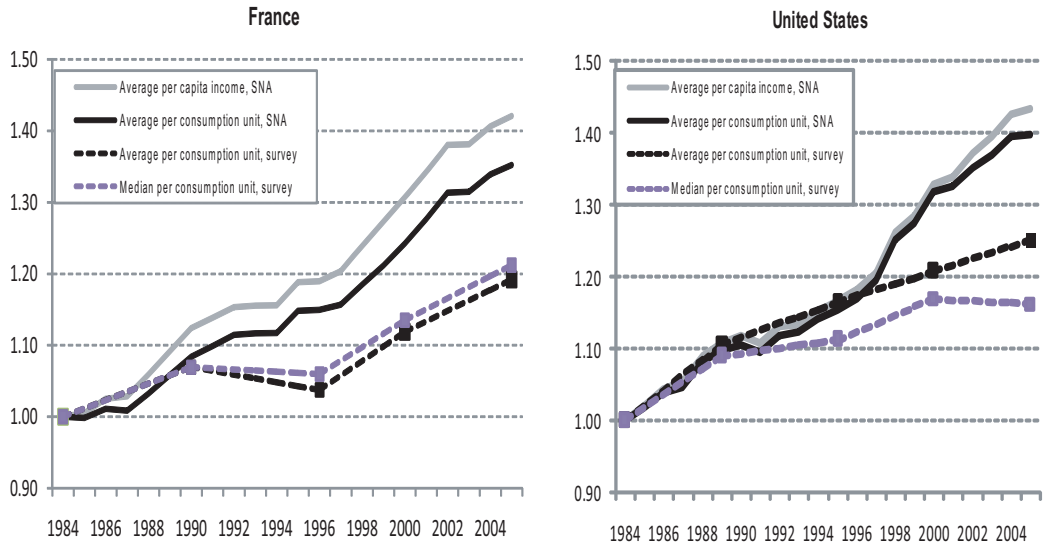
89. Against this background, we can consider the evolution of average and median household income in several countries 89. Against this background, we can consider the evolution of average and median household income in several countries show results for France and for the United States. These results are illustrations much more than definite data and Annex C provides more information on the statistical obstacles that needed to be overcome for this comparison. SNA measures of average household disposable income per

23. For the present purpose, the relation between the number of households and the number of consumption units was postulated as $CU = HH*(POP/HH)^{0.5}$. Thus, the number of consumption units CU equals the number of households HH, adjusted by the square root of the average household size POP/HH where POP stands for population. Note that this is only one of many possibilities to derive consumption units.

24. On the other hand, there are diseconomies of scale. The benefits of living in smaller units is, presumably, one of the reasons that as incomes have risen, household size has decreased. Our measure of "consumption units" does not take this into account, and may accordingly understate increases in living standards.

capita and per consumption unit diverge significantly in the case of France, reflecting a trend towards smaller household size. Survey measures also permit comparing trends in average and median income per consumption unit. In the case of France, these two items move in parallel: this reflects the broad stability of income distribution in France over this period. The picture is different for the United States, as evidenced by the widening gap between median and average income per consumption, reflecting the trends to a wider income distribution in this country. show results for France and for the United States. These results are illustrations much more than definite data and Annex C provides more information on the statistical obstacles that needed to be overcome for this comparison. SNA measures of average household disposable income per capita and per consumption unit diverge significantly in the case of France, reflecting a trend towards smaller household size. Survey measures also permit comparing trends in average and median income per consumption unit. In the case of France, these two items move in parallel: this reflects the broad stability of income distribution in France over this period. The picture is different for the United States, as evidenced by the widening gap between median and average income per consumption, reflecting the trends to a wider income distribution in this country.

Figure 9: Trends in different measures of household disposable income



Source: Computations based on OECD SNA and income distribution data (see Annex C).

90. Many measurement issues influence the above comparisons. An important source of discrepancy between micro and macro estimates is property income. If this aggregate is not well measured in micro estimates, and if the distribution of property income becomes more unequal than the distribution of other income, this could explain why average and median incomes in these estimates have a parallel evolution in France. Also, the international comparability between household surveys is far from perfect.

91. Another way in which national accounts and household data may be brought closer together is by looking more closely at the share of wages in total income. For example, Atkinson and Voitchovsky (2008) examine trends in the share in total Net Domestic Income of the wages earned by all workers, by the bottom 90 per cent and by the bottom 50 per cent

of workers based on data for the United Kingdom from the Annual Survey of Hours and Earnings/New Earnings Survey data. The evolution of this wage share is certainly different when one considers the bottom half of the distribution. In the 1950s, when the overall wage share was rising slightly, the bottom half suffered a reduction in their share. Between 1954 and 1964, the wage share rose by 1½ percentage points, but the share of the bottom half of workers fell by 2 percentage points. In 2006, the overall wage share is virtually the same as it was in 1954, but the share of the bottom 50 per cent was 4 percentage points lower. Of this fall, approximately a third can be attributed to the decline in the wage share and two-thirds to increased wage dispersion.

Table 3: Trend in real household income* by quintiles

	Average annual change mid-1980s to mid-1990s					Average annual change mid-1990s to mid-2000s				
	Bottom quintile	Middle three quintiles	Top quintile	Median	Mean	Bottom quintile	Middle three quintiles	Top quintile	Median	Mean
Australia	..	..	..	..	..	2.4	2.0	1.9	2.2	2.0
Austria[1]	2.5	2.7	2.8	2.8	2.7	-2.1	-0.5	-0.4	-0.6	-0.6
Belgium[1]	1.2	0.5	1.2	0.4	0.8	1.4	1.3	1.7	1.2	1.5
Canada	0.3	-0.2	-0.1	-0.2	-0.1	0.2	1.2	2.1	1.1	1.4
Czech Republic	..	..	..	..	..	0.4	0.6	0.7	0.5	0.6
Denmark	1.3	0.9	0.8	0.9	0.9	0.6	0.9	1.5	0.9	1.1
Finland	0.9	0.9	1.0	0.8	1.2	1.6	2.5	4.6	2.5	2.9
France	1.0	0.5	-0.1	0.5	0.3	0.9	0.7	1.0	0.8	0.8
Germany	0.4	1.4	1.6	1.2	1.4	-0.3	0.5	1.3	0.6	0.7
Greece	0.3	0.1	0.1	0.3	0.1	3.6	3.0	2.7	2.9	2.9
Hungary	..	..	..	..	..	0.9	1.2	1.0	1.1	1.1
Ireland[1]	4.0	3.0	2.9	3.2	3.1	5.2	7.7	5.4	8.2	6.6
Italy	-1.3	0.5	1.5	0.6	0.8	2.2	1.0	1.6	1.0	1.3
Japan	0.8	1.8	2.1	1.8	1.9	-1.4	-1.0	-1.3	-1.0	-1.1
Luxembourg	2.3	2.5	3.0	2.4	2.7	1.5	1.5	1.7	1.5	1.6
Mexico	0.7	1.2	3.8	1.1	2.6	-0.1	-0.1	-0.6	-0.2	-0.4
Netherlands	1.1	2.7	3.9	2.8	3.0	1.8	2.0	1.4	2.0	1.8
New Zealand	-1.1	-0.5	1.6	-0.6	0.3	1.1	2.2	1.6	2.3	1.9
Norway	-0.3	0.3	1.0	0.4	0.5	4.4	3.9	5.1	3.8	4.3
Portugal[1]	5.7	6.5	8.7	6.2	7.3	5.0	4.1	4.4	4.2	4.3
Spain[1]	4.4	3.2	2.4	3.2	3.0	5.2	5.1	5.0	5.5	5.1
Sweden	0.5	0.9	1.2	0.9	0.9	1.4	2.2	2.8	2.2	2.3
Turkey	-0.6	-0.7	1.4	-0.8	0.4	-1.1	-0.5	-3.2	-0.3	-1.9
United Kingdom	0.7	2.0	4.3	1.9	2.8	2.4	2.1	1.5	2.1	1.9
United States	1.2	1.0	1.9	1.0	1.4	-0.2	0.5	1.1	0.4	0.7
OECD-22[2]	1.2	1.4	2.1	1.4	1.7	1.5	1.8	1.9	1.9	1.8
OECD-20[3]	1.3	1.5	2.1	1.5	1.7	1.7	2.0	2.2	2.1	2.1

1. Changes over the period mid-1990s to around 2000 for Austria, the Czech Republic, Belgium, Ireland, Portugal and Spain (where 2005 data, based on EU-SILC, are not deemed to be comparable with those for earlier years).

2. OECD-22 refers to the simple average for all countries with data spanning the entire period (i.e. excluding Australia, the Czech Republic and Hungary, as well as Iceland, Korea, Poland, the Slovak Republic and Switzerland).

3. OECD-20 refers to all countries mentioned above except Mexico and Turkey.

*Note that income in this table is based on household surveys and consequently not comparable to national accounts-based income. Income flows have been deflated with each country's consumer price index.

Source: OECD (2008a); *Growing Unequal?* Paris.

92. From a perspective of living standards, the distribution of income and wealth determines who enjoys access to the consumption of goods and services produced within a society. One of the reasons why average per capita measures of income, consumption and wealth often fail to reflect peoples' perceptions of how their resources and consumption possibilities change over time is that the benefits of growth are not equally distributed – some people in society can be worse off even if average incomes have increased. Complementing

measures of average income, consumption and wealth by measures that reflect their distribution should become standard practice when annual national accounts figures are published. Ideally, such distributional measures should be conceptually compatible with average measures from the national accounts.

4.3. Price indices – temporal comparison

93. To convert nominal measures of income and consumption into real measures, price indices are required. For comparisons of living standards, price indices of bundles of consumer goods and services are the appropriate choice. It is useful to dwell on the nature of the price index to be employed for the measurement of real income and to briefly discuss the notion of cost-of-living indices.

94. Recently, the *Panel on Conceptual, Measurement, and other Statistical Issues in Developing Cost-of-Living Indexes* (Schultze and Mackie eds., 2002) presented a series of recommendations for the development of the US Consumer Price Index. Many points raised in this report are directly relevant for the questions at hand. In particular, the report refers to the concept of cost-of-living indices and to the distinction between conditional and unconditional indices. A cost-of-living index aims at measuring the relative change in expenditure that a household would have to make in order to maintain a given standard of living[25]. Every cost-of-living index has to be defined with respect to a particular domain, or scope (i.e. the goods and services included in the index), keeping other factors fixed. Thus, changes in the standard of living are measured with conditions outside the scope of the index held constant. What exactly should be within the scope of the index, and on which factors the index should be conditioned, depends on the analytical question at hand. The *Panel*, for example, recommended that for purposes of constructing the U.S. Consumer Price Index, only private (market) goods and services should be inside the scope of the index, i.e. changes in non-market commodities (leisure), environmental conditions (climate) and other societal factors (crime) should not influence the evolution of the index.

95. For other analytical purposes, however, the scope of cost-of-living indexes could be different and this is the case for the measures discussed in the present document. More specifically, for the purpose of comparing living standards, the domain of a cost-of-living index could be made commensurate with a broader notion of income, encompassing at a minimum, government-provided services such as health and education. A further extension (see below) would be to include non-market services that households produce themselves and, possibly, leisure. Accounting for goods and services provided for free or at subsidized prices by government is not without problems, however. The inclusion of such services introduces uncertainty about data quality as, by definition, there are no prices for non-market products. At the same time, non-market goods and services *are* important elements for the living conditions of citizens and their provision potentially affects international comparisons of living standards[26].

25. One particular feature of a cost-of-living index is that it takes into account the effects of substitution by consumers when the relative prices of goods and services change. This is mainly achieved by the choice of superlative index number formulae (Diewert 1976). Diewert (2001) and Triplett (2001) provide in-depth discussions of the consumer price index as a cost-of-living index.

96. Table 4: Price changes and real income of households and total economy 1995-20064 provides estimates of real income trends based of alternative choices for the relevant deflators. Several patterns stand out.

Table 4: Price changes and real income of households and total economy 1995-2006
Percentage change at annual rate

	France	United States	Finland
Disposable income* (current prices)	3.7%	5.2%	4.4%
Price index final consumption expenditure	1.3%	2.0%	1.6%
Real disposable income	2.4%	3.1%	2.8%
Adjusted disposable income* (current prices)	3.8%	5.2%	4.5%
Price index actual final consumption	1.6%	2.2%	2.0%
Real adjusted disposable income*	2.2%	3.1%	2.5%
National disposable income (current prices)	3.6%	5.6%	5.7%
Price index domestic demand	1.6%	2.1%	1.9%
Real national disposable income**	2.1%	3.5%	3.8%

*For private households and non-profit institutions serving households; net of depreciation.
**Note that deflation with a domestic demand deflator is but one possibility to obtain real income.

- First, the nominal change in disposable household income is close that for adjusted disposable income for France and the United States, but less so for Finland. . Similarly, the price indices of final consumption expenditure and actual final consumption are quite similar in France and in the United States but not so in Finland. Whether higher inflation in government-provided services than in market-provided services in Finland reflects economic reality or the statistical method employed to measure price (or unit cost) non-market services remains an open issue. But the results show that moving from one price index to another is not a trivial matter, and that getting price and volume measures for publicly-provided services right makes a difference for the assessment of households' living standards. More detail on this issue in conjunction with government provided health and education services is provided below.
- Second, (and confirming the result already highlighted in Figure 7: Real household disposable income and GDP Percentage growth at annual rate, 1996-20067) differences in the growth rates of household real income and economy-wide real income can be significant, as shown by the Finnish case (where real household income has grown by more than one percentage point less per year over the past decade than economy-wide

26. The cost-of-living index proposed here would have the following components: (i) prices of goods and services of private final consumption, i.e. market prices directly available from national accounts; and (ii) prices of individual services provided by government (such as health and education). For this second group of items, the lack of market prices implies that unit costs should enter the cost-of-living index. Whether these unit costs should refer to the costs per unit of output or the costs per unit of input is an important question and country practice varies on this point. When income or consumption measures are further extended to include the services that households produce for themselves, the scope of the cost-of-living index should be equally extended to cover the unit costs of producing these services.

real income). The gap is smaller for the United States and absent in France but likely to be present in many other countries.

97. While the concepts and problems that are involved in constructing good price indices are well understood[27], the rapid changes in relative prices, economic structures, and product design and characteristics have meant that conventional methods towards measuring price change may be inadequate in capturing these changes. For instance, in 1996 the US *Advisory Committee to Study the Consumer Price Index* (widely known as the Boskin Commission after its chair, Michael Boskin) concluded that the CPI overstated the cost of living by about 1.1 percentage points per year, implying that real growth was understated by the same amount. The report called attention to problems associated with adjustments for quality changes, to the appearance of new products (particularly relevant in fast changing sectors like health and IT, but also in retailing), and to problems in data collection of data (e.g. the increasing fraction of sales done over the internet, and at discount stores). Quality change is also an important topic in the report by the *Panel on Conceptual, Measurement, and other Statistical Issues in Developing Cost-of-Living Indexes* (Schultze and Mackie eds., 2002). Diewert (1998) estimates that for a measured price change of about 2% per year, the quality bias may be as large as 0.4 percentage points in the US CPI. Deaton (1998), on the other hand, questions the practical feasibility as well as some of the conceptual foundations for the need to capture quality change and to target a cost-of-living index. The reader is referred to the extensive literature on price indices that brings out the many conceptual and empirical issues involved in capturing quality change. It is evident that there is no simple or unique methodology.

98. A point of particular relevance from a welfare perspective is the question about 'whose' price index is evaluated. 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 the costs of an average bundle of goods. However, different people buy different bundles of goods (e.g. poor people spend more on food and less on entertainment) and they may buy their 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 may have become more marked and people at the bottom of the income distribution may have seen real incomes fall by much more than those at the top of the income distribution.

99. Deaton (1998) relates the income-specific price indices back to the problem of quality measurement. He argues that quality effects are income related and that the "benefits of quality upgrading and of new goods will only be distributionally neutral if the affected goods are neither luxuries nor necessities. While it is true that advances in the technology of consumption have benefitted people in all parts of the income distribution, it is hard to believe that many of the relevant goods are not luxuries. When new goods are consumed disproportionately by the rich, whose price indexes are weighted in the CPI according to their incomes, the quality-corrected plutocratic index can provide a very poor measure of prices for the average consumer."

27. For an overview of price index theory and price measurement see Diewert (1987), IMF et al. (2004) and ILO et al. (2004).

100. A price index for (actual) private consumption[28] of major groups in society (by age, income, or place of residence such as rural vs. urban population) is necessary if we are to appraise their economic situation. One of the recommendations of the *Commission mésures 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 renters and for households that are about to purchase dwellings (Ruiz 2009). Other relevant categories potentially exist. However, such prices indices are not available in most countries. While indices with different weights for different groups of the population are easily computed and exist in several countries (such as the United Kingdom, France, Germany), they tend to show relatively similar movements. A full development of price indices differentiated by socio-economic groups would require, however, the collection of different prices for different parts of the population, so that socio-economic aspects are taken into account in data collection design. This is likely to be difficult and costly. This development should constitute a medium-term objective – a recommendation that echoes a similar conclusion by Schultze and Mackie (2002)[29]. Such work would not only foster the quality of deflation procedures, but also make it easier for citizens to assess their personal situation through some of the income and price data released by statistical offices.

4.4. Price indices – spatial comparison

101. Just as there are price indices to deflate nominal income or consumption over time within a country, price indices are also required to deflate nominal income or consumption for comparisons in space, i.e. across countries or regions. Spatial price indices, or *Purchasing Power Parities (PPP), have* a long tradition in the OECD area[30]; recently, new world-wide PPPs have been released by the World Bank. Many of the issues outlined above for temporal price indices are shared by PPPs. For example, comparing different qualities of the same type of product across countries is inherently difficult. The equivalent of the problem of new goods in temporal price indices is the problem of goods existing in one country but not in another. In many respects, measurement issues are even more pronounced with PPPs than with temporal price indices.

102. For the purpose at hand – international comparisons of real income and consumption - PPPs need to be chosen to best reflect the different goods and services that are of interest for our comparison. A meaningful spatial deflator for disposable income (and of course for final consumption expenditure) is the PPP for final consumption expenditure, while a meaningful spatial deflator for adjusted disposable income is the PPP for actual individual consumption. We use 2005 benchmark PPPs for the comparison at hand.

28. For a discussion of cost-of-living indices in conjunction with the production of household services see Hill (2009).

29. The Schultze commission states: "BLS [the United States Bureau of Labor Statistics] should pursue an exploratory research program that would, initially only on a small scale, investigate and assess several alternative approaches – including, but not limited to, the use by survey respondents of handheld scanners and computers – for collecting prices in a way that allows them to be associated with household characteristics. A first objective might be the production of indexes for a few commodity categories and several demographic groups." Schultze and Mackie (eds.) (2002), p. 5.

30. See, for example, the OECD website under http://www.oecd.org/topicstatsportal/0,3398,en_2825_495691_1_1_1_1_1,00.html

103. Table 5: Real income of households, comparisons across countries, 20055 shows that in 2005 French per capita disposable household income was at 66% of the U.S. level, and Finnish per capita disposable income at 49% of the U.S. level[31]. When government-provided goods and services are taken into account, this gap is significantly reduced: real adjusted disposable income per capita in France is 79% of the U.S. level and Finland's at 68%. This, of course, reflects the fact that the two European countries have a larger share of social transfers in kind (which raise final comparison) than the United States. Finally, all household-related comparisons paint a picture that is quite distinct from one based on GDP per capita. We should add that even this real income comparison for households is incomplete from the perspective of living standards, as important non-market activities by households and leisure are neglected.

Table 5: Real income of households, comparisons across countries, 2005

		France	United States	Finland
Real disposable income* per capita*	USD	19338	29448	14395
	USA=100	66	100	49
Real adjusted disposable income** per capita	USD	25378	32110	21771
	USA=100	79	100	68
Real GDP per capita (total economy)	USD	30519	41740	30460
	USA=100	73	100	73

*For private households and non-profit institutions serving households; net of depreciation; converted with PPPs for final consumption expenditure
**Equals disposable income corrected for social transfers in kind and converted with PPPs for actual individual consumption

Source: OECD Annual National Accounts.

4.5. Risk and vulnerability

104. Considering wealth brings to the fore considerations about risk, exposure to risk and vulnerability. For example, when pension systems increasingly rely on private providers and when households invest in riskier financial assets than hitherto, this creates greater exposure of households' financial resources to volatility in financial markets. The simple indicator of riskiness of household financial assets for France (i.e. the share of households wealth held in assets with a high exposure to investment risks) shows a clear growing exposure of French households towards more risky assets. It should be noted that the figure reflects both direct and indirect exposure of French households, i.e. the exposure of portfolios of institutional investors who offer mutual fund shares and life insurance policies to households is also accounted for in the indicator (we refer to this adjustment as the "transparency" adjustment)

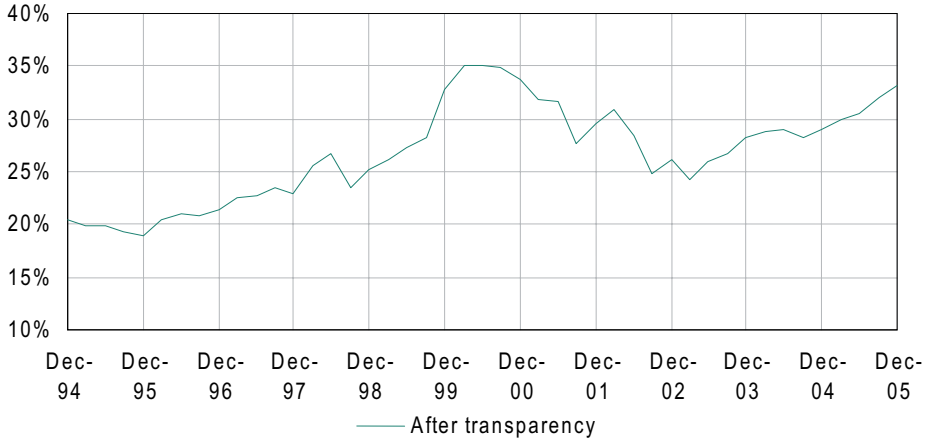
105. The riskiness indicator shown above concerns all French households. As with measures of income, it would be helpful to break this measure down by socio-economic group, thereby tracking which part of the population is most exposed to financial volatility.

106. On an annual basis, we could think of individuals as self-insuring, putting aside, say a 3% reserve for the expected annual loss from fire. Similarly, we could think of individuals

31. Income levels in France and Finland relative to the United States are higher than relative consumption levels because of higher savings rates in the two European countries.

setting aside a reserve against a portfolio loss. If individuals undertake greater risk, they should set aside greater reserves. In reality, individuals don't do that. If the income generated by undertaking greater risk is included in conventional income measures (for instance through performance-related bonuses for financial investors), while capital losses are not, this leads to an asymmetry that may bias long term measures. At the national level, the accounting consequences of greater exposure to risks become important in making inferences about the effect of different policies. Using metrics that do not reflect future expected losses may, for instance, give an excessively optimistic view of the effects of deregulations of the financial sector. As a general point, the current national accounting system is ill-equipped to reflect financial risk and this may lead to biased pictures of corporations' and households' exposure to risk. .

Figure 10: Share of risky assets in households' financial wealth, France



Source: Marionnet (2007).

4.6. Broader measures of households' economic activity

107. Production and income measures in the national accounts include all non-market *goods* produced by households[32], as well as one important *service* item, i.e. the value of rents that owners of dwellings 'pay to themselves'[33]. However, no other services that households produce for themselves[34] are measured in the standard accounts. This can generate biases on several fronts, one of which is to exaggerate the growth rates of developing countries (Deaton 2005) where this type of production is relatively important.

108. Service activities that households undertake for themselves include cleaning, cooking, childcare, driving to work and so on. In principle, the time spent on these productive activities can be measured, valued and integrated with the traditional national accounts and income

32. Not only households undertake non-market production. Government is another big non-market producer but, unlike households, the services that it provides are recognised in the standard national accounts.

33. In France, imputed rents account for about14% of final consumption expenditure. In the United States, the share is about 11%.

34. Nor is production and use of services recorded that households produce for free for others such as free software.

measures. Past studies[35] have shown that these household services are important and that their consideration is likely to change the level, distribution and growth of expanded measures of household income, consumption (and investment). As such services can substitute goods and services purchased in the market, even if they are not directly marketable, they can affect how individuals dispose of marketed income.

109. There have been large changes in the functioning of households and the society. For example, many of the services that people received from their family in the past are now purchased on the market. This shift translates into a rise of income, as measured in the national account, and this may give a false impression of a change in living standards, while it only reflects a shift from non-market to market provision of services[36]. A shift from private to public provision of a particular product should not affect measured output. By the same token, a shift of production from market to household production or vice versa, should not affect measured output. In practice, this invariance principle is not assured by current conventions on the measurement of household services.

110. 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 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 the same overall pay ($50,000 a year), 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.

111. A more complete picture of household production could be drawn by constructing a full set of accounts for households. Such an endeavor would not only consider the household services mentioned above but also include education, as an investment that increases the magnitude of human capital. Comprehensive accounting of this investment would, of course, have to take account of the depreciation of human capital, e.g. as a result of aging. Studies in the United States (Jorgenson and Fraumeni 1989, 1992) that used such a comprehensive measure of productive activity found that the resulting numbers were large: for example, investment in human capital turned out to be at least four times investment in non-human capital.

112. Finally, recognizing household production may alter our assessment of the pace of economic growth and of the distribution of income and consumption. A recent major report

35. For example, Rüger and Varjonen (2008); Landefeld and McCulla (2000), Landefeld, Fraumeni and Vojtech (2006).

36. See Folbre and Wagman (1993) and well as Wagman and Folbre (1996). Note that there may be extra social value in the fact that production takes place via market transactions rather than within households due to the simple fact that jobs are created and that jobs tend to constitute more than just sources of earnings to individual. In this sense, a shift between market and non-market production may not be neutral. Of course, the same point can be made with inverted signs – some household production activities may generate utility above and beyond the market equivalent value. Raising children comes to mind as an example.

on non-market activities and their measurement (Abraham and Mackie, eds., 2005) sums this consideration up as follows:

> "The key point in this discussion is that economic growth can alter the relative importance of home and market production. That in turn may lead to incorrect inferences about how fast average economic well-being is growing if only market GDP is measured. Since changes in the scope of home production will differ across income groups as their opportunities change, ignoring it in measuring incomes also will bias conclusions about how inequality is changing." (Abraham and Mackie, p. 62).

113. In what follows, we present some illustrative calculations of the implications of complementing conventionally-measured household income by non-market production of household services, i.e. moving towards a notion of "full income" (Becker 1965) and full consumption.

4.7. Time use

114. Table 6 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 emails and travel time related to all these activities. 'Personal care' is mainly sleeping, eating and drinking, whereas 'paid work' includes the time spend in paid work or study either at home or in the workplace, and the commuting time related to it. 'Leisure' was defined residually, and it includes sports, religious and spiritual activities and other leisure activities. The time spent on personal care activities in various countries has been normalized across countries to the lowest value because, to minimise the effect of international differences in survey designs on results[37]. At the same time, normalisation bears the risk of ironing over true differences in behaviour between countries.

115. Another ambiguity in the data presented relates to travelling. It constitutes a case of time use whose allocation to production or consumption is not straightforward. Adding up travelling and housework attenuates differences between some European countries and the United States in time spent on non-market production. The categorization we use follows the convention in the time-use literature and allocates time spent travelling to the activity that it is associated with. That is, travel time related to paid employment is added to paid work and travel time for child care and housework is added to time for non-market work. There is also some arbitrariness to the definition of what makes up leisure. In particular, eating and drinking are included in 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.

116. With the normalised data and the above caveats (for more detailed information, see Annex A), one finds that more time is spent on household production in European countries as opposed to the United States. The United States is also one of the few countries where more time is spent on paid work than on housework (Freeman and Schettkat, 2005). The picture concerning leisure is similar. More time is spent on leisure in Finland, France, Italy, Germany and the United Kingdom than in the United States (6). Note that these differences

37. More discussion of methods and limitations to international comparability of time use surveys are provided in OECD (2009).

would increase further if eating time were allocated differently. According to the time use surveys at hand, Americans spend on average 74 minutes per day on eating and drinking whereas the same activity takes up 135 minutes on average in France. 7 shows an alternative allocation of eating and drinking time – half of which has been allocated to leisure instead of personal care. Also, no normalisation has been made for the personal care time. Leisure time in France now turns out to be lower than in the United States. The time spent on unpaid household work now is nearly identical between the two countries. We conclude that the allocation of specific activities to time use categories as well as their international comparison leaves room for improvement and harmonisation.

Table 6: Time use of population (age 16 and older)
Minutes per day, normalized for personal care activities; latest years available*

	Personal care	Paid work			Unpaid work			Leisure			Unspecified time use	Total
		Total	Men	Women	Total	Men	Women	Total	Men	Women		
Germany	**616**	**196**	251	144	**219**	161	273	**394**	371	349	**15**	**1440**
Italy	**616**	**219**	300	143	**232**	107	352	**360**	343	264	**13**	**1440**
UK	**616**	**229**	289	171	**228**	160	291	**356**	355	322	**11**	**1440**
France	**616**	**215**	269	167	**215**	147	276	**358**	283	249	**36**	**1440**
Finland	**616**	**222**	261	185	**209**	159	251	**376**	376	348	**17**	**1440**
USA	**616**	**245**	291	203	**213**	161	261	**344**	332	294	**22**	**1440**

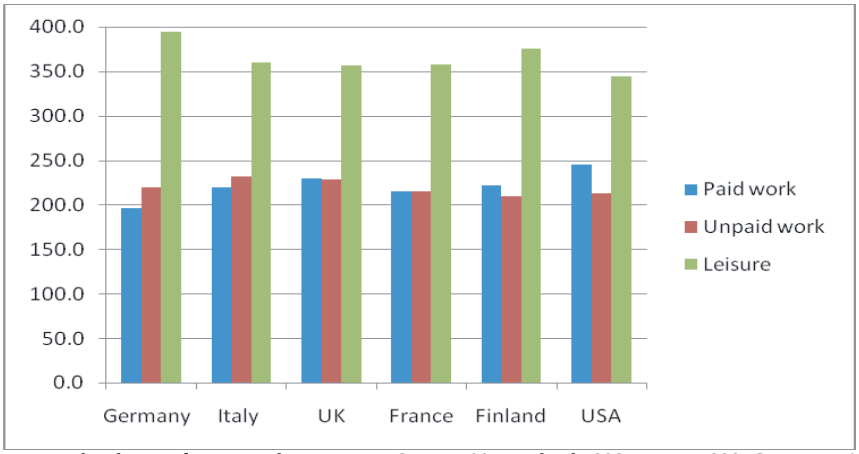
*Unites States: 2005, Finland 1998, France 1998, Germany 2002, Italy 2003, United Kingdom 2001.
Source: OECD (2009), based on HETUS and ATUS databases.

Table 7: Time use of population (age 16 and older)
Minutes per day, 50% of time for eating and drinking allocated to leisure; latest years available*

	Personal care	Paid work	Education	Unpaid work	Leisure	Unspecified time use	Total time
Germany	598	183	13	219	413	15	1440
Italy	360	207	12	232	616	13	1440
UK	592	220	9	228	381	11	1440
France	642	201	14	215	333	36	1440
Finland	591	207	15	209	402	17	1440
USA	611	227	18	213	349	22	1440

*Unites States: 2005, Finland 1998, France 1998, Germany 2002, Italy 2003, United Kingdom 2001.
Source: OECD (2009), based on HETUS and ATUS databases.

Figure 11: 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.

117. Gender differences in time use are significant. In each of the countries under consideration, men spend more time in paid work than women and the converse is true for unpaid work. Men also spend more time on leisure than women. The implication is that women provide household services but other members of the household benefit. This distributional issue within households gets lost when income or consumption per household are considered.

Table 8: Time spent in different activities, ratio men/women

	Paid work	Unpaid work	Leisure
Germany	1.74	0.66	1.36
Italy	2.10	0.62	2.17
UK	1.69	0.75	1.43
France	1.61	0.78	1.46
Finland	1.41	0.89	1.31
USA	1.44	0.95	1.32

*Unites States: 2005, Finland 1998, France 1999, Germany 2002, Italy 2003, United Kingdom 2001.
Source: OECD (2009), based on HETUS and ATUS databases.

118. 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 comparability between surveys is often 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 statistical task for the future will be to build up internationally comparable time series of information on how people use time to allow assessing trends. Such work is under way in the United States and in several European countries but lacking in many other parts of the world.

4.8. Valuing the production of household services

119. It is possible to provide an illustrative calculation for the value of household production for France, Finland and for the United States. The approach chosen here is simple[38]: the value of household services is measured by their costs[39]. Two basic cost items are considered, the value of labor inputs and the value of capital services from consumer durables. In what follows, we will also assume that the volume of household production varies with the volume of the labor and capital inputs. This assumption implies that there is no productivity change in the production of household services. Many of the points that were made earlier in conjunction with measuring the value of government services apply here as well, i.e. our method to quantify values and volumes of household production is input-based.

38. For more elaborate and focused studies, for example on childcare, see Folbre and Jayoung (2008).

39. Schreyer and Diewert (2009) show that this choice implies either that households are constrained in their possibility to deliver additional hours to the labour market or that households choose to engage in own-account production because the net benefit from spending time on household work is positive. Alternatively, household production could be given a market valuation if an equivalent product is offered on the market. This complicates matters, however, and has not been pursued here. For a recent discussion with indications of the quantitative impact of this methodological choice, see Fraumeni (2008).

120. The value of labor is estimated by applying a wage rate net of taxes and social security contributions 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[40]. Data on the wage rate for household workers are drawn from the 2005 OECD/Eurostat Purchasing Power Parities price collection. This wage rate is, however, only available for one year and for the present purpose, a time series has to be constructed. We do so by taking the rate of change in the hourly wage for each country's 'Other community, social and personal services' industry, based on the assumption that household workers' wages have evolved at the same pace as wages in the parent industry[41]. Thus, the value of household production will go up if these wage rates increase and/or if the number of hours spent in housework go up. The value of household production will also increase if the price and/or the quantity of capital services from consumer durables increase.

121. The quantity of labor input is captured by the time spent in household work. As argued above, there are also uncertainties about this evolution; furthermore, no consistent time-use surveys exist over time. We therefore relied on two hypotheses of how this quantity has changed over time: Variant A, where the time per day and per person allocated to household work remains constant; and Variant B, where the time allocated to household work changes at the same rate as the time allocated to paid work (which implies a decline in all countries). The main rationale for Variant A is its simplicity – for want of better information on the evolution of household work, we take it to be unchanged. Variant B implies that housework is not a substitute to paid work, and that work and leisure operate as substitutes. Various alternative assumptions could be made but uncertainty could only be reduced through consistent time-use observations.

122. The value of capital services is measured by constructing a stock of consumer durables and multiplying it by an imputed price of capital services for consumer durables[42]. The volume input of capital services from consumer durables is captured by the net stock of consumer durables. More details on the methods used to compute labor and capital services as well as results are available from Annex A.

123. A central implicit assumption in measuring the volume of household production by the volume of labor and capital inputs is that there is no multifactor productivity (MFP) growth. A growth in MFP would imply that household labor and capital are combined more efficiently

40. Landefeld, Fraumeni and Vojtech (2009) show alternative estimates of the value of household production for the United States. Their estimate of household production for 2004 excluding services of consumer durables and owner-occupied housing, and using a housekeeper wage rate to value time spent on household production is 19% of conventionally measured GDP. Our own estimate excluding services of consumer durables for the United States amounts to 22% of conventionally measured GDP for the year 2004.

41. There is no strong empirical justification for this choice. If the economy-wide development of wages is used to construct a time series, the resulting figures look quite different because economy-wide wages tended to rise quicker than wages in the community, social and personal services industry. The resulting estimates of the value of household production are directly affected by this choice.

42. The housing services that households which owe their own dwellings provide to themselves are already accounted for in income measures in the national accounts. To avoid duplication they are not included here. However, it should be clear that the value of household work as presented here is not a complete characterisation of household own account production of services.

in the production of household services. Note, however, that labor productivity may change if more or less services from durables are used per unit of labor input. Annex A presents therefore two variants, one without MFP growth, and one with an assumed MFP growth of 0.5% per year.

124. The value of household services presented here is a first approximation. It is apparent, consistently with previous studies, that imputations for own-account production of household services are sizeable matter. Under the various assumptions made, household production amounts to about 35% of conventionally-measured GDP in France (average 1995-2006), about 32% in Finland and 30% in the United States[43]. The value of household production comprises the value of household work (about 22% of GDP between 1995-2006 in the US, about 30% in France and 29% in Finland) and the value of capital services provided by consumer[44] durables.

125. Some of the time spent by households on unpaid work is devoted to produce goods for their own account such as growing vegetables, carrying out construction and maintenance activities for their own dwellings. As the production values of these activities are already recognized in the national accounts, our valuation implies an element of double-counting. For the three countries under consideration, however, the size of this bias is likely to be small. This may not be the case for countries in the developing world, where agricultural production carried out by households for their own consumption may be a more important item.

126. From a perspective of measuring living standards, it is also of interest to know by how much a broader measure of household production changes measured income and consumption. And there, real measures over time and in international comparison are telling indicators. 9 shows comparisons of real income growth over the last decade or so. Accounting for household production reduces the measured growth rates of real income in all three countries by a significant margin. The general pattern of growth for each country is not markedly modified nor is the relative growth pattern between countries. Differences are, however, more important when comparing income *levels* across countries, as shown in . The first lines in this table show a comparison based on real adjusted disposable income (that is, an established national accounts aggregate), the third and fourth line compare real income including household production. Because household production is more important in France and Finland than in the United States, the new measure narrows the per capita gap in household incomes between the two European countries and the United States.

127. It is worth reiterating that many assumptions enter the comparisons presented here. The resulting numbers are therefore not on very solid, given the absence of ‘hard’ data throughout several steps in their computation. Thus, much care needs to be exercised to avoid over-

43. There are no annual observations for time use surveys. We generated such a series by using information on annual hours actually worked (on the labour market), based on labour force surveys. The trend in hours per person was used to extrapolate backwards the hours spent on paid work as available from the ATUS and HETUS time use surveys. Time spent on personal care was kept constant. Given annual observations for time other than paid work and personal care, the distinction was made between Variants A and B – time spent on housework constant and time spent on housework changing at the same rate as time spent on paid work. Under each variant, the remaining time was allocated to the remaining activities in the same proportions as observed in the latest years of the ATUS and HETUS time use surveys.

44. Over longer periods, durable goods will make little difference, since we include the value of these goods in GDP, and over the service life of durables, their (discounted) value is equal to the value of the services provided.

interpretation of the results. This should not prevent official statistics from regularly estimating the full value of household production. Household satellite accounts would allow accommodating the lower data quality and frequency of these estimates.

Table 9: Real income of households, corrected for household production, 1995-2006
Percentage change at annual rate

	France	United States	Finland
Real adjusted disposable income*	2.2%	3.1%	2.5%
Real adjusted disposable income*, corrected for household production**	1.9%	2.9%	2.2%

*For private households and non-profit institutions serving households; net of depreciation.
**'Income from household production' equals the value of the estimed labour and capital inputs used by households to produce own-account services (except owner occupied housing); deflated with price index for actual individual consumption corrected for household production

Source: OECD estimates.

Table 10: Real income of households, corrected for household production, 2005

		France	United States	Finland
Real adjusted disposable income* per capita	USD per capita	25378	32110	21771
	USA=100	79	100	68
Real adjusted disposable income* per capita, corrected for household production**	USD per capita	35037	41904	29208
	USA=100	84	100	70

*For private households and non-profit institutions serving households; net of depreciation, converted with PPPs for actual individual consumption
**'Income from household production' equals the value of the estimed labour and capital inputs used by households to produce own-account services (except owner occupied housing); converted with PPPs for actual individual consumption corrected for household production

Source: OECD estimates.

4.9. Valuing leisure

128. Once one starts thinking about non-market income, one has to think about leisure. Time spent on generating income (market or non-market) provides the means to buy goods and services to meet our needs, in addition to a range of non-market returns. Time available for leisure affects well-being in more direct ways. If follows that changes in the amount of leisure over time, and differences between countries, represent one of the more important aspects for comparative assessment of economic well-being. Focusing only on goods and services can bias comparative measures of well-being 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 scale of economic production, especially of goods, beyond limit. Taxes and regulations will be imposed that will both discourage the production of goods and change the way they are produced. It would be a mistake, beyond the decrease of production and consumption as they are currently recorded, not to consider that an increase in leisure time can benefit well-being.

129. As society progresses, it is reasonable to expect that people will enjoy some of the fruits of that progress to in the form of leisure[45]. Stiglitz (forthcoming) observes that over the past 30 years or so, Americans have hardly increased their leisure time, despite the large increase in the volume of goods and services at their disposal. He also notes the growing difference in leisure patterns between the United States and other countries and questions what this implies for comparisons of living standards. Whatever the precise answer (which may well lie outside the realm of traditional economics), it is the case that different societies may respond differently to the consumption-leisure choice, and we do not want to bias our judgments (e.g. of success) *against* societies that choose to enjoy leisure by excluding leisure from our measurement system.

130. It is difficult to put a monetary value on leisure. Economists have worked around this issue by treating leisure time like a consumer good whose price is the value of income from work foregone[46]. Nordhaus and Tobin (1973) were among the first to make an explicit adjustment of national income measures for the value of leisure, recognising that more leisure adds to the well-being of persons. A number of measurement issues need to be addressed in this context (Boarini et al. 2006). For example, leisure of a person who is constrained in his/ her supply of additional labour on the labour market is not necessarily valued the same way as leisure of a person who is not constrained. The quantity of leisure is sometime difficult to identify, due to the difficulty in drawing the borderline between personal care (sleeping etc.) and leisure. The exact valuation is also a matter of debate – which wage rate should be chosen? Should the quantity of leisure be adjusted for higher productivity (e.g. due to greater availability of consumer durables)? Lastly, there has been some debate how exactly to integrate measures of leisure into standard measures of GDP and income[47].

131. Many of these problems raised by this discussion are little different from those discussed in earlier sections, for instance concerning home production. Index number problems are pervasive in all aspects of national income accounting, as are problems associated with imputations for non-market activities. Even for market activities, we simply *assume* competitive market valuations, knowing full well that these assumptions are not appropriate in many key sectors of the economy. Similarly, in the case of non-market (government) production, we made implicit assumptions about the relative pace of technological change (relative to that in market activity). Markedly different results would obtain under different assumptions. There are, to be sure, problems in the measurement of leisure – just as there are problems in the measurement of time spent in home production. When is cooking a leisure activity, and when is it a substitute for market production?

132. Our approach towards measuring the leisure starts from time use data. The value of leisure per hour is measured by its opportunity cost, i.e., the wage foregone because a person engages in leisure rather than in paid work. This section provides some illustrative

45. Keynes (1935).

46. Nordhaus and Tobin (1973).

47. The approach followed by Nordhaus and Tobin (1973) adds the value of leisure at constant prices to real GDP in both the initial and final year: as growth in leisure-time tends to be less than for real GDP (in the absence of productivity adjustments) the resulting leisure adjusted aggregate will typically increase by less than real GDP (the more so the larger the size of leisure relative to GDP). The approach suggested by Usher (1973) adds the change in leisure-time to real GDP in the final year: when leisure increases over time, growth of .leisure-adjusted. GDP will exceed that of real GDP, even in the absence of productivity adjustments.

calculations of the impact of leisure on our measurement system. These calculations are based on a number of assumptions. First, we only value the leisure of the working age population, assuming that the opportunity cost for other parts of the population is zero. Second, we ignore differences in the value of leisure for working age people with different characteristics (e.g. an hour of leisure for a person that is considered as involuntary unemployed is valued in the same way as for a person with a well-paid job working a long number of hours). Lastly, we value leisure based on the after-tax, average compensation of employees in each country. Multiplication of the average leisure time per day by the working age population and by the wage rate yields a measure of the total value of leisure.

133. There are two issues with this computation. The first is a result of lack of data: no distinction was made between employed, unemployed and inactive persons in their consumption of leisure time per day. Socio-economic breakdowns of time use surveys show that such differences exist. Employed persons tend to enjoy less leisure time than the average person and inactive persons spend more time on leisure. The second problem concerns the treatment of retired persons. Our method assumes zero opportunity cost for the population above 64 which is incorrect in all those cases where persons in this cohort could work if they so wished. Conversely, there are persons in this cohort who are active on the labor market and assigning a zero value to their leisure implies understating the total value of leisure. However, the empirical basis to discriminate between those retirees who are constrained in their choice of working and those who are not is weak.

134. A second shortcoming in the calculation is the failure to discriminate between full-time and part-time employees. If part-time employment is a choice, more leisure associated with part-time work should be valued in terms of the wage foregone. Conversely, if part-time employees are constrained in their choice, the additional leisure that these persons get should have a lower valuation than otherwise. But since individuals could always choose to work in home production, it is reasonable to assume that the value of leisure is at least equal to that imputed for home production – the opportunity cost of leisure.

135. Finally, it must not be forgotten that an important objective in valuing leisure is to make cross-country comparisons: a given real income in a society with more leisure will typically imply a higher living standard than in a society with the same income but less leisure. Whether or not the complications in the measurement of leisure play out in international comparisons, depends on the *relative* enjoyment of leisure between different groups – employed (and inactive) versus unemployed, part-time versus full-time employment etc. – and on the *relative* occurrence of imposed leisure. If these proportions vary significantly across countries, country comparisons will be affected. If the proportions are similar, cross-country comparisons will be affected to a lesser degree. But even “imposed” leisure may be valued. Societal regulations or union bargaining that require, say, so many weeks of vacation reflect societal preferences. There are large differences across countries in the number of vacation days. More research needs to be conducted to reduce uncertainty around these questions.

136. For the three countries at hand, the value of leisure adds between 20 and 30 % to net disposable income of households in nominal terms. More interesting than nominal income levels is the question how leisure affects *real* income and consumption trends across countries. Real measures are again obtained by applying a suitable price index to nominal measures. This price index now includes a component for the ‘price’ of leisure, i.e. wages

foregone on the labor market. Results for annual growth rates are presented in 11. In comparison with the measured rate of real income, the measure including leisure moves significantly more slowly. This is in line with earlier studies, where one finds that measures of income with a large non-market component (such as household production and leisure) tend to exhibit higher levels, slower growth rates and less volatility[48] than more narrowly defined, market-based measures of incomes and prices.

Table 11: Real income of households, corrected for household production and leisure
Percentage change at annual rate, 1995-2006

	France	United States	Finland
Real adjusted disposable income*	2.2%	3.1%	2.5%
Real adjusted disposable income*, corrected for household production and for leisure**	1.4%	2.3%	1.4%

*For private households and non-profit institutions serving households; net of depreciation.
**'Income from household production' equals the value of the estimed labour and capital inputs used
by households to produce own-account services (except owner occupied housing);
leisure has been valued with the average after-tax wage of the working age population;
income has been deflated with a price index for actual individual consumption, adjusted for houshold production and leisure.

Source: OECD estimates.

137. Cross-country comparisons of living standards that reflect both household production and leisure ('full income'- 12) are of interest because, *ceteris paribus*, societies with more leisure are better off than societies with less leisure: ignoring these differences means ignoring one of the main components of social progress. Inclusion of leisure and non-market production of household services into a broader measure of living standards paints quite a different picture of per capita income in the case of France relative to the United States – instead of 79%, the income level of France is now at 87%. The effect is, however, marginal in the case of Finland. This is not because Fins enjoy less leisure than Americans but because real wages per hour (our measure to value leisure) are lower in Finland than in the United States. 13, which compares the volume of leisure between France, the United States and Finland spells this out in more detail. The real value of leisure reflects several elements: real wage per hours worked, the leisure hours per working age person, the ratio between working age persons and the population as a whole. While there is about 10% more leisure per person consumed in Finland than in the United States, this is offset by the fact that Finnish real wages only amount to 59% of the U.S. level. One effect offsets the other, so that the real income levels including and excluding leisure are relatively similar.

48. Some of the lower volatility is simply methodological as imputations don't change; some is because, almost as a matter of definition, when market work is reduced, either non market production or leisure increases. This may understate true magnitude of volatility by ignoring the costs/consequences of capital market constraints, which are ignored in the analysis.

Table 12: Real income of households, including household production and leisure, 2005

		France	United States	Finland
Real adjusted disposable income* per capita	USD per capita	25378	32110	21771
	USA=100	79	100	68
Real adjusted disposable income* per capita, corrected for household production and leisure**	USD per capita	64708	74571	50021
	USA=100	87	100	67

*For private households and non-profit institutions serving households; net of depreciation, converted with PPPs for actual individual consumption

**'Income from household production' equals the value of the estimed labour and capital inputs used by households to produce own-account services (except owner occupied housing); converted with PPPs for actual individual consumption corrected for household production and leisure

Source: OECD estimates.

Table 13: Decomposition of the real value of leisure in international comparison, 2005

			France	USA	Finland
Value of leisure in real terms per capita		USA=100	91.9	100.0	65.3
	Real wage per hour	USA=100	89.2	100.0	59.5
	Leisure per (working age) pers	USA=100	106.1	100.0	110.5
	Working age/total population	USA=100	97.1	100.0	99.4
	Working age/total population	%	65.1	67.1	66.7

Source: OECD estimates.

138. The imprecision associated with the above estimates is again large. These are orders of magnitude and should not be over-interpreted. However, it is also clear that recognition of leisure as one element for living standards does 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.

139. Finally, we note that Krueger et al. (2008) and other researchers propose a different way of considering the effects of both the consumption of goods and service and of leisure. They develop a system of national time accounting where information on how people use time is combined with information on the emotional experience during these activities. This approach holds open the promise of being able to distinguish among household production activities which more rightly might be called "leisure" (cooking as an art form) and those which might more appropriately called production.[49] Certain activities concern household production, and others concern leisure. The broader national time accounting index is a way of bringing these elements together in a single measure.

4.10. Distribution of full income

140. The rationale for examining distribution holds not only for market income but also for broader measures such as full income. The recognition of own-account production of household services may not only affect the aggregate measures of income and production but also, in all probability, the established picture on its distributions a point made by Eisner

49. Note that in market production, no distinction is made between jobs that are enjoyable (and generate non-pecuniary benefits) and those that are not. Some of these differences are reflected in wages. If a change in working conditions that made a job less enjoyable were reflected in a lower wage, national income would go down.

(1988) twenty years ago[50]. A more recent survey of the literature on this topic (Folbre, forthcoming) confirms the importance of considering household work when assessing distributions of income and consumption.

141. Developing distributional measures of full income is, however, a formidable task. The single most difficult item is to allocate to various groups those income flows that have been imputed at the macro level. 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. All these measures are typically absent from micro-estimates of income.[51] Accardo et al. (2009) is a recent example of such a work in France, but limited to household disposable income. Household production and leisure are not taken into account.

142. One could go further and ask, if other dimensions than income are deemed to be important in determining the well-being of an individual or a household, whether it is possible to take them into account when trying to derive a more comprehensive picture of the distribution of living standards in society. This would be easy if some aggregate scalar representing most dimensions of individual well-being were available. But it is unlikely that such aggregation would be possible without arbitrary assumptions on imputed prices. Moreover, it is also unlikely that all the information to compute such an aggregate indicator at the individual level would be available in standard micro-data bases. Does this mean that considerations for distributional aspects of well-being should be limited to a few simple economic variables like income or consumption expenditures?

143. A multi-dimensional approach may help here by taking into account distributional aspects of various dimensions of well-being. One application of this approach would consist in crossing data in these other dimensions of well-being with the ranking of individuals or households in the distribution of (real)[52] money incomes, i.e. considering that money income is the main determinant of people's well-being. For instance, introducing distributional aspects into a non-monetary dimension like the consumption of public goods would simply require observing how much is consumed by people around the median of the money income distribution as opposed to consumption per capita. The same could be done with the consumption of leisure, own-account production, wealth or change in wealth, and more generally all individual or household characteristics in micro-data bases where information is available to determine people at the median or other quantiles of money income - including surveys on subjective matters like satisfaction and happiness. This would constitute an improvement over present practices of focussing on averages.

50. "Inclusion of non-market output may also bring significant changes to measures of income distribution, by size, by sex, by age, and by rural-urban status. It is reasonable to conjecture that the proportions of income related to non-market output in the home are larger among the poor and among women, the aged, and those on farms or in rural areas. Full imputation of the value of education would appear likely to bring further dramatic changes in our views of the distribution of saving and investment as well as of income." (Eisner 1988, p. 1613).

51. Income from owner-occupied housing is now being imputed in a number of European surveys.

52. Note that the definition of the median individual or household in terms of *real* money income raises the issue of the price deflator to use. Presumably, income-dependent price deflators have to be used. They differ across income levels because the weights of the various consumption goods vary with the level of income.

Main messages and recommendations

Recommendation 1: Look at income and consumption rather than production.

144. 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 mainly measures market production, though it has often been treated as if it were a measure of economic well-being. Conflating the two can lead to misleading indications about how well-off people are and entail the wrong policy decisions. Material living standards are more closely associated with measures of real income and consumption – production can expand while income decreases or vice versa when account is taken of depreciation, income flows into and out of a country, and differences between the prices of output and the prices of consumer products.

Recommendation 2: Consider income and consumption jointly with wealth.

145. Income and consumption are crucial for assessing living standards, but in the end they can only be truly gauged in conjunction with information on wealth. A vital 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 need to have comprehensive accounts of its assets (physical capital – and probably even human, natural and social capital) and its liabilities (what is owed to other countries). Balance sheets for countries are not novel in concept, but their availability is still limited and their construction should be promoted. There is also a need to "stress test" balance sheets with alternative valuations when market prices for assets are not available or are subject to bubbles and bursts. Note that measures of wealth are also central to measuring sustainability. What is carried over into the future necessarily has to be expressed as stocks – of physical, natural or human capital. Here too the right valuation of these stocks plays a crucial role.

Recommendation 3: Emphasise the household perspective.

146. While it is informative to track the performance of economies as a whole, trends in citizens' current material living standards are better followed through measures of household income and consumption. Indeed, the available national accounts data shows that in a number of OECD countries real household income has grown quite differently from real GDP, and typically at a lower rate. The household perspective entails taking account of payments between sectors, such as taxes going to government, social benefits coming from government, and interest payments on household loans going to financial corporations. Properly defined, household income and consumption should also reflect the value of in-kind services provided by government, such as subsidized health care and educational services.

Recommendation 4: Give more prominence to the distribution of income, consumption and wealth.

147. Average income, consumption and wealth are meaningful statistics, but they do not tell the whole story about living standards. For example, a rise in average income could be unequal across income groups, leaving some households relatively worse-off than others. Thus, average measures of income, consumption and wealth should be accompanied by indicators that reflect their distribution across persons or households. Ideally, such

information should not come in isolation but be linked, i.e. one would like information about how well-off households are simultaneously with regard to all three dimensions of material living standards: income, consumption and wealth. After all, a low-income household with above-average wealth is not necessarily worse-off than a medium-income household with no wealth. The desirability of disposing of information on the "joint distribution" of dimensions will be encountered once again in Recommendation 3 of the Chapter on the quality of life.

Recommendation 5: Broaden income measures to non-market activities.

148. There have been changes in how households and society function. For example, many of the services people received from other family members in the past are now purchased on the market. This shift translates into a rise in income as measured in the national accounts and may give a false impression of a change in living standards, while it merely reflects a shift from non-market to market provision of services. Many services that households produce for themselves are not recognized in official income and production measures, yet they constitute an important aspect of economic activity. While their exclusion from official measures reflects uncertainty about data more than it does conceptual dissent, more and more systematic work in this area should be undertaken. This should start with information on how people spend their time that is comparable both over the years and across countries. Comprehensive and periodic accounts of household activity as satellites to the core national accounts should complement the picture.

REFERENCES

Abraham, K. G. and Ch. Mackie (eds.) (2005); Beyond the Market: Designing Nonmarket Accounts for the United States; Panel to Study the Design of Nonmarket Accounts; The National Academies Press, Washington D.C.; http://www.nap.edu/openbook.php?record_id=11181

Accardo J., Balmy V., Consalès G., Fesseau M., Le Laidier S., Raynaud É. (2009), "Les inégalités entre ménages dans les comptes nationaux. Une décomposition du compte des ménages.", INSEE - L'Économie française

Atkinson Review (2005) *Measurement of Government Output and Productivity for the National Accounts.* Final Report, Palgrave Macmillan.

Atkinson A. B. (1970) "On the measurement of inequality"; *Journal of Economic Theory*; 2; pp. 244-263.

Atkinson A. B. and J. E. Stiglitz (1980), *Lectures on Public Economics*, McGraw-Hill, New York.

Atkinson, A. B. and Voitchovsky, S. (2008); *The Distribution of Top Earnings in the UK since the Second World War*, University of Oxford.

Becker, G. S. (1965); "A Theory of the Allocation of Time"; *The Economic Journal*, Vol. 75, No. 299 pp. 493-517.

Boarini, R., Å. Johansson and M. Mira d'Ercole (2006), "Alternative Measures of Well-Being", *OECD Economics Department Working Papers*, No. 476

Canberra Group (2001); *Final Report and Recommendations of the Expert Group on Household Income Statistics*; Ottawa.

Commission 'Mesure du Pouvoir d'Achat des Ménages' (2008) ; *Rapport remis à la Ministre de l'économie, des finances et de l'emploi*, Paris ; available under http:// lesrapports.ladocumentationfrancaise.fr/BRP/084000066/0000.pdf

Cowell, F. A. (2000); "Measurement of Inequality"; in A. B. Atkinson and F. Bourgignon (eds.), *Handbook of Income Distribution*, Volume 1, Elsevier, Amsterdam, pp. 87-166.

Cutler, D., A. Deaton and A. Lleras-Muney (2006); "The Determinants of Mortality"; *Journal of Economic Perspectives;* Volume 20, Number 3; Summer 2006; pp. 97–120.

Deaton, A. S. (2005); "Measuring poverty in a growing world" (or "Measuring growth in a poor world"), *Review of Economics and Statistics*, vol 87: pp.1-19.

Deaton, A. S. (1998); "Getting Prices Right: What Should be Done?"; *The Journal of Economic Perspectives* Vol. 12, No. 1 (Winter, 1998), pp. 37-46.

Deveci, N., K. Heurlén and H. S. Sørensen (2008); "Non-Market Health Care Service in Denmark – Empirical Studies of A, B and C Methods"; paper presented at the meeting of the *International Association for Research on Income and Wealth*, Slovenia.

Diewert, W. E. (2008), "Changes in the Terms of Trade and Canada's Productivity Performance", *Discussion Paper 08-05, Department of Economics*, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1, (March), 114 pp.

Diewert, W. E. (2001); "The Consumer Price Index and index number purpose"; *Journal of Economic and Social Measurement* 27; pp. 167-248.

Diewert W. E. (1998); "Index Number Issues in the Consumer Price Index"; *The Journal of Economic Perspectives*, Vol. 12, No. 1 (Winter, 1998), pp. 47-58

Diewert, W. E. (1987); "Index Numbers", in J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*.

Diewert, W. E. (1976), "Exact and Superlative Index Numbers"; *Journal of Econometrics* 4, 115-145.

Diewert, W. E. and A. Nakamura (2009); "Accounting for Housing in a CPI"; *Discussion Paper 09-08, Department of Economics*; University of British Columbia; available under http:// www.econ.ubc.ca/diewert/dp0908.pdf

Diewert, W. E. and D. Lawrence (2006), *Measuring the Contributions of Productivity and Terms of Trade to Australia's Economic Welfare*, Report by Meyrick and Associates to the Australian Government, Productivity Commission, Canberra, Australia, 114 pp.

Diewert, W. E. and Morrison, C. J. (1986). "Adjusting output and productivity indexes for changes in the terms of trade"; *Economic Journal* 96, pp. 659–679.

Diewert, W. E., H. Mizobuchi and K. Nomura, (2005), "On Measuring Japan's Productivity, 1955-2003", *Discussion Paper 05-22, Department of Economics*, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.

Eisner, R. (1988); "Extended Accounts for National Income and Product"; *Journal of Economic Literature*, Vol. XXVI, December, pp. 1611-84.

Folbre, N. (forthcoming); "Time Use and Inequality in the Household"; *Oxford Handbook of Economic Inequality*.

Folbre, N. and Y. Jayoung (2008). *The Value of Unpaid Child Care in the U.S. in 2003.* Forthcoming in *How Do We Spend Our Time? Recent Evidence from the American Time Use Survey*, ed. Jean Kimmel. Kalamazoo, Michigan: W. E. Upjohn Institute for Employment Research.

Folbre, N. and B. Wagman (1993); "Counting Housework: New Estimates of Real Product in the U.S., 1800-1860", *The Journal of Economic History*; 53:2 (1993), 275-88.

Fraumeni, B. (2008); "Household Production Accounts for Canada, Mexico, and the United States: Methodological Issues, Results, and Recommendations"; paper presented to the 30th *General Conference of the International Association for Research in Income and Wealth*; Potoroz Slovenia.

Freeman, R. and R. Schettkat (2005); "Marketization of Household Production and the EU-US Gap in Work"; *Economic Policy*, Vol.20, No.41, pp. 6-50, January 2005.

Hill, P. (2009, forthcoming); "Consumption of Own Production and Cost of Living Indices"; in: W.E. Diewert, B. M. Balk, D. Fixler, K. J. Fox and A. O. Nakamura (editors); *Price and Productivity Measurement*; Trafford Press.

Hicks, J. R. (1939); "Public finance in the national income"; *The Review of Economic Studies, Vol. VI, N°2, p.150*

IMF, OECD, UN-ECE, World Bank (2009); *Export and Import Price Index Manual*,Washington, D.C.;

ILO, IMF OECD, UN-ECE, World Bank (2004); *Consumer Price Index Manual, Theory and Practice*, Washington, D.C.;

IMF, ILO, OECD, UN-ECE, World Bank (2004); *Producer Price Index Manual, Theory and Practice*, Washington, D.C.;

Jorgenson, D. W. (1990); "Consumer Behaviour and the Measurement of Social Welfare"; *Econometrica*, Vol. 58; No. 5; pp. 1007-1040.

Jorgenson, D. W. and B. M. Fraumeni. (1989). "The Accumulation of Human and Non-Human Capital, 1948-1984," in *The Measurement of Saving, Investment, and Wealth*, R.E. Lipsey and H. Tice, eds. Chicago: University of Chicago Press, pp. 227-282.

_____ (1992) "The Output of the Education Sector," in *Output Measurement in the Service Sectors*, Z. Griliches, eds. Chicago: University of Chicago Press, pp. 303-338.

Keynes, J. M. (1935): "Economic Possibilities for our Grandchildren"; in J. M. Keynes, *Essays in Persuasion*; London: Macmillan.

Keynes, J. M. (1940); *How to pay for the war*; New York.

Kohli, U. (1991); *Technology, Duality, and Foreign Trade: The GNP Function Approach to Modeling Imports and Exports* (London: Harvester Wheatsheaf and Ann Arbor, MI: University of Michigan Press).

Kolm, S. C. (1969); "The Optimal Production of Social Justice"; in: J. Margolis and H. Guitton (eds.) *Public Economics*; London.

Krueger, A. B., D. Kahneman, D. Schkade, N. Schwarz and A. A. Stone (2008); "National Time Accounting: The Currency of Life"; paper presented at the first meeting of the Commission on the Measurement of Economic Performance and Social Progress, Paris 22-23 April 2008.

Kuznets, S. (1951); "Government product and national income"; *Income and wealth, Series 1*

Landefeld, J. S., B. M. Fraumeni, and C. M. Vojtech (2009); "Accounting for Nonmarket Production: A Prototype Satellite Account Using the American Time Use Survey", *Review of Income and Wealth*; series 55; pp. 205-225.

Landefeld, J. S. and S. H. McCulla. 2000. "Accounting for Nonmarket Household Production within a National Accounts Framework," *Review of Income and Wealth*, Series 46, No. 3 (September), pp. 289-307.

Marionnet, D. (2007); "The final financial investment of French households", *Banque de France*, March.

Meade, J. and R. Stone (1941); "The Construction of Tables of National Income, Expenditure, Savings and Investment"; *The Economic Journal*; pp. 216-233.

Nordhaus W. and Tobin J. (1973); "Is Growth Obsolete?" in: *The Measurement of Economic and Social Performance, Studies in Income and Wealth*, National Bureau of Economic Research, vol.38.

OECD (2009); *Society at a Glance*; Paris.

OECD (2008a); *Growing Unequal? Income Distribution and Poverty in OECD Countries*; Paris.

OECD (2008b); *Towards measuring the volume of health and education services: Draft handbook;* available under http://www.oecd.org/document/52/0,3343,en_2649_34245_40617524_1_1_1_1,00.html

Ruger, Y. and J. Varjonen (2008); *Value of Household Production in Finland and Germany*. Working Paper 112, National Consumer Research Centre, Finland.

Ruiz, N. (2009); "La question du pouvoir d'achat"; *Cahiers Français* no 347.

Sen, A. (1985), *Commodities and Capabilities*, North-Holland, Amsterdam.

Sen, A. (1979); "The Welfare Basis of Real Income Comparisons"; *Journal of Economic Literature*, Vol XVII;

Sen, A. (1976); "Real National Income"; *The Review of Economic Studies*, Vol. 43, No. 1 (Feb., 1976), pp. 19-39.

Schreyer, P. and W. E. Diewert (2009); “Household production, leisure and living standards”; unpublished manuscript.

Schultze, Ch. L. and Ch. D. Mackie (eds.) (2002); *At What Price?; Conceptualizing and Measuring Cost-of-Living and Price Indexes*; National Academy Press, Washington D.C..

Stiglitz, J. (forthcoming); “Towards a General Theory of Consumerism: Reflections on Keynes' Economic Possibilities for Our Grandchildren”, in *Revisiting Keynes: Economic Possibilities for Our Grandchildren*, G. Piga and L. Pecchi, ed., Cambridge : MIT Press.

Triplett, J.E. (2006); *Handbook on Hedonic Indexes and Quality Adjustment in Price Indexes: Special Application to Information Technology Products*; OECD, Paris.

Triplett, J. E. (2001); “Should the cost-of-living index provide the framework for a consumer price index?”; *Economic Journal* 111(June); pp. 311-334.

UN-ECE, OECD and Eurostat (2009); *Measuring Sustainable Development*; United Nations; New York and Geneva.

Usher, D. (1973); comments on Nordhaus and Tobin (1973).

Vanoli, A. (2002); *Une histoire de la comptabilité nationale* ; Paris, La Découverte.

Wagman B. and N. Folbre (1996), “Household Services and Economic Growth in the U.S., 1870-1930,” *Feminist Economics* 2:1 (1996), 43-66.

Weitzman, M. L. (1976); “On the Welfare Significance of National Product in a Dynamic Economy.” *The Quarterly Journal of Economics* 90: 156-162.

Yitshaki, S. (1979); “Relative Deprivation and the Gini Coefficient”; *The Quarterly Journal of Economics*, Vol. 93; 2; pp. 321-324.

CHAPTER 2 - QUALITY OF LIFE

Introduction

The traditional approach followed by economists to measure human well-being focuses on the resources that individuals have at their command, which are usually assessed in terms of either money income or assets or the goods and services that they consume. However, although resources are clearly important for human well-being (they are discussed in Chapter 1 of this report), they are also clearly insufficient, for reasons that are detailed in Box 2.1. Human well-being depends on what resources enable people to do and to be, and the ability to convert resources into a good life varies across people. This suggests that indicators that go beyond being measures of income, wealth and consumption and incorporate the non-monetary aspects of quality of life (hereafter referred to as QoL) have an important role to play. The variety of these measures, and the lack of an obvious metric to compare developments in the various dimensions, constitutes both the main advantage and the main limit of these indicators.

What constitutes a "good life" has occupied leading philosophers since Aristotle, and dozens of definitions of the "good life" are discussed in the literature: none of these definitions commands universal agreement, and each corresponds to a different philosophical perspective. The Commission's goal in this area has not been to reach a shared understanding of what "quality of life" means but, rather, to identify areas where credible measures, based on clear conventions and definitions, could be established.[1] As a result, this report does not aim to be comprehensive: attention is limited to areas where members of the Commission had specific competencies and where available indicators allow drawing a more comprehensive assessment of QoL.

Box 2.1. Is command over resources an adequate metric to assess human well-being?

The answer to this simple question is "no". Indeed, all the approaches to quality of life described in this chapter share the view that people's *opulence* (*i.e.* the quantity and features of the commodities that they have at their disposal) does not offer an adequate perspective on human well-being. The standard economic argument that changes in (suitable deflated) measures of real income or wealth lead to changes in the same direction of consumer satisfaction provides only a weak connection between resources and human well-being: it says *something* about the direction of change, but it does not inform about its magnitude or about the level of well-being of individuals with different preferences. The standard argument that people's command over resources is the relevant metric for determining how well-off people are similarly ignores that people with different characteristics will have different capacities to transform income or wealth into actual well-being, and that these differences cannot be ignored.

1. The term "quality of life" is used in this chapter to refer to those aspects of life that shape human well-being beyond the command of economic resources.

Quality-of-life approaches regard income or wealth, even when extended through imputations of additional items, as an inadequate proxy of human well-being for a range of reasons:

- First, many resources are unequally accessible to individuals, either because they are not marketed or because they come with different prices. Even when pricing is possible, the prices will differ across individuals, which makes it problematic to compare real income across people (for example, when valuing leisure for individuals with unequal wage rates).
- Second, many of the determinants of human well-being are not monetary resources but aspects of people's life circumstances (health, social networks, quality of institutions) or activities (home activities, quality of work, leisure). It would be far-fetched to describe them as resources with imputable prices, even if individuals do make trade-offs among them.
- Third, most approaches to human well-being reject the idea that command over resources is a suitable metric even in the absence of the two previous points. Resources are means that are transformed into well-being in ways that differ across people: people with greater capacities for enjoyment or greater abilities for achievement in valuable domains of life are better-off even if they command fewer economic resources.

These arguments suggest that conventional, market-based measures of income, wealth and consumption are insufficient to assess human well-being. They need to be complemented by non-monetary indicators of QoL.

In guiding its work, the starting point of the Commission has been an understanding of the most important features that give life its value. QoL is often tied to the opportunities available to people, to the meaning and purpose they attach to their lives, and to the extent to which they enjoy the possibilities available to them. QoL research has identified a rich array of attributes – such as belonging, fulfilment, self-image, autonomy, feelings, and the attitudes of others – that are associated with QoL. Some of these attributes are intangible and difficult to evaluate. Others, however, have a more tangible character, and can be measured in reasonably valid and reliable ways. In all cases, measuring QoL requires consideration of a multidimensional array of indicators, which lack a unique metric that would allow simple aggregation across dimensions: alternative approaches to the aggregation of this multidimensional array of indicators are discussed at the end of this chapter.

In thinking about how best to measure QoL, it is necessary to choose some methodological criteria. This chapter relies on a number of them.

- First is the emphasis on people, on what they value as important for their daily life, and on the environment in which they develop. Taking the individual as the fundamental unit of analysis does not imply neglecting communities and institutions, but requires evaluating them in virtue of what they bring to the QoL of the people participating in them. This perspective also implies focusing on the "ends" of various human activities, while recognizing that their achievement can matter both intrinsically and instrumentally (*i.e.* to achieve some other goal).

- Second is a recognition of the diversities and inequalities in human conditions. This diversity implies that societal well-being depends on both the aggregate level of the various items that shape people's lives and on how they are distributed in society. The relative importance attributed to these two aspects will depend on alternative conceptions of social justice.

- Third is the view that quality of life depends on a variety of factors, with none of them claiming absolute priority over the others. The multi-dimensional nature of quality of

life (as opposed to the scalar nature of income) increases the complexity of the analyses and raises a number of measurement issues. These pertain to whether to use comparable dimensions across people and countries, how to assess the importance of each dimension, which indicators to use to describe accomplishments in the various dimensions, how to present these indicators (e.g. in their raw form, or by standardizing them in various ways), and whether and how to aggregate them.

- Last is a focus on the present, rather than on the QoL of future generations; while the sustainability of QoL is an important issue, the factors that bear on this assessment are described in Chapter 3.

The next section describes the main approaches to QoL that have the potential to deliver concrete measures. The third section then reviews a number of objective domains that are both broadly relevant for QoL and where credible measures should be developed through appropriate tools and adequate investment. The last section describes three cross-cutting themes in the measurement of QoL, *i.e.* assessing the links between various domains, accounting for inequalities in individual experiences in each QoL domain, and providing a more parsimonious description of QoL in each society by aggregating its various components.

Approaches to measuring quality of life

Three main conceptual approaches have proved useful in thinking about the measurement of QoL: the first is based on the notion of subjective well-being, the second on the notion of capabilities, and the third on economic notions drawn from welfare economics and from the theory of fair allocations. Each of these approaches informs different measurement strategies. While these approaches may represent opposite poles in intellectual terms, they all have a role to play in measuring QoL.

Subjective well-being

A long philosophical tradition views individuals as the best judges of their own conditions. In economics, this approach is closely linked to the utilitarian tradition, which argues that QoL is reflected exclusively in the subjective states of each person. An approach based on subjective self-reporting has broad appeal due to the strong presumption in many streams of ancient and modern culture worldwide that making people "happy" and "satisfied" with their lives is a universal goal of human existence. Only a few years ago, the idea of measuring people subjective states would have appeared incongruous. Today, several methods have made subjective well-being amenable to systematic quantification (Kahneman, Diener and Schwartz, 1999). The greatest strength of this approach is its simplicity: relying on people's own judgments is a convenient shortcut and potentially provides a natural way to aggregate various experiences in a way that reflects people's own preferences. Further, this approach makes it possible to reflect the diversity of people's views about what is important in their lives.

Distinct aspects of subjective well-being

Subjective measures of well-being have been very prominent in recent discussions about measuring QoL, but this popularity has also generated some ambiguities and misconceptions. Most prominent among these is the assumption that all dimensions of subjective well-being

can somehow be reduced to the single concept of "happiness". In reality, as argued by Diener (1984), subjective well-being (SWB) is best understood as encompassing three separate aspects:

- life satisfaction, *i.e.* a person's overall judgment about their life at a particular point in time;
- the presence of positive feelings or affect, *i.e.* the flow of positive emotions (such as feeling happiness and joy, or a sense of vitality and energy) from moment to moment;
- the absence of negative feelings or affect, *i.e.* the flow of negative emotions (such as feeling angry, sad or depressed) from moment to moment.

Life satisfaction, positive affect and negative affect are separate aspects of SWB, and accord with different conceptions of QoL. Satisfaction with life as a whole (and with particular domains, such as work, housing and family life) involves an *evaluative judgment* of how one's life is faring, which requires making an effort and remembering past experiences. Conversely, positive and negative affect involve measuring the *hedonic experiences* experienced by people in real time, or shortly after these experiences have occurred.

These three aspects of subjective well-being are distinct. People who experience unpleasant feelings or physical pain could still be very satisfied with their lives if they value the contributions that they feel they are making to society or to some other personal goal. Likewise, the presence of positive affect does not imply the absence of negative affect. Across individuals, the correlation between life satisfaction and positive affect is only around 0.40; even after adjusting for day-to-day variability in the reports of life satisfaction and affect, the correlation is still below 0.60 (Krueger and Schkade, 2008). The correlation among different indicators of negative affect, such as anger and sadness, is also low at the individual level.

Which of these dimensions of SWB matters more – and for what – is still an open question. Much evidence suggests that people act to achieve satisfaction in their choices, and that these choices are based on memories and evaluations. These memories and evaluations, however, can also lead to systematic errors and to choices that, in many fields, don't improve people's QoL. In any case, it would be problematic to neglect people's momentary feelings when considering their behaviour: some choices may be made unconsciously, rather than by weighing the pros and cons of all available alternatives, and relying on the current experiences of other people may sometimes lead to choices that are better at promoting one's own well-being than relying on predictions of our own emotional futures (Gilbert, 2005). The issue of which of these aspects of SWB has the bigger effect on health is similarly unresolved, with some findings suggesting that the presence of positive affect is a more significant determinant of health than the absence of negative affect (with the exception of depression, Janicki-Deverts *et al.*, 2007; Cohen *et al.*, 2006).

Measurement approaches

Each of these aspects of subjective well-being needs to be measured in the most appropriate way, as different survey questions can lead to different numerical results:

- Data on life-evaluations have been collected in several representative surveys. In several cases (e.g. various waves of the *World Values Survey*), these measures are based on qualitative responses, such as feeling "quite" or "fairly" happy with one's life, or on

other scales that assess life satisfaction. However, results based on qualitative response categories are potentially affected by biases that limit cross-country comparability. Conversely, the use of a visual (ladder-of-life) scale, with explicit reference points (10, for the best possible life, and 0 for the worst possible life) has proved more effective in eliciting a cognitive evaluation from respondents that is less vulnerable to comparability problems. Even this formulation, however, does not assure the full comparability of answers, as reference points may differ across time and people (Deaton, 2008).

- Hedonic experiences are measured by a report that is made by each person either in real time or shortly after an event has occurred. These measures have been collected less frequently than life evaluations. Indeed, the gold standard method for collecting data on hedonic experiences in real time (*Experience Sampling*) has never been applied to a representative population sample because it is burdensome.[2] However, alternative methods of collecting data on hedonic experiences, such as a telephone version of the *Day Reconstruction Method*, are less costly to implement, and investments should be made to implement these approaches with representative samples.[3] It is also important that multiple dimensions of affect – such as feeling happy, sad, angry, tired or in pain – be separately measured because they are distinct emotions, and that these measures be collected periodically, so as to allow assessing changes over time.

Some survey data that capture these three aspects of SWB already exist. For example, the *Gallup World Poll* is a nationally-representative survey fielded in around 140 countries that aims to better assess people's experiences and well-being. Questions about life-evaluations are based on the (0 to 10) ladder-of-life scale, and the survey also includes questions on feelings, both positive and negative, experienced the previous day. Across OECD countries, the correlation between mean life-evaluations and the mean prevalence of positive affect (computed first by counting the number of positive experiences of each respondent and then by averaging these scores at the country level) is positive (0.67), while that between the mean prevalence of positive affect and of negative affect is negative (-0.26). However, in both cases, there is much variation across countries: in other words, the fact that most people in a country report a high life-satisfaction does not imply a high prevalence of positive affect, while a high prevalence of positive affect can go in hand with a high prevalence of negative affect.

As research progresses, it is likely that other aspects of SWB will be developed and measured. In this sense, both life satisfaction and hedonic experiences provide an incomplete

2. The term "experience sampling" refers to techniques involving the recording of participants' experiences during their daily lives; the term is sometimes used more broadly to designate any procedure involving: the assessment of experiences in natural settings; measurements that take place soon after the occurrence of the experience reported; and measures on repeated time occasions. One concrete application of this method uses reports made by participants in response to a random signal (*e.g.* emitted by a pager) or at pre-determined times during the day.

3. The Day Reconstruction Method asks respondents to reconstruct the previous day by completing a structured self-administered questionnaire. Respondents first reconstruct the previous day by producing a short diary consisting of a sequence of episodes; next, they draw on their confidential diary to answer a series of questions describing each episode (*i.e.* when it began and ended, what they were doing, where they were, whom they were interacting with, and how they felt on multiple affect dimensions). Respondents also answer a number of questions about themselves and the circumstances of their lives (*e.g.*, demographics, job characteristics, personality measures).

description of SWB. Nonetheless, they contain information that is not reflected in other common indicators, such as income. And they highlight patterns across people living in each country that differ sharply from those based on income measures: for example, in most developed countries younger and older people report higher levels of life-satisfaction than prime-age individuals (Oswald, 1997), a pattern that contrasts sharply with the levels of income of the same groups (which rise until prime age and then decline following retirement). This suggests that these measures can play a useful role in measuring the QoL of people and groups as complements of other indicators.

Different aspects with distinct determinants

One of the most attractive promises of research on subjective well-being is to deliver not just a good measure of the level of QoL, but also a better understanding of its determinants, as affected by a variety of objective features (such as income, health status and education).

The set of determinants that is most relevant will depend on which aspect of subjective well-being is considered. For example, across individuals, indicators of life circumstances such as household income and marital status are more strongly related to life satisfaction than to positive or negative affect, while features of daily experiences, such as time pressure at work, are more strongly correlated with affect than with (work) satisfaction. The same conclusion applies across countries. For example, in a sample of over 130 countries surveyed by the Gallup Organization, the correlation between people's income and their individual ladder-of-life scores is 0.46; a similar correlation is found when looking at the national average ladder score and the GDP per capita of each country. In contrast, the correlation between income and reports of affect is much lower, both across people within each country and between countries. The effects of world regions (defined by culture as well as geography) on the ladder-of-life scores are mediated by GDP, but culture also has a direct effect on expressions of affect. Relative to their GDP, the former communist countries are quite joyless, as are the Islamic countries; conversely, Latin American countries feature high levels of affect – both positive and negative. These results also indicate that the amount of time that people spend in contact with others is a prime determinant of the positive affect they report (see below).

When considering the determinants of subjective well-being, one critical feature that matters for the interpretation of these measures is *adaptation*. According to one view, each person has a genetically-established personality, with given "set points" for various aspects of subjective well-being; in this view, changes in external circumstances may lead to temporary changes in subjective well-being, but subsequent adaptation will always force subjective well-being back to each individual's set point. While this reading of human nature would suggest that there is very little that policies could do to ameliorate QoL for the human lot, most proponents of this set-point view now seem to be leaning towards the view that the adaptation is less than complete. Even partial adaptation, however, has implications for subjective measures of well-being, as the "hedonic treadmill" makes affect and evaluations somewhat immune to objective life conditions. This, in itself, is not a weakness of the measure and could well reflect a fundamental feature of human nature (*i.e.* adaptability and resilience). However, it also suggests that measures based on subjective well-being may not be sufficient for all social evaluations, especially when the desires and expectations of those who are persistently deprived adjust to what they perceive as feasible: it is possible to be less satisfied with a better life, if one has a more ambitious standard; similarly, when standards

evolve in the same way as realizations, satisfaction levels at any given rank of society will remain roughly stable no matter what happens to the objective features that shape QoL.[4]

Another determinant of subjective well-being that has generated much research and controversy relates to the importance of *peer effects and relative comparisons*. Discussions in this field have been driven by the observation, by Easterlin that long-term gains in income and material prosperity were not followed by similar gains in life-evaluations. One way to explain this paradox is the hypothesis that income gains relative to other people within a community matter more for these life-evaluations than country-wide improvements in absolute income. Better measures of SWB have helped to dispel some of these old controversies. An important result of recent research is that the Easterlin paradox does not apply to cross-country comparisons of life evaluation at the world level (Stevenson and Wolfers, 2008; Deaton, 2008). In other words: i) countries with a higher *level* of GDP per capita do report higher life-evaluations; ii) the relationship between life-evaluations and the logarithm of GDP is broadly linear (*i.e.* it does not flatten out at higher income levels beyond the flattening that is implicit in the log-linear relation between the two variables); and iii) the relation between country-level GDP and average life-evaluations is similar to the one that applies to individuals' income and their own life-evaluations in each country.

Similarly, in a set of 18 countries, Diener and Kahneman (unpublished work) found that the correlation between *changes* in GDP and *changes* in the ladder-of-life scores about 30 years apart was 0.58; and, further, that the regression of national ladder scores on GDP was almost exactly the same at the two points in time. This suggests that the standard for "the best possible life" is not only shared across countries but is also remarkably stable over time. Other researchers, however, report evidence that relative income matters, based on different measures of subjective well-being, different surveys, and different selections of countries (Luttmer, 2005; Clark, Frijters and Shields, 2008; Helliwell, 2008; Layard *et al.*, 2008; Frank, 2008). Further, the Easterlin paradox may still apply to affect, implying that as countries get richer people will not necessarily report lower prevalence of negative feelings, or a higher prevalence of positive ones. More generally, while the debate on the ways in which income influences subjective well-being is likely to continue as better data became available, the results provided by Diener (1984), Di Tella *et al.* (2003) and Wolfers (2003) highlight that a range of factors beyond income shape life-evaluations: these factors include both personal characteristics (age, gender, marital status, perceptions of corruption and availability of social support) and macro-economic and contextual effects (as measured by national averages of variables already included at the individual level). Yet many puzzles remain, including the fact that, at least in some contexts, the health status of respondents does not appear to affect their life satisfaction (Deaton, 2008; Deaton, Fortson and Tortora, 2009). In other terms, survey measures of subjective well-being do not always appear to behave in ways that we would expect them to in order to trust them at face value.

One aspect where all research on subjective well-being does agree concerns the high human costs associated with unemployment. Unemployed people report lower life-evaluations than do the employed even after controlling for their lower income, a pattern that

4. "Combined with other evidence on the validity of life satisfaction measures, the widespread evidence on adaptation suggests to us that subjective well-being is not mainly a matter of income and consumption opportunities." (Kahneman and Krueger, 2006). This conclusion makes it harder for life satisfaction to be a consensual indicator of QoL, as many people resist the view that mental comfort is all that matters in life.

holds both when looking at cross-sectional data (Clark and Oswald, 1994; Blanchflower, 2008) and when following the same person over time (Winkelman and Winkelman, 1998); this pattern suggests the existence of non-pecuniary costs of unemployment, such as loss of friendship, meaning and status. There is also evidence that high unemployment also negatively affects people who retain their jobs, and that the gap in life-satisfaction between the unemployed and the employed has increased slightly in European countries (Di Tella et al, 2003). Evidence for a large sample of Europeans and non-Europeans also shows that the unemployed report feeling more sadness, stress and pain than the employed (Krueger and Mueller, 2008), with sadness being particularly high when searching for jobs and watching television (an activity that is conventionally classified as "leisure" in time-use surveys). Some research also finds that unemployment is more costly than inflation in terms of the impact on subjective well-being (Blanchflower, 2008): while there is much debate on the robustness of some of these findings, they do suggests *prima facie* that the cost of recessions (in terms of higher unemployment) could exceed those implied by more conventional macro-economic measures, such as the "misery index" (*i.e.* the sum of the unemployment and inflation rates).

A general difficulty for the study of the determinants of subjective well-being is to distinguish between *causes* and *correlates*. For example, the fact that the unemployed report lower life-evaluations as a whole may be because the experience of being unemployed lowers these evaluations (in which case, unemployment causes life-evaluations); or that people reporting lower life-evaluations are more likely to quit their jobs or to remain unemployed for longer (in which case, cognitive evaluations cause unemployment); or that people more affected by neurosis (a personality trait) are unstable in their jobs and report lower life-evaluations than others (in which case, a third factor shapes both cognitive evaluations and the experience of unemployment). The same difficulty applies when studying the determinants of a range of other features of QoL (e.g. health) discussed in this chapter.

Strengths and weaknesses

One intrinsic limitation of all measures of SWB is that reported data cannot be validated against objective measures of the same phenomena, for the simple reason that no obvious external benchmark for these phenomena exists. However, indirect attempts to validate measures of SWB, such as by comparing people's self-reported SWB with the frequency and intensity of people's smiles or with reports provided by other persons, do confirm that these subjective measures have some validity against these benchmarks (Krueger *et al.*, 2008). A similar conclusion can be drawn from evidence that many of the coefficients relating subjective well-being measures to various determinants are quite similar across countries.[5]

Two of the most important weaknesses of SWB relate to *interpersonal comparability* and (in the case of life-evaluations) the possibility that *external events may disturb evaluations and their measures*.

- Different individuals could use the scales differently. This may not be very important for the estimation of "averages" of SWB over culturally homogeneous populations but is more of a problem for the measurement of dispersion within a country. The problem of heterogeneous standards could be partly alleviated by asking respondents to describe

5. There are, however, exceptions, in terms of both the significance of some determinants (*e.g.* religion) across countries and the strength of their association with subjective well-being (*e.g.* age).

the standards that they use when judging a specific situation, but the application of these methods remains rare.

- Making a global evaluation of one's life is not always easy: people need to make a cognitive effort to provide an answer, and some surveys may fail to put respondents in good conditions to ponder the issues. Also, answers may be polluted by current moods (finding a dime a few minutes before the interview, or the state of the weather) or by the order in which questions are posed (asking about dating before posing questions on life-evaluation). More knowledge about the validity of these subjective measures could also be developed by studying how evaluations vary with the time people are given to think about the issue, and the detail with which they are asked to justify their evaluations.

But, overall, the overwhelming conclusion that we derive from existing analyses of various aspects of subjective well-being is that these measures tap into QoL in meaningful ways, despite the persistence of many unresolved issues about the relationship between the various measures and the aspects of human experience that are not captured by these subjective measures. The type of questions that have proved their value within small-scale, unofficial surveys should start being included in larger-scale surveys undertaken by official statistical offices.

Capabilities[6]

While psychological studies of QoL focus on people's own feelings, other approaches broaden the information set relevant for valuing people's lives, beyond their self-reports and perceptions. The most prominent of these approaches is rooted in the notion of "capabilities" (Sen, 1987b, 1993). This approach conceives a person's life as a combination of various "doings and beings" (functionings), and assesses QoL in terms of a person's freedom to choose among the various combinations of these functionings (capabilities).

- ***Functionings*** is a broad term used to refer to the activities and situations that people spontaneously recognize to be important. These can also be conceived as a collection of the observable *achievements* of each person (e.g. their health, knowledge or having a meaningful job). Some of these achievements can be quite elementary, such as being safe and well-nourished, and others quite complex, such as being able to express oneself in public without shame. As people in different places and times have different values and experiences, the list of the most relevant functionings depends on circumstances and on the purpose of the exercise. In this perspective, the well-being of a person is a summary index of the person's functionings.

- ***Freedom*** requires expanding the range of information relevant for assessing people's lives beyond their observed achievements, to the full range of *opportunities* open to them. The limits of focusing on achievements for assessing QoL become obvious when considering cases where a low observed functioning (e.g. low calorie intake) reflects a choice (as in the case of fasting) or where a high level of functioning reflects the choices of a benevolent dictator. The concept of freedom emphasises the importance of empowering people to help themselves, and of focusing on individuals as the actors of their own development.

6. This section draws on a paper prepared for the Commission by Sabina Alkire (Alkire, 2008).

The intellectual foundations of the capability approach include a number of notions. First is a focus on human ends, and on the importance of respecting people's ability to pursue and realise the goals that he or she values. Second is the rejection of the economic model of individuals acting to maximize their self-interest heedless of relationships and emotions, and a recognition of the diversity of human needs and priorities. Third is an emphasis on the complementarities between the various capabilities for the same person (while valuable in themselves, many of these capabilities are also means of expanding others, and leveraging these interconnections increases quality of life) and their dependence on the characteristics of others and on the environment where people live (e.g. illness may spread from one person to another and be influenced by public health and medical programmes). A last feature of the capability approach is the role played by moral considerations and ethical principles, and its central concern with justice, in the form of either bringing each person above a given threshold for each capability, or assuring equal opportunities to all in the "capability space" (Alkire 2003).

The implications of the capability approach are not limited to the measurement of QoL but extend to the evaluation of policies. Policies supportive of human development should expand the opportunities available to people, which would be valuable irrespectively of the effect on people's subjective states.[7] While subjective states may be considered as being part of the set of capabilities considered, the capability approach emphasises that people may adapt to their life-circumstances, and that this adaptation makes subjective feelings inadequate as the *sole* metric for assessing QoL.

The practical implementation of the capability approach requires taking a number of steps. A first one is choosing among dimensions. While some authors have strongly argued in favour of specifying a single list of "central" capabilities as a necessary step to make the capability approach operational[8], others have argued against "freezing" a list of capabilities based on expert views. In practice, most of the methods used in empirical applications of this approach select dimensions (or capabilities) based on the following: what type of data are actually available; *a priori* assumptions about what people do or should value; existing lists that have achieved some degree of political legitimacy (e.g. universal human rights, the Millennium Development Goals); surveys on what people value as important; and participatory processes that periodically elicit people's values and perspectives (Alkire, 2008).

7. This will apply even when measures of subjective states do reflect changes in objective features of QoL.

8. Nussbaum (2000) lists the following ten "central human functional capabilities": i) *life, i.e.* being able to live to the end of a human life of normal length; ii) *body health, i.e.* having good health and shelter, and being adequately nourished; iii) *bodily integrity, i.e.* being able to move freely, being secure against assault and violence, having opportunities for sexual satisfaction and choice in matters of reproduction; iv) *senses, imagination and thought,* which include the ability to do things in an informed and cultivated way, to use one's mind in ways protected by guarantees of freedom, to have pleasurable experiences and avoid unnecessary pain; v) *emotions,* such as to love, grieve, experience longing, gratitude and anger; vi) *practical reason, i.e.* being able to form a conception of the good and to engage in critical reflection about the planning of one's life; vii) *affiliation,* in terms of being able to live for and towards others and having the social basis of self-respect and non-humiliation; viii) *other species,* in the sense of being able to live with concern for and in relation to the world of nature; xi) *play, i.e.* being able to laugh, play and enjoy recreational activities; and x) *control over one's environment,* in terms of being able both to participate in political choices that govern one's life and to hold property, both formally and in terms of real opportunities.

A second practical step in the implementation of this approach is getting data and information on these various dimensions. Here a practical difficulty is that most data generally refer to functionings (*i.e.* description of individuals' states) rather than to capabilities (*i.e.* the set of opportunities that are available to each person). However, many functionings, such as health and education, also determine capabilities (to consume, to move, to participate), while some data may directly refer to people's rights and freedoms (e.g. to participate in political decision-making, to join organizations at work and in society). In addition, information on capabilities is sometimes available from surveys that probe respondents for their reason for not doing something (*i.e.* whether people did not consume more of a good because of preferences or constraints) or through additional information on the extent of choice that people have. More generally, one can imagine a broader framework in which both capabilities and achieved functionings serve to describe individual situations (Sen, 1985 and 1992).

A third step in the implementation of this approach involves valuing the different capabilities. This valuation allows converting the vector of functionings and capabilities into a scalar measure of well-being or advantage. While this step raises a number of difficult issues (which are further discussed in the last section of this chapter), the capability approach emphasises that several sources of information can be used, separately or jointly, for that valuation (e.g. survey-data on SWB provide evidence on valuation), that people may differ in their valuation of a given vector of functionings and capabilities, and that such differences may imply recognizing the "partial" nature of these rankings (*i.e.* two persons may agree that both states A and B are superior to C, while disagreeing on the ranking of A and B); in these conditions, the intersection of these partial orderings may reflect the minimum that could safely be said while respecting both the incompleteness and conflicts of people's valuations (Sen, 1987b).

In practice, a large body of empirical research has drawn inspiration from the capability approach. The "human development index" launched by the UNDP in 1990 is rooted in a notion of development conceived as a process of enlarging people's choices and opportunities. More recently, the *Oxford Poverty and Human Development Initiative* has aimed to develop specific survey questions (on employment, empowerment, safety and security, meaning and value, and ability to go without shame) to give operational content to a measure of poverty in a space of capabilities.[9]

Economic approaches: welfare economics and fair allocations

Subjective well-being and capabilities have developed in close connection with disciplines other than economics (psychology for the former, moral philosophy for the latter). In the economics field, the welfare economics tradition and the theory of fair allocation propose other ways of addressing the issue of how to include non-market aspects of QoL into a broader measure of well-being. The fundamental goal of these approaches is to respect individual preferences in the weighting of the various dimensions of QoL. These approaches are based on consumer theory, in which preferences are described by indifference sets (*i.e.* sets containing all situations in which the individual is indifferent). While consumer theory

9. See http://www.ophi.org.uk/ .

usually bears only on the consumption of goods and services, the same approach can be extended to other aspects of QoL.

Welfare economics has traditionally relied on the notion of "willingness-to-pay" to extend the scope of monetary measures to non-market aspects of life (Boadway and Bruce, 1984). People make trade-offs between different dimensions of their situation, and this allows relating changes in their QoL to changes in their income that are equivalent according to their own preferences (*i.e.* their willingness-to-pay to achieve a given level of health, education or exposure to pollution). This approach, however, has been severely criticized because of the potential inconsistency in the conclusions it produces and because evaluations based on total willingness-to-pay may disproportionately reflect the preferences of those who are better-off in society (Box 2.2).

Box 2.2. Welfare economics and the sum of people's willingness-to-pay

It is common in cost-benefit analyses to add up people's willingness-to-pay (WTP) in order to evaluate whether a change in their situations is good or not from the perspective of society as a whole. This approach has been used to evaluate past improvements in health and life expectancy in terms of equivalent increments in wealth or income (Becker *et al.*, 2005; Murphy and Topel, 2006).[10] For example, one application of this approach concludes that the value of the gains in life expectancy realized in the United States over the past century is about as large as the value of all other consumption goods and services together (Nordhaus, 2002). Some of the main weaknesses of this sum-of-WTP approach are as follows:

- First, the references it takes for the non-income dimensions of life may vary with the comparison at hand (typically, either the initial or the final values).[11] This implies that it may be possible that the total WTP is positive when considering a move both from situation A to B as well as from situation B to A.
- Second, by adding WTP across individuals, this approach does not give any priority to the worst-off. This approach favours the rich, whose WTP is greater because of the presence of income effects.

Because of these drawbacks, most welfare theorists regard this approach as inappropriate, although it remains influential in applied cost-benefit analysis, as used for industrial policy, international economics and health economics. More generally, studies showing the total WTP for a change in QoL are informative about the relative importance of changes in various non-monetary dimensions of QoL compared to income. However, they cannot be used to evaluate changes globally or to guide policy decisions.

The theory of fair allocations overcomes some of the weaknesses of welfare economics by explicitly referring to equity criteria.[12] This theory has also inspired several methods for comparing individuals' multidimensional situations in a way that respects their preferences (Moulin and Thomson 1997; Maniquet 2007). In this tradition, a particular method (the

10. For example, the change from a state characterised by an average annual income of €15,000 and by 65 years of average life expectancy, to a state where average income is €30,000 and average life expectancy is 75 years, might be equivalent, in terms of summing up individuals' willingness-to-pay, to moving to a situation where income rises to €55,000 and life expectancy remains at 65 years.

11. When the initial values are taken as a reference, this approach computes the "equivalent variation" of income; when the final values are taken as a reference, this approach computes the "compensating variation", *i.e.* what sacrifice in income would bring the individual back to the state of initial satisfaction.

12. The theory of fair allocations studies the allocation of resources among people with different tastes and abilities, subject to a number of fairness criteria, such as "no-envy" (*i.e.* no agent should prefer another's bundle), "solidarity" (*e.g.* no agent should be hurt by an increase in available resources) and "lower-bounds" (*e.g.* no agent should prefer the equal-split solutions, Fleurbaey, 2008b).

"equivalence approach") might be the most suitable for applications. This method consists in selecting a "reference" set of individual situations that are easy to rank from better to worse (typically selected so that a better situation dominates a worse one in all aspects of QoL); then the conditions of any given person are assessed by identifying the particular situation in this reference set that is *equivalent* to his current situation according to his own preferences. For instance, if two individuals are indifferent with respect to their current situation and the same situation in the reference set, they are considered to be equally well-off. This approach requires, in addition to information about preferences, the choice of a specific reference set. An example consists of taking as a reference a set of situations that are identical in all non-market aspects of QoL and differ only with respect to income; in this case, people could be compared in terms of their "equivalent income", *i.e.* the income they would have in their equivalent situations (which differ only in respect to income, and share the chosen reference values for the non-monetary aspects of QoL). This approach – which builds on the Bergson-Samuelson approach to welfare economics[13] – makes it possible to use a monetary metric for measuring non-market aspects of QoL. The economic approaches described in this subsection require data on individual preferences that are not easy to assess: various methods for collecting suitable data are described at the end of this chapter.

Relations between the various approaches

All the approaches described here address the difficulty of bringing together the various aspects of QoL. The subjective well-being approach focuses on subjective and mental states, and incorporates all other aspects of QoL in the analyses as potential determinants of well-being, by measuring how they impact individual perceptions. The approaches drawn from the tradition of welfare economics and from the theory of fair allocations also build on subjective views about the various aspects of QoL but seek to evaluate QoL itself, taken as an object of individual preferences rather than as a subjective output.[14] Conversely, the capability approach recognizes genuinely distinct, plural and incommensurable achievements. Because of this, proponents of this approach are generally reluctant to engage in the construction of indices that summarize the evaluation of an individual situation into a single number, although some methods of aggregation that draw on this perspective are discussed by proponents of this approach. Subjective well-being may also be viewed as multidimensional if one considers that evaluative judgements, positive affect and negative affect are all distinct aspects of subjective well-being.

Choosing between these approaches is ultimately a normative decision. It depends on what aspects of life are deemed to be most relevant in the evaluation of QoL. Subjective well-being

13. Neither Bergson nor Samuelson committed themselves to specific indexing methods, although money-metric utilities and other indices of the equivalent sort were provided as examples by Samuelson. The equivalent income (or "money-metric utility") approach has been advocated by Deaton and Muellbauer (1980) and by King (1983). For a discussion of specific criticisms that this method has attracted from social choice theorists, see Fleurbaey (2008a).

14. To illustrate, if one thinks of any given aspect of subjective well-being (life-evaluation, positive or negative affects) as measuring some kind of "utility function" $U(x)$, where x represents all the aspects of an individual's situation that matter to the individual, the subjective well-being approaches work with the units of U (which may differ depending on the specific aspects of subjective well-being under consideration), while the economic approaches evaluate x. So does the capability approach, provided that x is sufficiently comprehensive and incorporates what the individual is able to do, not just what he achieves.

may be considered as encompassing all key capabilities, while focusing on their impact on people's subjective states. Conversely, the capability approach treats subjective well-being as just one aspect of QoL among the many capabilities that people have reasons to value. The economic approaches are somewhere in between, as they seek to weight the various aspects of QoL without assuming *a priori* that subjective well-being takes full weight; further, contrary to the capability approach, these approaches insist on relying on people's own preferences. It should be stressed, however, that the difference between the capability approach and the one based on notions of fairness is rather thin, as both share a focus on equality across all members of society. Finally, subjective well-being, capabilities, and fair allocations have developed fairly recently. The progress achieved in these fields provides a real opportunity today to make progress in measuring QoL.

Features of quality of life

In practice, all approaches to QoL share an emphasis on a range of features in people's lives that are important either intrinsically, as objective expressions of a good life, or instrumentally, to achieve valuable subjective states or other objective goals. Some of these features may be conceived as referring to particular functionings (*i.e.* descriptions of people's doings – *e.g.* working, commuting – and beings – *e.g.* being healthy or educated) while others may be conceived as freedoms in particular domains (*e.g.* political voce and participation).

Determining which elements should belong to this list of QoL features (*i.e.* the choice of a "space" for QoL evaluations) inevitably depends on value judgements about which aspects are of greater importance at a given place and time, and on the purpose of the exercise (*e.g.* whether we want to describe developments within each country, or to compare countries at different levels of development) but also on judgements.[15] In principle, this judgement can only be based on deliberative processes, whereby people in various communities identify those aspects that most directly bear on their living conditions. In practice, practical experiences with these deliberative processes highlight a number of themes that are shared across a large number of constituencies; a similar degree of consistency also emerges when comparing the "frameworks" developed under various initiatives that focus on measuring broad concepts such as "well-being", "human development", and "societal progress".[16] In most cases, these themes include not just the measures of people's subjective states described above, but also measures of their health and education, their daily activities, their participation in political processes, and the social and natural environment in which people are embedded and which shape their sense of security. The existence of these recurrent themes opens the way to comparisons of quality of life across people and countries.

Health

Health is perhaps the most fundamental component of capabilities as, without life, none of the other components has any value. And yet, as noted by Wolfson (1999), we know far more

15. As argued by John Hicks: "In order that we should be able to say useful things about what is happening, before it is too late, ... we must concentrate our attention, and hope that we have concentrated it in the right place" (quoted in Alkire, 2009).

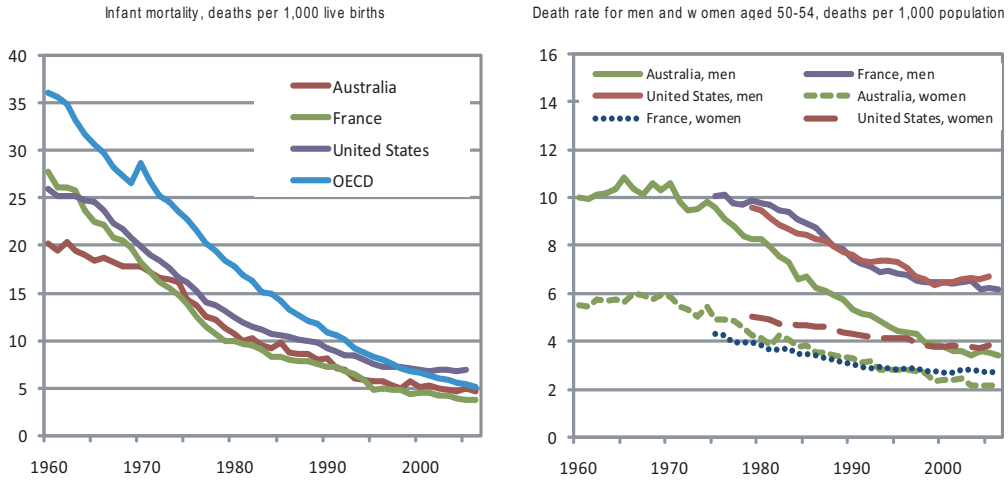
16. See, as an example, the taxonomy developed in the context of the OECD-hosted "Global Project on Measuring the Progress of Societies", available at www.oecd.org/progress/taxonomy.

about the costs of health care, and the number of patients treated, than we know about the impacts of the treatment, and the health status of the population in general. Health impacts on both the length of life and its quality. While a fairly wide range of data are available on the former, measurement challenges remain much more important on the latter. While a basic infrastructure of health statistics is in place in most of the world's countries, this is by no means universal. Further, adequate measures are still lacking in several domains.

Mortality

Mortality is better-measured and less subject to error than other health measures. Mortality statistics hence constitute the basic indicator of a population's health status, and many researchers rely only on mortality-based measures to look at population health. Since 1960, the mortality rates of both infants and adult men have declined significantly, with a more moderate decline for adult women, partly reflecting differences in smoking behaviour (Figure 2.1). The gains are even sharper over a longer time-period: for example, infant mortality rates in the United States declined from 100 deaths per 1,000 live births in 1915 to around 7 by the mid-2000s, while in France the decline over a similar period was from above 140 to around 4. However, even mortality statistics have problems.

Figure 2.1. Trends in infant and adult mortality in selected countries



Source: OECD Health data.

First of all, mortality rates by age and gender require complete vital registration systems. These are available in all developed countries as well as in some countries in Latin America (whose registration systems are close to complete), but not elsewhere. In poor countries, infant mortality rates are reasonably well estimated from surveys, and some countries (such as India) have a sample registration system or ask questions in their population censuses (China). But for much of the world there is a huge gap in available measures of adult mortality rates, a gap that is a giant problem for thinking about HIV/AIDS in Africa, for example. Adult mortality rates there and other statistics based on them are *modelled*, rather than measured, by international agencies, typically based on infant mortality rates and on

assumptions about AIDS mortality.[17] The wide margins of error implicit in these estimations are not always recognized. It follows that for much of the world international measures of health status that depend on adult mortality are deeply problematic, making it impossible to monitor current efforts to improve global health. Historical comparisons even for developed countries (such as those mentioned above) raise similar problems; the United States, for example, has had a complete registration system at the federal level only since the 1930s (Hetzel, 1997).

A second problem is how best to construct summary statistics based on mortality data. Mortality rates are vectors, not scalars, *i.e.* one number per age and gender. To get an overall population measure requires the use of some aggregation formulas. In general, crude mortality rates (mortality divided by population) are not useful indicators of population health, as they depend on the age structure of the population. Several types of aggregation formulas are currently used, each with its own advantages and disadvantages.

- ***Life expectancy at birth*** is the imaginary expectation of life for someone born today who will experience today's age-specific mortality rates throughout life. While this measure is the most common indicator of health conditions, it understates how long someone born today can expect to live if mortality rates continue to fall (and, conversely, it overstates it in those countries where mortality rates are projected to rise). Also, the effects of lower mortality on life expectancy depend on the age at which the mortality decline takes place. In particular, changes in life expectancy will be larger the earlier the age at which the decline in mortality takes place. This has implications for time-series comparisons. For example, in the 1950s, vector control and immunization campaigns in poor countries led to big declines in infant and childhood mortality, whereas the big declines in mortality in rich countries have been among the middle-aged and the elderly. These developments had the effect of sharply narrowing the gap in life expectancy at birth between rich and poor countries. However, this narrowing also hides an important philosophical issue, *i.e.* whether a decline in mortality rates among newborn infants is indeed "better" than a decline in mortality among middle-aged people. Answers to this question also depend on the fertility declines that typically follow lower mortality rates in poor countries (Deaton, 2006).

- ***Standardized mortality rates***. Mortality rates of different age groups should be combined in ways that do not unduly reflect differences in age structure between countries and periods. However, different methods to achieve this standardization can be used, and results may differ depending on the properties of the underlying data. Two frequently used measures are "direct" and "indirect" standardization. Both methods apply the mortality structure of a place or time to a reference population. For the United States, for example, the mortality rate of a given state (*e.g.* Florida) could be "directly standardized" by calculating the fraction of the population that would have died had

17. Infant mortality (annual deaths under age one, expressed per 1,000 live births) is a long-standing health indicator. It is of value in that it reflects the effects of economic and social conditions on the health of mothers and newborns, as well as the effectiveness of health systems. It is commonly included in all evaluations of living standards, being inversely related to per capita GDP. Infant mortality is an important indicator in developing countries and a cornerstone of the United Nations *Millennium Development Goals*, for example, but less so in developed countries, where small cross-country differences around very low levels may reflect differences in medical and recording practices.

Florida's age-specific mortality rates held but the US age-structure of the population applied. Or it could be "indirectly standardized" by using the age structure prevailing in Florida, and comparing Florida's *actual* mortality rate to the one that would have been expected had the overall US age-specific mortality schedule been applied to its actual population.[18] These two approaches, however, will lead to different numerical values, and even different rankings among countries and population groups, when the survival curves of the two countries being compared cross each other. In these circumstances, both types of standardization should be used.

- ***Median life expectancy*** is sometimes useful and behaves very differently from *mean* life expectancy. Age-specific mortality curves are high in the first years of life (especially in poor countries), dip to a minimum and then rise slowly from about age 30. As a result, reductions in infant mortality, by reducing the mortality peak at low ages, sharply reduce interpersonal inequality in years of life lived, with implications for some of the measures of health inequalities that are discussed below. In the highest mortality countries – Mali in 1960, for example – median life expectancy at birth was around 5 years, though mean life expectancy was around 37 years. The use of median life expectancy has other implications. For example, as child and infant mortality falls, the median age of death suddenly jumps from young to old; as soon as that happens, median life expectancy exceeds mean life expectancy while, after that, the mean will typically rise faster than the median if mortality continues to fall. These properties of median life expectancy point to the complexity of reducing the age-specific mortality structure to a single measure of population health, given its typical properties, and to the problems posed by using standard measures of life expectancy when comparing countries with differences balances between child and old-age mortality.

The different properties of these various measures used to summarize the information provided by the vector of age- and gender-specific mortality rates stress the importance of using a set of these synthetic measures for assessing mortality.

Morbidity

Measures of morbidity, or non-fatal health conditions, are even more varied than for mortality. These measures are relevant measures of the functioning of people who are alive and, as such, they can be used at the individual level, not just for populations. In other words, we can talk about the distribution of morbidity across people in a way that is hard to do for life expectancy. While some scholars argue that population morbidity moves in parallel with population mortality, there is no guarantee that it must always do so, and there have been times when it did not. For example, Riley (1997) argued that the decline of mortality in Britain in the XIXth century was accompanied by large increases in morbidity, at least initially; in his terms, British working men were "sick, not dead". In some circumstances, the relation between mortality and morbidity can vary across countries and the population covered. For example, a recent review of trends in severe disability among the elderly in 12 OECD countries found evidence of a decline in Denmark, Finland, Italy, the Netherlands and the United States; of an increase in Belgium, Japan and Sweden; of stability in Australia,

18. These two measures are closely analogous to Laspeyres and Paasche price indices.

Canada; and of changes in different directions, depending on the source used, in France and the United Kingdom (Lafortune et al., 2007).

The available indicators of morbidity are based on either administrative and register data, or on self-reports by people. The available metrics, however, remain limited in important ways, and they often provide an inadequate basis for benchmarking morbidity across countries and for assessing its evolution over time. Some of the most familiar measures of morbidity include the following:

- ***Anthropometric measures.*** These measures are especially useful for children. Height and weight allow calculating measures of stunting (long-run nutritional deprivation) and wasting (short-run nutritional deprivation), which are central data in health measurement in poor countries. Adult height is strongly affected by nutrition and disease in the first three years of life, and it does not change (much) in adulthood. As a result, even a single cross-section of adults classified by height can tell a great deal about the childhood nutritional and health history of the current population. Adult height provides a measure of the combined effects of disease and income during childhood: taller people earn more and are better educated and more satisfied with their lives than shorter people; they also experience more enjoyment and happiness, less pain and depression, and lower blood-pressure, cholesterol and mortality. These individual effects, however, can only be observed in large samples. Both height and weight data are needed to compute measures of obesity (using the body-mass index, BMI), which is becoming a public health issue in most OECD countries.

- ***Disease-specific information.*** Information on the prevalence of different diseases can be based on disease registries, registries of general practitioners, hospital discharge data, pharmaceutical databases, and health interview surveys. The information collected is usually focused on a shortlist of diseases, typically selected among those that contribute the most to the burden of disease in different countries. Health interviews often include questions of the form: “has a doctor ever told you that you have X”, followed by a shortlist of diseases. The advantage of this type of question is that it seeks information on diseases that have been diagnosed by a health professional, as opposed to relying on a purely subjective assessment; the drawback is that responses will be affected by the extent to which people have access to medical doctors. Some population-based surveys also collect information on biomarkers, including blood pressure and anaemia. Information from these surveys has allowed important (and surprising) comparisons of health status across nations. For example, Banks *et al.* (2006) conclude that the US population in late middle age are less healthy than their British counterparts with respect to diabetes, hypertension, heart disease, myocardial infarction, stroke, lung disease and cancer. This conclusion applies when relying on biological markers of various diseases, and holds true even after controlling for a standard set of behavioural risk-factors (*e.g.* smoking, overweight, obesity, and alcohol consumption). Surveys are increasingly being put in place in Europe and the United States (as well as India and China) that would support such comparisons, particularly among people over age 50.

- ***General measures of self-reported health***. A commonly-asked question in national health surveys is of the type: “How is your health in general?” Responses to this type of question can be collected easily and quickly in surveys, and they predict subsequent mortality even after controls for a physical examination by medical personnel. They are related to a wide range of diseases and conditions in a cross-section of the population,

and they decline with people's age, though perhaps not as much as expected. Data on self-reported general health highlight other interesting features. For example, women have worse self-reported health status than men, though they live longer. This is not necessarily a flaw in the measure, as women may get sick while men get dead, so rankings of mortality and morbidity would differ (Case and Paxson, 2005). A more serious problem is that mean self-reported health status has been stable over time, in spite of large reductions in mortality. This suggests that these measures of self-reported general health may be subject to the sort of adaptation that would make them unsuitable as an objective measure of morbidity. A critical problem for the international comparability of these measures is the cultural bias that arises when answering these questions, as well as the use of different scales across countries: while self-reported health is usually reported on a "Likert scale" of 1 (excellent) to 5 (poor), there are exceptions (*e.g.* Japan).[19]

- ***Vignettes*** have recently been proposed to try to remove any systematic biases in self-reported answers to questions on general health status. In this approach, respondents are asked to rate a standard scenario concerning someone whose health status is described, with these reports then used to standardize the self-reported measures of health status. The extent to which the use of vignettes solves the underlying problem of the non-comparability of individual answers is still unsettled. Another approach to correcting possible biases in self-reported information is to carry out tests of people's ability to see, hear, walk, think and remember. Such tests are, however, costly, as they involve face-to-face encounters and long interviews and examinations.

- ***Specific measures of self-reported health.*** Several national health surveys also include more specific questions on key dimensions of health status. These questions often relate to limitations in the ability to see, hear, walk, grasp objects, think and remember. These surveys may also include questions about pain and mood. An important issue for cross-country comparisons is the lack of standardization of survey questions and response categories. However, some initiatives have been recently undertaken to develop a small set of harmonized measures of health status in national surveys.[20] A further issue for these measures relates to cultural biases that might affect the comparability of responses across countries.

- ***Activities of daily living and instrumental activities of daily living.*** Health surveys also collect data on the ability of people to perform various activities of daily living (ADLs) or instrumental activities of daily living (IADLs).[21] These are direct measures of functionings in various dimensions, and they appear to correspond well to actual

19. For example, in European countries the response scale is usually "symmetric", with response categories ranging from "very good, good, fair, poor, very poor"; conversely, in non-European countries response categories are often "asymmetric" (*i.e.* "excellent, very good, good, fair, poor"). The wording of questions also matters. For example, *World Values Survey* questions ask about "satisfaction with health" (on a 0 to 5 scale) rather than about "self-reported health"; based on this question, satisfaction with health declines steadily with age. However, when Canadian respondents to another survey were asked to rate their health satisfaction *relative* to other people of the same age, this downward trend disappears. This suggests that measures of health satisfaction are less likely to reflect health conditions than are subjective assessments of health status (Helliwell, 2008)

20. One recent example is the *Budapest Initiative on Health Status Measures* http://www.unece.org/stats/documents/ece/ces/2007/6.e.pdf.

functionings. They do not suffer from the problem that the poor say they are less sick than the rich, in circumstances where other evidence suggests that this is far from the case. A drawback is that they are useful only for older people, since almost all young people can perform all of these activities.

Measurement challenges are also high when moving beyond physical health to include various forms of mental disorders. Most research based on this subject reports high prevalence of mental disorders, For example, data from the WHO *World Mental Health Survey Initiative* show that the share of people who have experienced some type of mental health disorder (*i.e.* anxiety disorders, mood disorders, disorders linked to impulse control or the use of alcohol or drugs) in the previous year ranged from 9% in Italy, Japan, Spain and Germany, to 18% in France and 26% in the United States, with much higher levels when looking at life-time prevalence (close to 40% in France and 50% in the United States, Kessler *et al.*, 2007). Most of these cases are classified as "mild" or "moderate", and most of them go untreated. In European countries, people who are young, female, unemployed, disabled and unmarried have higher risks of mental disorders (Alonso *et al.*, 2004). It is more difficult to assess whether the prevalence of mental disorders is rising, with various studies arriving at different conclusions[22]. More research on mental health – based on national representative samples, common diagnostic questions, and surveys repeated at regular intervals – is clearly needed, and governments should assist by undertaking more systematic surveys.

Combined health measures

The various dimensions of people's health status have led, in recent years, to the development of a variety of measures that try to combine mortality and morbidity into a single measure of health status. All of these measures aim to calculate the average number of years of life spent in good health, after adjusting for years lived with some form of illness or disability. Examples include the following:

- ***Disability-Adjusted Life Year*** (DALY) is typically disease-specific: the burden of each disease in the population is obtained by adding the number of years of life lost through death and the number of "equivalent" years lost to disability. The equivalence is based on a set of weights for a long list of medical conditions (*e.g.* a year with an untreated fractured skull to someone aged 45-59 is valued as equivalent to 42% of a healthy year, an amputated foot loses 30%, and schizophrenia loses 57%). There are, however, ongoing debates about the best method to elicit the valuation of different health states.[23]

21. *Activities of daily living* are the things normally done in a typical day, such as bathing, dressing, eating, walking, using toilets and moving from a bed to a chair. *Instrumental activities of daily living* are not necessary for fundamental functioning but enable the individual to live independently within a community: they include items such as preparing meals, taking medicines, using a telephone, shopping, managing money and performing light housework.

22. For example, Weissmann *et al.* (1992), based on a study of major depression in a large number of countries, conclude that "more recent birth cohorts are at increased risk of major depression". Similarly, studies based on data over long periods of time report a rising prevalence of mental disorders in the Netherlands (Hoidiamont *et al.*, 2005; Verhaak *et al.* , 2005), Belgium (Wautericks and Bracke, 2005) and the United Kingdom (Sacker and Wiggins, 2002; Oswald and Powdhavee, 2007) but stability in the United States (Murphy *et al.*, 2000; Kessler *et al.*, 2005).

23. These measured were originally developed for the World Bank's 1993 *World Development Report.*

- ***Disability-Free Life Expectancy*** (DFLE) quantifies the remaining years a person of a certain age can expect to live without disability. Its computation requires mortality data from standard life tables, and estimates of the prevalence of disability among population groups, usually self-reported measures of general health based on a simple dichotomous variable (with a value of 0 if people report being moderately or severely disabled, or 1 if they report not being disabled).[24]

- ***Health-Adjusted Life Expectancy*** (HALE) measures the number of years that an individual can expect to live in full health if current morbidity and mortality conditions prevail. It adjusts life expectancy by giving a larger weight to the years of life lived in good health compared to the years lived in poor health. The measures of health status usually come from questions on key dimensions of physical and mental functioning drawn from health surveys. The weighting of years of life depends on the valuation of different combinations of health status, with values ranging from 0 (in the case of death) to 1 (in the case of full health).[25]

Regular monitoring of summary measures of population health across large numbers of countries is still uncommon. These measures are easy to interpret, as they are based on recognizable concepts of illness, disability and mortality, but their calculation is complex, since they rely on different data sources — *e.g.* a measure of morbidity or disability, usually from a health or household survey, and a measure of mortality, often from a census cohort linked to a mortality registry. More fundamentally, however, their construction raises a number of ethical problems: for example, DALYs value a disabled life at less than a life without disability (Anand and Hanson, 1997)[26]; this implies that these measures, if used to design health-care delivery, would implicitly disfavour the treatment of people with disabilities. Similarly, the legitimacy of the weights used in the construction of these combined measures is not always clear; for example, one of the major "findings" of the global burden of disease studies — the heavy burden of mental illness — may owe as much to assumptions as to measurement.[27] In spite of these problems, combined health measures respond to a genuine need, and the drive to develop them is expected to continue. Both Eurostat and the World Health Organization have recently intensified their efforts to reach a consensus on which indicator to use, and to harmonize underlying data sources and methods for their construction. In any case, however, "combined health measures" should be conceived as part of a coherent system of health statistics, which should include the capacity to analyze developments for individuals and population groups.

---

24. DFLE has achieved some currency since being officially adopted by the EU as one of the *European Structural Indicators*. In the United States, an indicator of life expectancy free of activity limitation (an equivalent concept to DFLE) is one of several summary measures of population health monitored under the *US Healthy People 2010 Initiative*.

25. Estimates of HALE have been produced and are being regularly updated in some countries (*e.g.* Canada). WHO has also reported estimates of HALE for all countries in its 2004 *World Health Report*.

26. This problem is alleviated when some priority for the worse-off is introduced in the evaluation. If the disabled are among the worse-off, it is more worthwhile to expand their life expectancies even if this does not add much to their total health as measured by such indices.

27. An alternative to this type of "macro level" aggregation is provided by two-steps aggregation methods that, first, generate some summary health score for each person, and then aggregate these individual level summary scores to the population level (Wolfson, 1999). This individual-level aggregation is the analogue, in the health sphere, of the methods based on "aggregating indicators for each individual" discussed later in this chapter in the section dealing with aggregation among QoL dimensions.

Health inequalities

The variety of existing health measures raises challenges not just for cross-country comparisons of average health status but also for within-county inequalities. Analyzing the size and determinants of these inequalities in health status has become a growing focus of epidemiological research in recent year. While some of these studies are based on simple measures of dispersion in the age of death – or, alternatively, in the length of life – among individuals in different countries (Edwards and Tuljapurkar, 2005)[28], these measures of dispersion cannot be used to classify people currently living, and they lack cogency. 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.

Most of the research on health inequalities has looked at between-group differences in either mortality or morbidity, highlighting a number of consistent patterns. People from lower occupational classes who have less education and income tend to die at younger ages and to suffer a higher prevalence of various health problems within their shorter lifetimes (Mackenback, 2006). Further, these differences in health conditions do not merely reflect worse outcomes for people at the very bottom of the socio-economic ladder but extend to people throughout the socio-economic hierarchy: for example, life expectancy in the United Kingdom increases when moving from unskilled manual workers to skilled ones, from manual to non-manual workers, and from lower-ranked office workers to higher-ranked staff (ONS, 2006). In European countries, the mortality rates of less-educated people are, on average, around 50% higher than for their more-educated counterparts, with lower gaps for women compared to men, and for the elderly compared to younger people (Mackenback, 2006). Further, these inter-group health inequalities do not consistently narrow over time, and are even widening in some countries. Inequalities in average life expectancy are also significant between ethnic groups.

While these patterns have an obvious relevance for assessing QoL in various countries, systematic data collection in this field remains sparse. It is not possible, for example, to compare the size of these health inequalities across countries in the same way as is conventionally done for income (and education). This reflects differences across studies in the measures of health outcomes used, in the individual characteristics considered (education, income, ethnicity), and in the reference population and geographic coverage of the various studies. Improvements in the measurement of health inequalities (in both absolute and relative terms), and especially the linkage between health outcomes and socio-economic status, should be recognized as a priority for statistical work in the years to come.

28. These measures highlight some significant differences across countries and over time. For example, the standard deviation in the age of death above 10 (a measure that abstracts from the decline in child mortality that affected all OECD countries) was highest in 2003 in the United States and France, and lowest in the Netherlands and Sweden. Japan, which started from a level close to that of the United States in 1960, recorded large declines until the early 1990s, and increases since then; Denmark, which started from low levels close to those of Sweden, recorded strong rises until 1990 and declines thereafter.

Education

A long tradition of economic research has recognized the skills and talents embodied in the population as a critical input into economic production. This human capital is created by investing in education and training, combined with parental inputs (*e.g.* supervision and mentoring) and societal resources (*e.g.* libraries, museums, etc). Approaches to the measurement of human capital based on frameworks similar to those used for economic accounting (described in Annex 1 of Chapter 3 of this report) figure prominently in research on non-market accounts (Abraham and Mackie, 2005).

What is recognized less often is that education, literacy, reasoning and learning are also important for QoL. They are important in their own right, as better cognitive functioning expands individuals' freedoms and opportunities, independently of any effect that it may have on people's earnings or on countries' economic activity. Indeed, as argued by Adam Smith, the father of human capital theory, the role of education and skills acquisition for the quality of life is *distinct* from its contribution to economic output. This implies that the neglect of education for a boy who is put to work at an early age has important non-economic consequences: "When he is grown up he has no ideas with which he can amuse himself." Smith also noted that the lack of education deprives people from participating in religious activities and can lead to "riot and debauchery" (*Lectures on Jurisprudence).*

Broader impact of education on QoL

Beyond its intrinsic value, education is also instrumental in achieving a variety of outcomes that matter for QoL. Some of these outcomes are monetary and accrue to the person making this investment (*i.e.* higher earnings, income and wealth). But others are non-monetary, and they benefit both the individual and society at large. Evidence indicates that individuals who attended school for longer, or who achieved higher educational qualifications, are more likely to report greater subjective well-being, to participate more actively in society and to enjoy better health. While the available research does not always allow determining whether the relation between education and these other dimensions of QoL is "causal" (rather than reflecting the importance of some other factor bearing on both education and other outcomes), the existence of an educational "gradient" (*i.e.* higher education contributes to various features of a good life) is in itself an important finding that warrants attention.

The available evidence on the broader effects of education relates to a number of QoL dimensions:[29]

- *Subjective well-being*. People with a tertiary education report higher evaluations of their lives as whole, and this effect holds even after controlling for income (Helliwell, 2008). Evidence on the causal effect of education on subjective well-being is provided by research on the effects of higher compulsory schooling in the United States, Great Britain and Ireland, which suggests that students who take an extra year of schooling experience higher subjective well-being; these effects are robust to a wide array of specification checks (Oreopoulos, 2007).

29. This evidence is analyzed in OECD (2007) and in follow-up work at the OECD Centre for Educational Research and Innovation.

- *Health.* There is compelling evidence of a correlation between education and health. In European countries, the mortality rates of less-educated men are, on average, 50% higher than those of more-educated men, while the difference is 30% for women (Mackenbach, 2006). The relation between education and health holds both for serious health problems, such as diabetes and high blood pressure, and for more common illnesses, such as colds, headaches and aches and pains from other causes (Stone *et al.*, 2008). Evidence on the "causal" nature of these effects (*i.e.* whether education causes health effects, or vice versa, or the existence of some other factors that shape the relation between the two) is sparser, and varies with the health dimension considered. For example, most evidence suggests that schooling reduces both the likelihood of smoking and the number of cigarettes smoked (see De Walque, 2007 and Grimard and Parent, 2007; see also Fuchs and Farrell, 1982 for an earlier analysis reaching an opposite conclusion), while the evidence is mixed when considering effects on alcohol consumption and obesity.[30] Education also influences access to health care, at least for some categories of services (*e.g.* specialist and dental care, van Doorslaer *et al.*, 2004), although these inequalities in access are believed to offer only an incomplete explanation of the observed gradient in health status. Overall, the consensus seems to be that the effects of education on health are positive and causal (Grossman, 2008; Cutler *et al.*, 2008). Groot and Maasen van den Brink (2007) calculate the value of these health improvements due to education as equivalent to between 15% and 60% of the financial returns to education.

- *Civic and Social Engagement* refers to the broad range of behaviours that influence people's participation in political and social activities, as well as their attitudes with respect to trust and tolerance of others. In all OECD countries, better-educated people are more likely to vote, although, across countries, voting turnout has declined despite higher educational attainment. The evidence on the effect of education on political engagement is mixed, with evidence of a causal effect in the United States (Milligan et al, 2004; Dee, 2004), but not in Europe (Milligan et al,, 2004; Siedler, 2007; Touya, 2006). Conversely, evidence based on changes in compulsory schooling and child labour laws suggests that education does not have an impact on volunteering rates, participation in groups, interpersonal trust and tolerance (Dee, 2004).

Indicators

Several indicators allow describing the state of education. Some of these indicators refer to inputs (*e.g.* school enrolment, educational expenditures and school resources), while others refer to throughputs and outputs (*e.g.* graduation rates, expected number of completed years of schooling, standardized test measures of student and adult achievement in terms of literacy and numeracy). Which of these indicators is more relevant depends on a country's stage of development and on the goal of the evaluation exercise.[31] For example, indicators of

30. Arendt (2005) and Spasojevic (2003) use reform of the minimum school-leaving age to identify the effect of education on BMI in Denmark and Sweden: in both countries the change lowered the probability of being overweight, but not significantly in the case of Denmark. Kenkel *et al.* (2006), using changes in high-school graduation laws in the United States, do not find any significant difference in the BMI of high-school drop-outs and high-school graduates. Lundborg (2008), based on data referring to monozygotic (identical) twins in the United States, reports that less-educated twins report lower self-reported heath and worse chronic conditions, but no difference in health behaviours, including BMI.

enrolment in elementary and secondary school are less informative in countries where the vast majority of students graduate from secondary school; similarly, some countries may combine widespread basic education for all with limited opportunities for higher education (*e.g.* China), while others may have a more developed system of higher education alongside widespread illiteracy (*e.g.* India). Finally, indicators of school enrolment inform about access to education, but they may provide a misleading picture of outcomes if schools do not provide effective instruction.

All these indicators, beyond informing about the average conditions in each country, can also provide important information about disparities in the distribution of educational opportunities. For example, if girls are afforded fewer educational opportunities than boys, or if ethnic or racial minorities are forced to attend separate (or lower-quality) schools, or are excluded from education altogether, then society as a whole may be worse-off. These concerns about inequality have become more important in the policy agendas of several countries. In the United States, for example, concerns in recent years about the lower educational opportunities for low-income and minority students have motivated the provisions of the No Child Left Behind Act. Educational disparities can be measured in many different ways, such as by comparing indicators across groups with differences in demographic features, parental income or education or across regions with different characteristics, or by computing a measure of the dispersion in the distribution of the educational measure of interest. All of these inequalities are simply lost by indicators that aggregate people's diverse experiences into an average measure of school performance or educational achievement.

Some of the most relevant indicators for assessing the role of education for QoL are measures of people's competencies (*i.e.* direct measures of what people have learned in various settings). Several tools have been developed in recent years to provide standardized measures of these competencies, such as:

- The OECD *Programme for International Student Assessment* (PISA), which has been conducted in three waves, covering 43 countries in 2000, 41 countries in 2003 and 58 countries in 2006, with a new wave now in the field. This survey covers 15-year-old students with tests focusing on reading, mathematics, and science.
- The *Trends in International Mathematics and Science Study* (TIMSS), which has now been conducted in four waves (1995, 1999, 2003 and 2007). This survey was developed by the *International Association for the Evaluation of Educational Achievement* (IEA), and covers 4th and 8th graders, with a focus on mathematics and science
- The *Progress in International Reading Literacy Study* (PIRLS), also developed by the IEA. This survey covers 4th graders, and focuses on reading literacy.

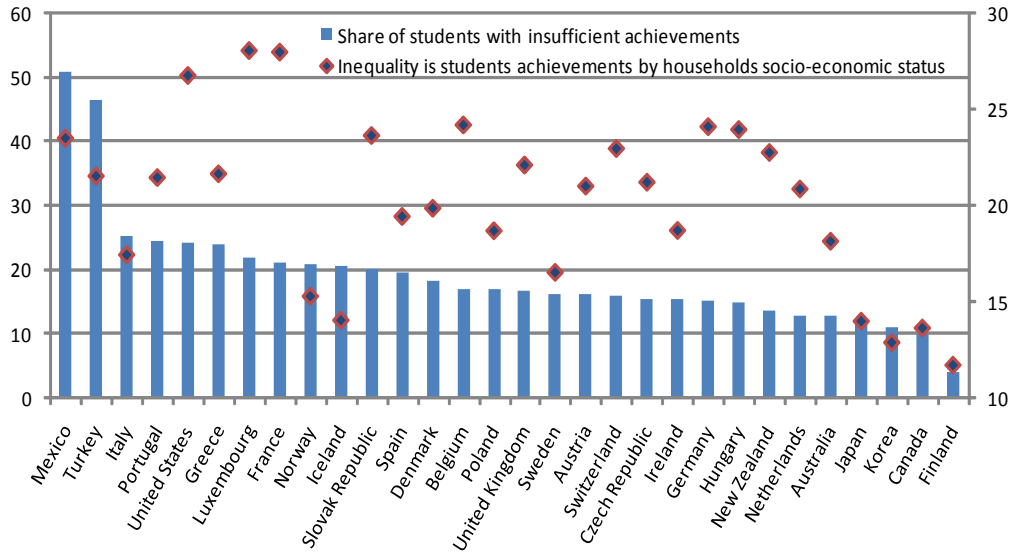
31. Education – as measured by adult literacy and by the combined enrolment rates for primary, secondary and tertiary education – is included in the Human Development Index. The particular weighting of the components of education in the HDI (two-thirds for literacy, one-third for enrolment rates) is somewhat arbitrary. These educational indicators have little discriminatory power among developed countries.

- The *International Adult Literacy Survey* (IALS), which tested adults in prose, understanding of documents and quantitative skills. This survey was conducted in various years from 1995 to 1998, covering 22 countries and regions around the world.

- The *Adult Literacy and Life Skills* survey (ALL), which tested people aged 16-65 in the fields of prose and document literacy, numeracy, as well as reasoning and analytical problem-solving, in two waves (2002 and 2006) covering twelve countries (or regions thereof).

To a large extent, the origins of these large-scale comparative testing programmes lie in the experiences pursued in the United States and Canada in the 1980s and early 1990s on the large-scale assessment of people and adults. The data collected as part of these international comparison programmes tend to be more comparable across countries than data on school enrolment or educational attainment. Indeed, even within the United States, there are significant differences in the way in which different states calculate high-school graduation rates, which significantly limit comparability. These test data have highlighted large differences in average test scores across countries,[32] as well as large inequalities in achievement among students. Inequalities in science scores are described in Figure 2.2 in terms of the share of students with insufficient levels of achievement (level 1 or 0), and the differences in scores between students from households with different socio-economic characteristics.

Figure 2.2. Inequalities in science literacy scores of 15-year-old students in 2006



Source: OECD Programme for International Student Assessment, 2006.

32. For example, PISA test scores in mathematics in 2006 in the United States and France were, respectively, 14% and 10% lower than in the country topping the OECD league (Finland).

Despite this variety of surveys, existing measures of competencies remain limited in important ways. Beyond their limited geographical coverage, some of these tools were developed for the purpose of assessing educational policies, which required focusing on a more narrow set of measurable competencies than those that are potentially relevant for measuring people's capabilities. Even for literacy surveys that have a broader focus than schools *per se*, it has not always proved possible to construct measures of sufficient reliability for a number of competencies (such as teamwork, practical cognition and use of information and communication technologies), and for which assessment would disproportionately increase the response burden and require the use of different tools than self-completed instruments. Despite these limits, some of the surveys developed in recent years have moved beyond a perspective focused on a narrow set of specific competencies, and this trend is expected to accelerate with the development of the OECD *Programme for International Assessment of Adult Competencies* (Box 2.3).[33] More generally, a QoL perspective on education would suggest the importance of looking beyond excellence in specific competencies, to include openness to other cultures, a capacity for self-expression and reasoned discussion, tolerance for other people's views, and the enjoyment that students derive from classrooms.

Box 2.3. The OECD *Programme for International Assessment of Adult Competencies*

The *Programme for the International Assessment of Adult Competencies* (PIAAC) aims to: i) identify and measure differences between individuals and countries in competencies believed to underlie both personal and societal success; ii) assess the impact of these competencies on social and economic outcomes at individual and aggregate levels; iii) gauge the performance of education and training systems in generating required competencies; and iv) help to clarify the policy levers that could contribute to enhancing competencies.

To meet these goals, a questionnaire will aim to measure the interest, attitude and ability of individuals to appropriately use socio-cultural tools, including digital technology and communications tools; to access, manage, integrate and evaluate information and construct new knowledge; and to communicate with others. PIACC will assess adult competencies in the fields of literacy (both prose and documents), reading, numeracy and problem-solving. In addition, PIAAC will collect information from respondents concerning their use of key work-skills in their job (building on the experience of the UK *Skills Survey*) as well as comprehensive background information (*e.g.* educational attainment, family and linguistic background, labour-market status, income, health status, civic participation, literacy and numeracy practices at home and at work, as well as participation in education and training and familiarity with ICT tools). By providing a direct measure of key cognitive skills, PIAAC will give a picture of the distribution of the population's proficiency according to the types and levels of cognitive tasks that they can perform together with the levels of formal education and training achieved. PIAAC will also allow describing changes over time in some of these competencies, by linking to previous international adult skills assessments.

PIAAC will be a multi-cycle programme that covers a range of policy concerns over time. The administration of the survey will take place in 2011, with the results to be released in early 2013; a field trial is foreseen in 2010 with development work occurring over the period 2008-2009. As of December 2009, 29 countries planned to participate in PIACC.

Source: Thorn (2009).

33. PISA: for example, defines reading literacy as the ability of students to use written information in situations which they encounter in their life: this implies a capacity to understand, use and reflect on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society. This definition goes beyond the traditional notion of decoding information and giving a literal interpretation of what is written and is oriented towards more applied tasks. Reading literacy used in PISA is defined by the format of the reading material, the type of reading task or aspects, and the situation or use for which the text was constructed.

A second limit relates to the coverage of existing assessment tools. Schooling is only one of the inputs that lead to individual knowledge, skills development and improvements in QoL. The importance of out-of-school educational experiences is well documented but often ignored in empirical studies, because of data limitations. For example, while some tools allow assessing the non-cognitive aspects of child development (*e.g.* motor coordination, social interactions), few large-scale surveys focus on young children (*e.g.* their caring arrangements, the time spent in them, the structure of these arrangements and the soft competencies that are developed) – despite growing evidence that what is learned in early childhood has significant impacts on later learning outcomes. Similarly, measurement tools remain inadequate to assess the participation and competencies learned in adult education and training, and how these programmes benefit people with different characteristics.

More generally, however, the main problem for indicators in this domain is not the lack of more detailed information on education *per se*, but rather the lack of surveys providing measures of both education and other outcomes that matter for quality of life at the individual level. The availability of such surveys, designed in a way that allows comparisons across countries, would permit a better understanding of the relationship between education and other dimensions of QoL, including the pathways through which education operates and what educators and policy makers can do to enhance the role of education in promoting QoL. The design of effective educational policies and programmes that could have beneficial spill-over effects on other dimensions of QoL crucially depends on knowing what works and what does not, what we could change and what we could do better.

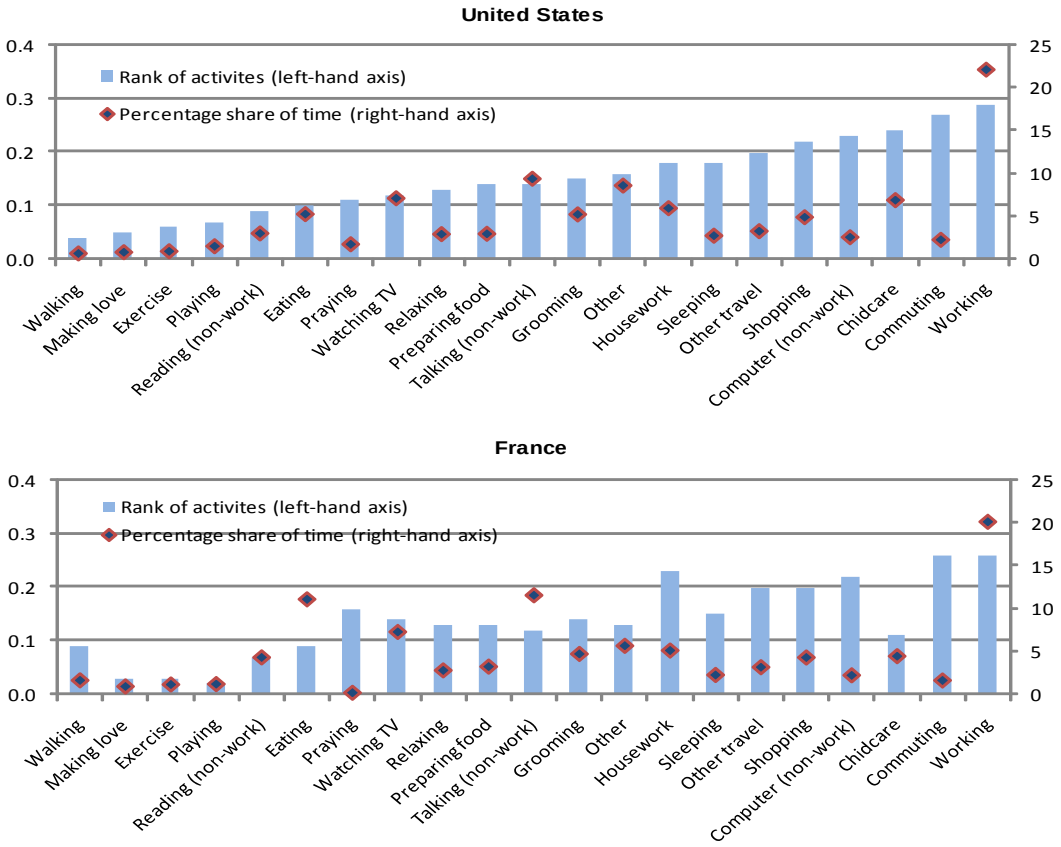
Personal activities

Why do they matter for quality of life?

Personal activities matter for QoL in a many ways. First, and most obvious, the activities that people engage in have predictable effects on both their hedonic experiences and their evaluative judgements. As an example, Figure 2.3 shows the ranking of various daily activities for a sample of women in two cities in the United States and France who report on how they spend their time on a typical day and on the enjoyment they derive from them. In both countries, making love, exercising and playing are the most enjoyable activities, while commuting and (paid) work are the least enjoyable ones; despite some differences between the two countries, these ranking are remarkably consistent. The main activity of each person, such as having a good job or being unemployed, also maps consistently into people's life-evaluations.

Figure 2.3. Ranking of different personal activities based on women's their hedonic experiences and share of time devoted to them, in selected cities in the United States and France

Activities ranked in decreasing order of enjoyment in the United States



Note: The ranking of activities is based on information on the proportion of 15-minute intervals in which the hedonic experience of "stress", "sadness" or "pain" exceeded that of "happiness". Data refer to a sample of women interviewed in Columbus (Ohio, United States) and Rennes (France), interviewed in 2006 with the *Princeton Affect and Time Survey.*

Source: Krueger *et al.* (2008).

More generally, it cannot be assumed that people choose among these various activities in the same way as they allocate their budget among various goods and services, as they may have little say in them (*e.g.* poor families who may prefer to send their kids to work rather than school). Choices over various personal activities can also have externalities on other family members (as in the case of choices about work and leisure) and on people living in the same community (as in the case of commuting). Moreover, the capability approach underlies the intrinsic value of some of these personal activities and their contributions to QoL.

One pressing question is to identify the most important personal activities that affect QoL. Those retained below are paid work, commuting, unpaid work and leisure time. As many of these activities are performed at home, and as this home setting is an important determinant of the quality of non-working life, measures of "housing" are also discussed here. While this list of personal activities is partial, it reflects political demands and the feasibility of providing concrete and comparable measures. Several objective indicators can be used to measure these personal activities and the balance between them at different periods of people's lives (such as the balance between paid work and family life for women with young children). These

indicators can inform not only about the total quantity and quality of various activities but also about how they are distributed among the population at large, particularly by gender.

Before discussing indicators pertaining to each individual activity, a more general consideration is in order. For all personal activities, time-use surveys (TUS) provide a crude but essential quantitative indicator of how people spend their time in various tasks. These surveys also provide essential information from the perspective of integrating non-market work into conventional economic accounts. However, such surveys remain one of the less-developed areas in the statistical system of most countries: they are undertaken only occasionally (if at all), based on protocols that differ significantly across countries, and they lack uniform international standards. Because of their roles in providing direct measures of how people spend their time, such surveys should be undertaken more regularly and be based on frameworks that allow valid cross-country and inter-temporal comparisons.

Paid work

Paid work contributes to QoL both positively and negatively. Paid work provides income as well as identity and social interactions, but it may also be a source of negative experiences and risks. Most prime-aged people in OECD countries have a paid job today, but only a smaller proportion hold jobs that match their inclinations and skills and that provide opportunities for development. While all dimensions are missed by simple employment counts, several national and international organizations provide information on qualitative aspects of paid work. Among these, some of the most relevant for any assessment of QoL are the indicators developed by the *International Labour Office* (ILO) in various studies dedicated to the concept of "decent work". Decent work is an essential determinant of QoL, and ILO tries to promote this concept in all the world's countries.

Decent work is a multidimensional concept, which includes all the elements listed in Box 2.4. These elements, which also contribute to the economic and social security of workers and their families, include the balance between paid work and other activities (as in the case of "combining work and family life") and a dimension of social dialogue (*i.e.* the collective conditions that shape the QoL and well-being of workers in the workplace). Also, decent work refers to rights, whether or not they exist yet (*i.e.* are internationally recognized) or have been institutionalized.

Box 2.4. Substantive elements of the "decent work" agenda

- Employment opportunities (1 + 2)
- Work that should be abolished (1 + 3)
- Adequate earnings and productive work (1 + 3)
- Decent hours (1 + 3)
- Stability and security of work (1, 2 + 3)
- Combining work and family life (1 + 3)
- Equal opportunity and treatment in employment (1, 2 + 3)
- Safe work environment (1 + 3)
- Social security (1 + 3)
- Social dialogue and workers' representation (1 + 4)

Numbers in parenthesis refer to the ILO strategic objectives to which the elements of "decent work" are related: 1. Standards and fundamental principles and rights at work; 2. Employment; 3. Social protection; 4. Social dialogue.

Source: ILO compilation

Measures of decent work can be compiled from a variety of sources. Beyond the various reports and data collections undertaken by the ILO, the *European Foundation for the Improvement of Living and Working Conditions* (Dublin Foundation) conducts surveys every five years on working conditions in the EU and accession countries.[34] These surveys (whose practical use is, however, limited by the small sample size for each country) provide data on many aspects of decent work, such as non-standard employment, gender gaps in employment and wages, discrimination in the workplace, lifelong learning and training, employment of disabled persons, working time and "unsocial hours", perceived work-life balance, work accidents and physical risk factors, work intensity, work-related health problems, social dialogue and workers' autonomy. National surveys of working conditions also exist in several countries. These should be extended to other countries in the world, based on a consistent design, so as to allow measuring both less-tangible aspects of paid work and the diversity of experiences across workers.

Each of these aspects of "decent work" needs to be measured with several potential indicators. While not everything can be included, the Commission recommends retaining a limited number of key indicators linked to the concept of decent work, to be selected in consultation with relevant stakeholders and international agencies. These indicators should be adapted to both developed and developing countries, and highlight the gender division of tasks and risks in special sectors such as agriculture. Some jobs carry greater health consequences than others: for example, transplanting during rice cultivation implies exposure to water-borne diseases and to chemicals, and women – who do most of this work – retain toxic chemicals for longer. Measures of "decent work" should also be developed at the sub-national and regional level.

The measurement difficulties in this field are well illustrated by workplace accidents, which represent the most visible manifestation of the hazards of paid work. Work accidents are sudden and sometimes violent events that occur during the execution of work and lead to health damage or loss of life for the worker involved. International comparisons of work accidents are difficult because of differences both in record-keeping – *e.g.* statistics sometimes only record "compensated" accidents in workplaces of a sufficient size and exclude minor injuries – and in data-sources – insurance companies, social security registers, labour inspectorates, establishment censuses and special surveys. Moreover, firms may under-report work accidents in countries where the social contributions that they pay depend on the frequency of accidents. Comparability has improved since the adoption in 1998 of an ILO Resolution on "Statistics on occupational injuries resulting from accidents at work", which sets out standards for data collection and presentation, but problems remain. For example, the reporting of non-fatal injuries is limited to injuries causing absences from work of more than three days in European countries and Japan, of six or more days in Australia, and of one or more days in other countries. In all countries, available statistics exclude absences that cause lower working hours rather than an outright absence from the workplace.[35] Changes in the

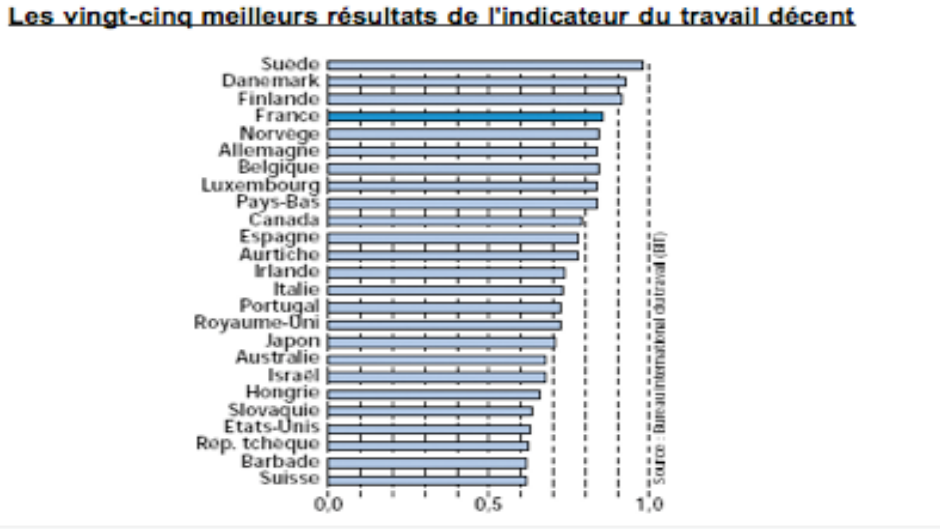
34. See, for example, the *Annual Review of Working Conditions in the EU 2007–2008*; the *Fourth European Working Conditions Survey, 2007*; and the *First European Quality of Life Survey: Time Use and Work–Life Options over the Life Course, 2007*.

35. In 2003, fatal work accidents were most frequent in Turkey, Korea and Mexico and least frequent in the United Kingdom and Sweden; their frequency has declined since 1995. Non-fatal accidents are more common, and they have also declined in most OECD countries (OECD, 2006a).

frequency of reported accidents from administrative and insurance sources might also reflect changes in insurance rules that alter employers' incentives to under-report minor accidents or to offer reduced work hours to injured employees. All these differences in reporting practices limit statistical comparability in this field.

Because of its multidimensional nature, several attempts have aimed at constructing composite indices of decent work. Figure 2.4 provides an example that shows a country ranking among the 25 countries with the best rank, as published in *International Labour Review* (2003). This composite index is based on seven sub-indicators (themselves composite), whose choice was partly determined by data availability: i) security of the labour market (various indicators of the availability of jobs); ii) employment security (the opposite of precariousness); iii) professional or skills security (availability of skilled jobs); iv) security at work (working conditions, including hours worked); v) security linked to education and training; vi) income security; and vii) workers' voice. Ways of conveying information on complex, multidimensional concepts such as decent work through composite measures should be more systematically explored.

Figure 2.4. A composite measure of decent work



Source: International Labour Review, 2003.

Unpaid domestic work

Most of the existing indicators used to inform about QoL neglect issues related to unpaid domestic work, such as shopping, family chores, and care of children and other dependent members of the household. This is an important omission. As for other personal activities, a direct measure of the quantity of time devoted to these domestic chores could be compiled based on time-use surveys. Such data, in turn, would allow computing measures of the inequality between men and women in the distribution of these unpaid domestic activities. These data would also allow better comparisons over time and across countries, and shed lights on many other inequalities (for example, in terms of employment and leisure).

In principle, indicators should also allow assessing the quality of unpaid domestic labour, although few objective criteria exist in this field. However, some countries (such as Sweden) undertake regular surveys of the stress and strain of domestic work and of their consequences

for health. Developing direct measures of these strains would also be important for developing countries, as both the technology (open or primitive stoves) and the fuel used for home cooking (bio fuels) carry much higher health risks for women and children than for men.

Unpaid domestic work is especially important for the QoL of families with young children. As most of the time spent caring for children is usually delivered while performing other tasks, this time is poorly measured in time-use surveys that do not comprehensively record "secondary" activities. Also, as most young children spend their time in a variety of child-care arrangements (formal and informal), measures are needed of how these arrangements are combined (i.e on the share of time that children spend in each). Indicators of access to childcare facilities and its costs are also important to assess the time-crunches faced by parents of young children and to better measure the costs of children.

Commuting

People with paid jobs have a higher QoL when their commuting time is lower. In several OECD countries, the combination of higher house prices and stagnant earnings has recently pushed many workers and their families to live further away from their places of work. This development will weigh heavily on their QoL. A partial but essential indicator of commuting is the number of hours spent travelling to and from work during a specified period (week, month). This type of indicator could be routinely compiled from regular TUS. The information already available shows large differences in the amount of commuting time across countries (*e.g.* employed people spend 70 minutes per day commuting, or the equivalent of 18 full days per year, in Korea, 36 minutes in France and 29 minutes in the United States).

Other dimensions of commuting besides its length are also important for QoL. These include the accessibility of transport and its affordability, which may affect people's right to daily mobility. Measuring accessibility and affordability is complex. While accessibility is partly captured by measures of commuting time, measuring affordability would require some measure of household costs and some benchmark to which these costs are compared. Commuting costs may also reflect preferences, as some people may opt for more costly private cars over cheaper public transport. Also important to a sustainability perspective are modes of transport, *e.g.* availability of bicycle lanes, etc. But statistical feasibility is a concern.

Leisure time

Leisure time is a critical dimension of QoL, and a long tradition of research (going back to Tobin and Nordhaus) has aimed at integrating estimates of its "value" within a broader measure of money-based well-being. While these efforts raise difficult questions in terms of the appropriate price for valuing leisure, the difficulties in identifying suitable indicators are also important for measuring its quantity. In fact, despite much research about the importance of leisure and recreation in shaping QoL, few objective measures of leisure time enjoy widespread use.

Some possibilities do, however, exist. The most obvious is a simple measure of the quantity of leisure time and of the differences between groups and individuals (*e.g.* by gender) in how it is distributed. Estimates of this quantity may be derived from the time-use surveys

that are periodically conducted in various countries, but even this raises some problems. One problem is that estimates of leisure time are based on somewhat arbitrary classifications of the various activities that survey participants report in their time diaries. A second problem is that the amount of time that people allocate to "necessary" activities (such as sleeping, eating and personal care) differs across people and countries, and that these differences will affect the estimates of the quantity of leisure time enjoyed in a typical day. When corrected for this factor, available estimates highlight some sizeable cross-country differences in leisure time, both across countries[36] and across groups (men enjoy more leisure time than women, with large differences in Italy and negligible ones in France; leisure-time is also U-shaped with respect to age, OECD, 2009). A last problem, already raised when discussing hedonic experiences, is that different people (*e.g.* people who are involuntarily unemployed and people with jobs that they like) will derive very different "pleasure" from the same activity classified as "leisure time" in conventional time-use surveys: measures of both the amount of time spent in various activities and of the degree of enjoyment derived from them are needed to account for inequalities in leisure time across people.

Other indicators of the quantity of leisure time also exist. One example is provided by indicators of participation/attendance in various leisure activities, such as recreation, sports and cultural events. These indicators are available in several countries, and could be collected through comparable surveys, although no common classification or dedicated survey instrument exists in this field. A different venue could take the form of indicators of "poor leisure", such as the proportion of individuals, families or children that cannot afford a week of holidays away from home at least once a year. This type of indictor has a long tradition in European countries but remains rare elsewhere.[37]

Beyond its "quantity", measures of the "quality" of leisure time are also important, as they may shed light on some important differences over time and between groups. For example, in the United States highly-educated individuals have less leisure time than less-educated individuals (*i.e.* they are "money-rich" but "time-poor"). However, when considering other measures of leisure quality (such as the number of leisure episodes or whether leisure is undertaken with other persons or combined with other activities), the quality of leisure appears to be higher for the highly educated (Gimenez Nadal and Sevilla Sanz, 2007). Research also shows that, even when men and women have the same amount of free time, its quality may be uneven, especially when fears of physical insecurity or assault push women to stay at home rather than go outside. While issues related to leisure time may be less pressing for poor countries, these are critical concerns for developed countries.

Housing

Housing is an important determinant of QoL as well as of a variety of other social outcomes (such as children's education). It is hence important to identify indicators describing both its quantity and quality that could be used to benchmark countries'

36. Among OECD countries, the amount of leisure time in a typical week is lowest in Mexico and Japan (at less than 5 hours per day) and highest in Belgium and Germany (at more than 6½ hours per day).

37. Among EU countries, close to 10% of households in the Netherlands and in most Nordic countries report that they could not afford a week away from home, as compared to levels above 50% in some countries in Southern and Eastern Europe.

performance in this field. Housing statistics are indeed available in most countries, and are also collected by some international and regional organizations. Eurostat publishes reports and data on housing within the framework of its "urban audit"; and information is also collected by various UN agencies.[38] But comparability in this field is a challenge, and no core set of housing indicators currently exists for international comparisons of QoL.

Developing a set of housing indicators that could be used in the context of QoL studies would require taking steps in several dimensions:

- First, all countries would need to develop appropriate measures and definitions of people who are homeless or using emergency shelters. As people in these conditions are unlikely to be reached through regular surveys, countries need to use the information available through public programmes and voluntary groups that provide emergency relief to people in these conditions.
- Second, national and international statistical offices need to develop and make operational a concept of "decent housing" (the negative side of which is "poor housing"), with different specifications for developed and less-developed countries. Decent housing indicators would help secure recognition of the importance of housing rights worldwide.
- Third, measures of housing quality should be put in place. The criteria for quality should differ between countries at different stages of development. For example, recent research in Brazil describes adequate housing through (census) data covering overcrowding (more than two persons per bedroom); availability of safe drinking water; waste collection; and sewage disposal facilities. In developed countries, suitable indicators could include, in addition, units lacking complete plumbing, waiting time for subsidized housing, and exposure to noise and pollution. Information on foreclosures on primary residences during the ongoing financial crisis should also be gathered.

Political voice and governance

What is it and why does it matter?

Political voice is an integral dimension of QoL, having both intrinsic and instrumental worth. Intrinsically, the ability to participate as full citizens, to have a say in the framing of public policy, to be able to dissent without fear and to speak up against perceived wrongs – not only for oneself but also for others – are essential freedoms and capabilities.[39] Instrumentally, citizens' voices can provide a corrective to public policy, ensure the accountability of governments and public institutions, reveal what people need and value, and call attention to significant human deprivations both in emergency situations (*e.g.* famine, floods and hurricanes) and over the long term (*e.g.* poverty, malnutrition and hunger). Institutions that promote participation and public discussion help citizens to make informed choices on many aspects that impinge on their QoL (such as health, education, the

38. See, for example, the UN sustainable indicators programme (http://www.un.org/esa/sustdev/natlinfo/indicators/isdms2001/isd-ms2001isd.htm) and the reports of the UN Centre for Human Settlements (*e.g.* on the causes of inadequate housing in Latin America, http://www.habitat.org/lac_eng/pdf/causes.pdf).

39. On aspects of this, see also Sen (1999), Nussbaum (2006), Dreze and Sen (2002), and Alkire (2002).

environment, legal rights and civic duties) and to play a role in shaping public values.[40] Political voice also reduces the potential for social conflict and enhances the prospect of building consensus on key issues, which can increase economic efficiency, social equity, and inclusiveness in public life. Political voice can be expressed both individually (such as by voting) and collectively (such as joining a protest rally). In both forms, the opportunities for expression and the degree of responsiveness of the political system (*i.e.* the extent to which the individual's voice is actually heard and acted on) will depend on the institutional features of each country. Prominent among these features are the presence (or absence) of a functioning parliamentary democracy, universal suffrage, a free media, and civil society organizations.[41]

Legislative guarantees and the rule of law impinge on and enhance the scope of political voice, but they are also independently important and, once again, have both intrinsic and instrumental worth. Intrinsically, the legislative guarantees provided by a government indicate to its citizens how fair, just and humane is its vision of society, and how low is its tolerance for human deprivation and the lack of basic freedoms. Such guarantees can take various forms, such as constitutional rights prohibiting discrimination by gender, race, religion or ethnicity, and social rights to education, housing, old-age pensions and health care. In addition, laws outside the constitutional framework can provide rights that enhance material and non-material welfare. Cases in point are laws that promise a living wage and various forms of social security, protection for women against domestic violence, property rights, and a right to information that citizens can invoke to ask public institutions about their functioning and hold them accountable for their decisions. Whether or not a person or group uses the available legal guarantees in practice, the very fact of living in a society that promises such guarantees defines the opportunities available to every citizen and thus impinges on their *perception* of their QoL. Instrumentally, legislative guarantees can affect both the economy and society. The structure of laws, for example, can affect a country's investment climate and thus have an impact on market functioning, economic growth, job creation and material welfare.

To realize their potential, however, the guarantees promised by the law require effective implementation. They require a rule of law that can ensure the provisioning of *substantive* justice and legal implementation, by going beyond paper guarantees to their practical realization. Unfairly or inadequately implemented laws will not only fail to enhance people's QoL but will undermine trust in the state both within the country and internationally. Effective implementation, by contrast, will benefit both actual well-being and the perception of well-being. Successful implementation depends, however, on how various state institutions, such as the police, the judiciary and other administrative services, function; whether they are free from corruption, political interference and social prejudice; and whether they can be held accountable for their decisions.

40. According to Sunstein (1991) and Young (1993), the process of deliberation can also produce other-regarding preferences among the privileged.

41. Using data for 46 countries, Owen *et al.* (2008) find a positive correlation between democratic institutions and individual levels of subjective well-being. Similar evidence, referring to Swiss cantons, is reported by Frey and Stutzer (2002a and 2002b).

The accountability of the police and the judiciary is critical to the enforcement of the rule of law, while accountability of administrative services is crucial for non-judiciable claims to public policies and programmes. Moreover, the attitudes of judges towards the disadvantaged, their perception and understanding of people's lives and circumstances, are important for how the laws are interpreted and whether the legal process provides substantive justice.[42] Politically appointed judges can bring pre-determined positions and attitudinal biases to the implementation process. Successful implementation of the law also depends on mechanisms to ensure that citizens are aware of their legal rights and that legal aid is provided for those unable to afford the cost of litigation. Civil society institutions and a free media can also prove crucial in holding public institutions accountable and in promoting the rule of law.

Political voice, legislative guarantees and the rule of law matter not only individually but also interactively. Political voice can lead to the establishment, formulation and expansion of a legal framework that embraces all citizens, by calling attention to missing or faulty laws, and lobbying the state to change discriminatory or ineffective laws and instituting new ones. Citizens with freedom of voice and assembly can also put pressure on governments for the fair implementation of laws, and identify cases where a law has been unfairly applied or justice denied. Legislative guarantees in turn can strengthen political voice by providing the right of free speech, free association and assembly, and the free exchange of information. All these factors matter both in developed and developing countries.

Indicators

Indicators are needed to measure where a country stands in terms of political voice and democratic governance, as well as legislative guarantees and the rule of law.

- Citizens can assert voice through a range of institutions and rights – multiparty parliamentary democracy and universal suffrage; decentralized government; free media; academic freedoms; and freedom to form and join civil organizations, social institutions, trade unions and professional bodies. Citizens' voices can thus be expressed through both political institutions and institutions that fall outside the remit of formal politics. In addition, indicators are needed to assess what rights, freedoms, opportunities and voice can be exercised by non-citizens – whose numbers have been rising in many countries with growing international migration.

- Legislative guarantees include rights embedded in constitutions (*e.g.* non-discrimination on the basis of race or gender; right to food, education, etc.); laws that promote civil and criminal justice, equality, inclusion and accountability (*e.g.* freedom of information); laws on affirmative action; international covenants that uphold principles of human rights and basic freedoms ratified by a country; and other legal guarantees of various kinds.

- The rule of law includes institutions that ensure *de facto* rights and not simply *de jure* rights. An independent judiciary, free from corruption and political influences, the speed with which justice is delivered, economic and social access to the law for all citizens (including women and religious and ethnic minorities) are all potentially informative indicators.

42. For elaboration, see Nussbaum (2007).

A list of useful indicators on the dimensions mentioned above is provided in Table 2.1. Each of these dimensions could potentially be measured through both objective and subjective indicators (*i.e.* people's perceptions of the items listed). In using subjective indicators, however, caution is needed to avoid the possibility that perceptions may provide a more positive reading of a situation than actual experience might indicate (or vice versa). A substantial literature on adaptive preferences highlights, for instance, that disadvantaged people may adapt their preferences to their unfavourable circumstances and claim to be better-off (*e.g.* in better health) than indicated by their objective conditions[43].

Table 2.1. Potentially useful indicators to measure political voice, legal guarantees and the rule of law

Citizens' Voice	Legislative Guarantees	Rule of Law
Democratic participation : Multiparty democracy; universal suffrage; free, fair and regular elections; voter turnout	*Constitutional guarantees:* number of guarantees built into the constitution	*Judicial institutions* : existence of an independent judiciary
Participation in governance : Decentralised institutions of governance; representation of women, minorities, etc. in political and executive institutions of government	*Legal guarantees for basic economic needs* : employment, food, education, health care, housing, etc.	*Equal access to and treatment for all:* by ethnicity, religion, race and gender, etc. (needs sub-indicators of access)
Support from outside government : Existence and number of secular civil society organisations; free media	*Political and social rights guaranteed by law* : right to information, freedom from domestic violence, etc.	*Functioning of legal institutions* for civil and criminal redressal
	Ratification of international treaties for equality, human rights, etc.	*Arrests and detention of political prisoners*

Despite the importance of political voice for QoL – and the array of international organizations and universities that have identified indicators and collected information on them, some on a regular basis and others occasionally – reliable measures remain limited. Most of the existing indicators are compiled by bodies outside the boundaries of national statistical systems (Appendix 2.1). The information collection methodology typically involves drawing on expert opinions about how countries are performing in terms of democracy, corruption and freedoms. Cases in point are the *Freedom House* and *Polity IV* indicators, and the set of governance indicators regularly compiled by the *World Bank Institute* (Kaufmann, Kraay and Mastruzzi, 2008)[44]. While experts' assessments are useful in some fields, such as concerning the existence of particular institutions of governance or legislative guarantees, they are also clearly inadequate for assessing how adequately or fairly such institutions function, or how people perceive them. To measure these aspects requires population surveys that provide information on citizens' perceptions of the functioning of

43. For conceptual discussions on adapted preferences, see Sen (1987) and Nussbaum (2001); for an empirical assessment, see Burchardt (2005). On how people's preferences are shaped by existing endowments, see Sunstein (1993).

these institutions. Such surveys are rare[45]. Also, there is very limited information on inequalities (predicated on race, gender, ethnicity or minority status) in access to legal redress, on differences in perceptions of the ways in which political, legal and executive institutions function, and on the trust that disadvantaged groups in the population place in these institutions.

Comparisons based on existing indicators of political voice, democratic governance, legislative guarantees and the rule of law highlight vast differences between countries. Although many countries have moved away from dictatorship and authoritarian regimes over the past two decades, their transition towards establishing the full range of democratic freedoms, rights and the rule of law has been difficult and slow. Effective indicators can help monitor the consolidation of democracy in these countries: to this end, several recent initiatives are aimed at improving these countries' capacity to develop indicators in these fields, which could be used to formulate and evaluate national policies.[46]

Available indicators that focus on political institutions show little or no variation among the established democracies.[47] This, however, is more a reflection of the data collected than of the views of their residents (Ringen, 2007). Even in the developed world, 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. For example, in a survey of OECD countries, on average only 44% of respondents reported high trust in the civil service, and less than 40% reported high trust in the legislative and executive branches of government, so in some countries the levels were even lower (OECD, 2006a). The exercise of political voice also differs systematically across countries and groups of people: for example, younger, less-educated and lower-income people are less likely to vote (OECD, 2006a); and, with higher immigration, rising numbers of non-residents lack fundamental rights and opportunities for political participation. The functioning of political systems in developed countries also provides plenty of examples of how rights are transgressed, such as by unequal representation of people before the law, biases in legal procedures, the purchase of political power through direct payments to decision-makers and contributions to campaign financing, and lobbying by sectoral interest groups (Okun, 1975). Suitable indicators are thus needed to assess the quality of democratic governance within established market economies as well.[48]

---

44. The *Polity IV* project measures the "degree of democracy" on a year-by-year basis for all independent states on a scale from +10 (most democratic) to -10 (least democratic). The indicators cover: i) the existence of a functioning central political authority; ii) openness or closedness of political institutions; iii) durability of the central political authority (number of years since the last regime transformation); iv) institutionalised procedures for the transfer of executive power; v) competition in executive recruitment; vi) independence of the chief executive; vii) institutional structures of political expression; and viii) competitiveness of participation.

45. One example is the survey undertaken by "Reporters without Borders" to measure press freedom. At the regional level, another example is provided by the survey conducted in South Asia to map citizens' opinions and attitudes (SDSA, 2008): this survey includes questions about the perceived gap between constitutional guarantees and their realization, perceptions about political and executive institutional functioning, freedom from fear or from a feeling of insecurity, and freedom from want (hunger, poverty).

46. Prominent among these initiatives are the *Metagora* project, launched in 2004 under the auspices of the OECD/PARIS21 network (OECD, 2008a); and the UNDP *Global Programme on Capacity Development for Democratic Governance Assessments and Measurements.*

47. Among the 28 OECD countries covered in Polity IV, one has a score of 7, four have a score of 8, one has a score of 9, and the 22 others all have a score of 10.

Social connections

Like political voice and the rule of law, social connections and the attendant norms of trust and trustworthiness are important for people's QoL. These social connections are sometimes subsumed under the heading of "social capital". Definitions of social capital (as were other forms of "capital" at an equivalent stage in their conceptual development) have been much debated, but there is now convergence towards a "lean and mean" definition: social networks and the associated norms of reciprocity and trustworthiness.[49] Since it is impractical to measure social networks at large geographic levels, researchers generally rely on proxies for these networks (*e.g.* number of close friends, political participation, membership in voluntary associations, religious involvement, doing favours, etc.). The core insight of the concept of social capital is that, like tools (physical capital) and training (human capital), social connections have *value* for QoL.[50]

Social connections have value, first, to the people who are in the networks. For example, labour markets are permeated by networks, so that most people are as likely to get their jobs through *whom* they know as through *what* they know. Similarly, social connections bring benefits for health: as a risk factor for premature death, social isolation rivals smoking (Berkman and Glass, 2000). Evidence also suggests that social connections are powerful predictors of (and probably causes of) subjective well-being. Finally, the same personal activity can have different impacts on subjective well-being, depending on whether it is conducted alone or with others.

All these are "internal" effects of social networks, since they represent ways in which social networks benefit people *in* those networks. Social connections, however, also have "externalities", *i.e.* implications for bystanders. The literature on "social capital" has brought out clearly a number of examples of positive externalities. For example, neighbourhood networks can deter crime (Sampson 2003), and this effect also benefits residents who sit at home in front of their TV. The performance of democratic government and even the pace of economic growth may also depend on the quality of social connections within a jurisdiction.[51]. Finally, several (mainly US) studies suggest that both child welfare (infant mortality, teen pregnancy, low birth-weight babies, teen drug use, etc.) and school performance (drop-out rates, test scores) are robustly predicted by measures of community social capital.

48. Ringen (2007) describes differences in democratic governance among developed market economies through indicators of the *strength* of the democratic process (*i.e.* the timing of introduction of universal suffrage, and a measure of freedom of the press); of its *capacity* (*i.e.* a composite indicator of government effectiveness and a qualitative indicator of protection against the political use of economic power); of the people's *trust* (*i.e.* trust in parliament and the civil service); and of the *security* that the political system provides to people (*i.e.* against risks of poverty and the affects of illness).

49. While various forms of "social trust" fall outside this narrow definition, they are included in discussions of social capital because they are common, empirically important by-products of dense social networks. Generic measures of social trust are widely used as proxies for social capital in international comparisons.

50. A rich literature also underscores the value of networks for economic development and the functioning of markets.

51. These claims, first suggested by Putnam (1993), have been confirmed by a number of more rigorous researchers, both in Italy and elsewhere.

However, the "externalities" stemming from networks can also be negative. A strong sense of belonging to one group can strengthen a sense of a unique personal identity in terms of the group to which he or she belongs (Sen, 2006). This may generate fissures in national communities, and breed a climate of violence and confrontation. More generally, a long-standing tradition in economics has stressed the potential for groups to generate benefits for insiders that weigh heavily on the opportunities and QoL of outsiders. In other words, groups can foster bonds among participants but also erect walls with respect to outsiders and members of other groups. To account for these multiple effects, research distinguishes between two types of social capital, "bonding" and "bridging", but practical implementation of this distinction in empirical research remains a challenge.

In short, a rich literature from several disciplines shows that social connections benefit people in the networks, with effects on non-participants that depend on both the nature of the group and the effects being considered.[52] In some cases, like health, studies have demonstrated that social connections can have positive effects at both the "individual" and "aggregate" level: people with more friends live longer in part because of the biochemical effects of social isolation, and in part because public health systems are more effective in areas of higher social capital. Many of these "causal" claims are yet to be tested with an experimental or quasi-experimental design, but even on this score progress is slowly being made. A high priority for research in this field is more work on causal linkages using natural or randomized experiments.

Social connections increase quality of life

Much evidence at both the aggregate and individual level suggests that social connections are among the most robust predictors of subjective measures of life satisfaction. Social connections have a strong independent effect on subjective well-being, net of income.[53] Moreover, the available evidence also suggests that the externalities of social capital on well-being are typically *positive*, not negative (Helliwell, 2001; Powdthavee, 2008). In other words, increasing my social capital increases both my own and my neighbours' subjective well-being, and thus represents a coherent strategy for improving QoL for the country as a whole.

The analysis of the effects of social connections on subjective well-being is in its infancy. Much of it does not account for unmeasured individual characteristics, and most of it relies on cross-sectional data. That said, recent analyses have strengthened the case that the link between at least some forms of social connections and subjective well-being is causal. Krueger, Kahneman *et al.* (2008) [54] report that, when controlling for individual fixed effects (such as personality traits), most pleasurable activities involve socializing — religious activities, eating/drinking, sports, and receiving friends. Similarly, in a recent large-scale US

52. For a summary of some of this literature, see Putnam (2000).

53. Helliwell and Putnam (2004) summarize their review by saying: "Social connections, including marriage, of course, but not limited to that, are among the most robust correlates of subjective wellbeing. People who have close friends and confidants, friendly neighbours and supportive co-workers are less likely to experience sadness, loneliness, low self-esteem and problems with eating and sleeping…[S]ubjective wellbeing is best predicted by the breadth and depth of one's social connections. In fact, people themselves report that good relationships with family members, friends or romantic partners — far more than money or fame — are prerequisites for their own happiness."

panel survey on religious attendance and subjective well-being, Lim and Putnam (2008) found that religious attendance at $time_1$ (or $time_2$) predicted subjective well-being at $time_2$, controlling for levels of subjective well-being at $time_1$, as well as many other covariates; the essential mechanism involved in this relation is neither theological nor psychological, but rather the strong effect of "friends at church" on well-being.[55] Fowler and Christakis (2008) also report evidence suggesting that subjective well-being can spread in a beneficially "contagious" way from one person to another. For no other class of variables (including strictly economic variables) is the evidence for causal effects on subjective well-being probably as strong as it is for social connections.

Indicators

As research on social connections is relatively new, national statistics are still rudimentary. Most researchers have relied on unofficial sources. One proxy of social connections often used is the number of associations in civil society or church to which each person belongs. However, the fragility of such measures is, by now, well recognized. A formal organization with a name and address may not correspond to any actual individual members, much less to social networks among those members. Moreover, the role of associations differs from country to country. Because of these reasons, measures of organizational density are generally *not* good measures of social connections, despite their frequent use for that purpose.

A related approach measures the activities assumed to be the result of social connections, such as altruistic behaviour. Thus, some research has used blood donations, membership in voluntary organizations or charitable giving as proxies. Other studies have relied on some aggregate measures of individual behaviour like voting turnout, based on the argument that, even though balloting is a private activity, participation in voting is higher in countries with a dense network of political parties or civic organizations, and that in all countries members of these organizations have a higher probability of voting. Other studies have used proxy measures of social connections based on information on family ties, such as marriage rates, though social changes throughout the world have made these an imperfect measure of enduring interpersonal ties.

Ultimately, however, all these indicators are inadequate proxies of social connections, and reliable indicators can only be constructed through survey data. Only personal reports allow measuring the many and evolving forms of social connectedness. In recent years a number of statistical offices (in the United Kingdom, Australia, Canada, Ireland, the Netherlands, and most recently, the United States) have begun to gather and report survey-based measures of various forms of social connections. As an example of these endeavours, Appendix 2.2 presents the list of the questions included (since early 2008) in an annual Supplement to the November US *Current Population Survey,* which has traditionally probed respondents about voting in national elections.[56] These questions have been selected after extensive vetting by

54. The paper used random time-sampling to elicit emotional reactions to various activities. Other cross-sectional analyses have controlled for personality types and found persistent effects of social capital on well-being, *e.g.* Helliwell (2005).

55. Chaeyoon and Putnam (2008) report that while SWB generally rises with the number of one's "close friends", friends at church (at least in America) seem to be "super-charged" friends, in the sense that the effect size on SWB is roughly twice as great as for "close friends".

the Census Bureau and the Bureau of Labor Statistics for reliability, intelligibility, and inoffensiveness; they cover several manifestations of civic and political engagement, as well as other forms of social connections (such as number of friends, or frequency of contacts and favours done for neighbours).[57]

Measuring social connections, however, goes beyond measuring these particular aspects. Also needed are suitably tested questions assessing people's experience in a range of other domains. Some of the most important domains that could be explored through dedicated surveys include the following:[58]

- ***Social trust***. Despite only moderate test-retest reliability, the canonical social trust question has been asked thousands of times in many countries: its behaviour is well understood, and it allows many comparisons across time and space.[59] Moreover, at the aggregate level (*e.g.* states/nations), responses are highly stable over time (even when individual-level stability is low), suggesting that this question measures a very predictive characteristic of communities. Data on social trust are also significant determinants of subjective well-being. Compared to this canonical question, questions about "lost wallets" are potentially more reliable, since they are more specific and quasi-behavioural.[60] Opportunities for comparison across time and space are, however, scant, and we still lack studies of its variability across time for the same person.

- ***Social isolation.*** Lack of contacts with other people in normal daily living is both a symptom and a cause of social distress, and it can lead to a downward spiral affecting morale and reducing social and economic opportunities. Social isolation can be measured through questions asking people about the frequency of their contacts with others or about how often they spend their time socializing with family members, friends and work colleagues or with other people in sports, religious and cultural associations. Social connections are also a function of living arrangements (*i.e.* living alone) and employment status (*e.g.* having a job). Research has highlighted strong associations between the degree of social isolation of each person and measures of their well-being, self-assurance, ability and power of action, and activity (Ringen, 2008)

- ***Informal support.*** Questions about the availability of social support in case of need have been used in many countries. The *Gallup World Poll* includes a yes/no question about friends or relatives "you can count on"; answers to this question are highly

56. These questions about political engagement might, in fact, be best seen as measures of "political voice", as discussed earlier in this report.

57. Questions on local voting are often found to be more reliable and to display higher variance than voting in national elections, probably because it is less subject to "social desirability" response bias. Empirically, such questions have proved to be among the best single measure of political engagement more generally.

58. This list largely overlaps with that proposed by an *ad hoc blue ribbon* interdisciplinary working group, and reproduced in Appendix 3; these items (and the associated questions) have not yet been accepted by the Census Bureau for inclusion in the *Current Population Survey*.

59. "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"

60. "If you lost a wallet or a purse that contained two hundred dollars, and it was found by a neighbour, how likely is it to be returned with the money in it? Would you say very likely, somewhat likely, somewhat unlikely or not at all likely?"

predictive of subjective well-being but have little discriminatory power (about 90% of respondents answer this question affirmatively). This suggests that alternative formulations or more nuanced responses may be needed (*e.g.*, "if yes, how many?", or specifying the type of help expected in various specific contingencies).

- ***Workplace engagement.*** For many people in many countries, a large fraction of all their social connections are with workmates, either in the workplace or outside of it. Helliwell and Huang (2005) found that trust in workmates is a robust predictor of life satisfaction. Questions about trust in workmates have been tested in some countries (*e.g.* North America).

- ***Religious engagement***. While there is general agreement that religion is an important form of social connections, the introduction of specific survey questions (even when extensively vetted) raises issues of political sensitivity. Identifying suitable formulations of these questions is important as, in most countries, religious engagement is a robust predictor of subjective well-being and (in many countries outside Europe) it is a large fraction of all social networks.

- ***Bridging social capital*** (*i.e.* friendships across lines of race, religion, class, etc.) is the most important under-measured form of social connections for many outcomes. The informal advisory committee to the US *Current Population Survey* recommended that a suitable measure of bridging social capital could take the form of follow-on probes of the form, "Of these close friends, how many are... (White, Black, Asian, Hispanic, Catholic, Jewish, college-educated, etc.)". This bridging question is least susceptible to response bias and political correctness.[61]

Aggregation

One important issue for the measurement of social connections is aggregation. Social connections come in many forms that are useful in many different contexts. These forms, however, are heterogeneous and not fungible, *i.e.* they are good only for certain purposes and not others. Further, the exact pathways between networks and benefits (both internal and external) differ from case to case, and much work remains to be done on those mechanisms. Last, although networks can affect our ability to get things done, nothing guarantees that what gets done will be socially beneficial.[62]

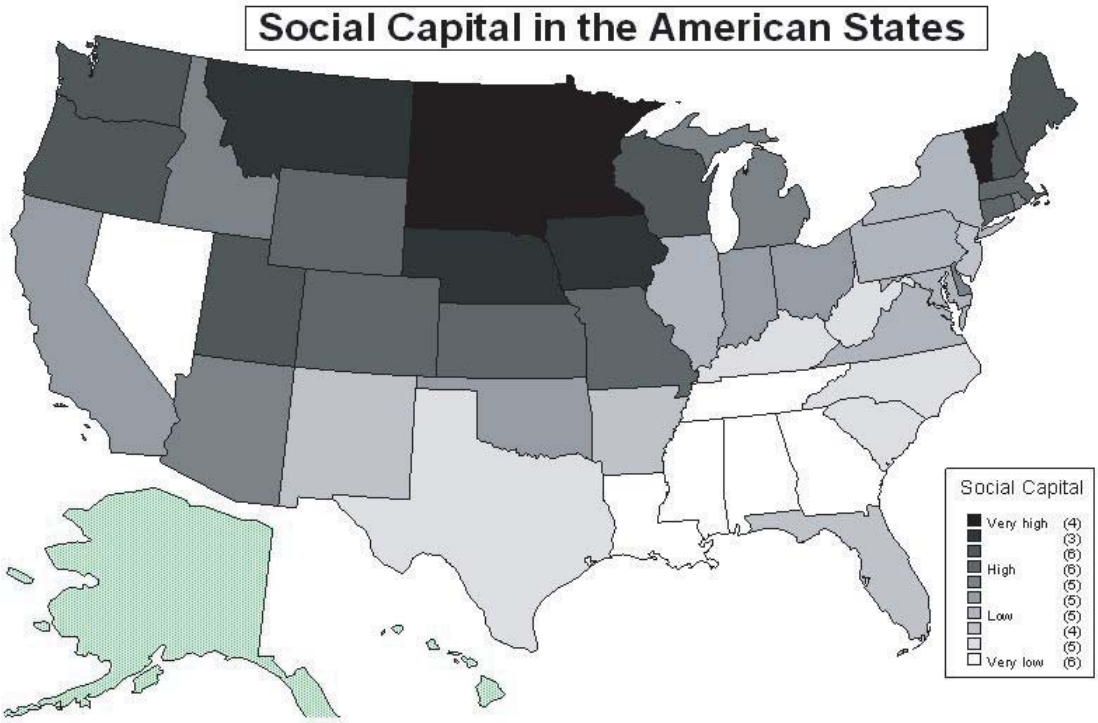
Because of these considerations, it is not obvious whether all different forms of social connections can be simply "added up" to produce a summary for a given community or country. Nevertheless, composite indices, based on various indicators, can sometimes generate useful information about differences in social capital across space. Figure 2.5 shows a composite measure of social capital across US states based on 14 different indicators.[63] While these measures are fairly closely inter-correlated, this state-level index is a strong

61. Appendix 2.3 presents the list of questions concerning social trust, religious engagement and local voting that, while not included in the CPS supplementary module, could be used for measuring these other forms of social connections.

62. Al Qaeda, for instance, is an excellent example of social capital used for disruptive purposes, enabling its participants to accomplish goals that they could not accomplish without that network. The same is true of human and physical capital, both of which can be put to malevolent purposes (as Al Qaeda's attacks also illustrate).

predictor (after controlling for other variables such as income, race and education) of child welfare; credit defaults; crime (especially murder); government efficiency; happiness; hospital quality; mortality rates; public corruption; school test scores; sexually transmitted diseases; support for civil liberties; tax evasion; teen pregnancy; and Social Security fraud. While causality remains an unresolved issue, this fairly crude measure of "social capital" clearly is picking up *something* that is correlated with an astonishing array of measures of individual and collective QoL. A composite measure of social connections should be explored to monitor developments over time and to benchmark performance across countries.

Figure 2.5. A composite measure of social capital in the United States



Source: Robert Putnam.

Summing up

Overall, while much remains to be done in the area of measuring social connections, significant progress has been achieved, at least in some countries and for selected areas. Social connections have been measured parsimoniously, based on replicable and standardized measures. This suggests that better and more comparable measures of social connections could be developed, building on the experience accumulated by a number of countries.

63. Some of these 14 indicators are survey-based (*e.g.* "How many times last year did you entertain at home?"), some are organizational (*e.g.* number of civic and social organizations per 1000 people), and some measure individual behaviour (*e.g.* electoral turnout). This composite measure has been computed by Robert Putnam.

Environmental conditions

Environmental conditions are at the centre of concerns about the sustainability of current development paths (they are discussed in Chapter 3 of this report), while some of the key interactions between the economy and the environment have led to efforts (mentioned in Chapter 1) to extend conventional accounting tools. Environmental conditions, however, also affect the QoL 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 the health of ecosystems). Secondly, people benefit from environmental services, such as access to clean 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 damage both the property and the lives of the affected populations. Measuring the effects of environmental conditions on people's lives is, however, complex; further, the strength of these relations is often underestimated due to limits in current scientific understanding and to the various time scales over which these effects manifest themselves.

Environmental conditions and human health

While much progress has been made in OECD countries in reducing some health risks related to environmental conditions, several surveys highlight persistent concerns of the population about the impact of various pollutants on their own health and that of their offspring. The burden of disease due to environmental factors is estimated at 24% of the total burden of disease (WHO, 2008). The prevalence of chronic conditions (*e.g.* birth defects, cancer and respiratory diseases) with links to environmental exposure is also rising. Although individual choices about smoking, diet and exercise are strong contributors to chronic disease, exposure to pollutants also matter. In most cases, cost-effective responses to these concerns are available; in addition, both research and precautionary measures are needed to better assess these long-term effects and to guard against their possible consequences for human health.

The measurement of environmentally-related diseases and disorders has been piecemeal. There have been improvements in several countries, however, as witnessed by large reductions in health risks related to lead and to environmental tobacco smoke (ETS). In other cases, the measurement of the effects of these environmentally-related diseases and exposure to environmental risks requires improvements in the collection of data, aggregation across various environmental factors, and the sharing of information (Box 2.5). Some environmental conditions affect the entire population in similar ways, while others are concentrated on specific groups, such as children, the elderly and the poor. For instance, children bear a disproportionate share of the burden of environmentally-related disease, as they ingest or breathe more pollution per unit of body weight than adults exposed to the same concentrations; children are also more exposed to some contaminants because of diet and lifestyle factors, with consequences for asthma, birth defects, cancer and cognitive development.

Box 2.5. Box Environmental effects on human health

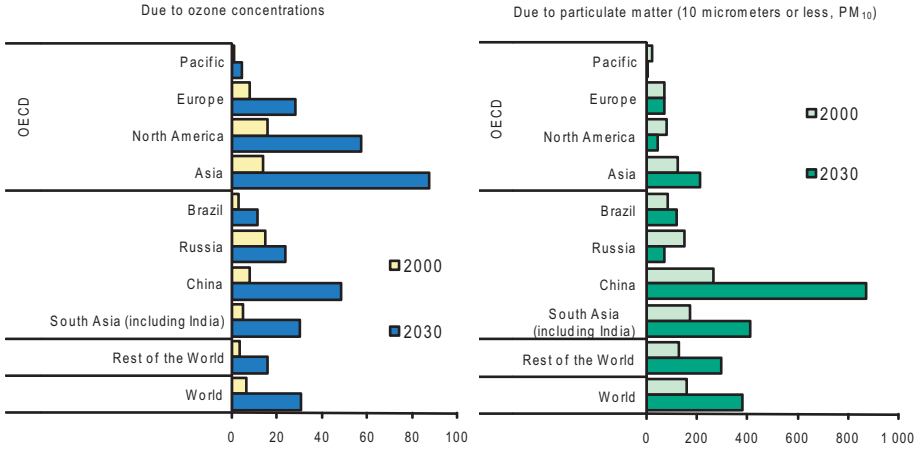
A range of environmental conditions affects human health (*e.g.* air pollution, water pollution, hazardous substances, noise). Some of the most important include the following:

- ***Outdoor air pollution:*** conventional pollutants such as particulate matter, sulphur oxides, nitrogen oxides, ground level ozone and lead have a range of health effects, such as irritation of the respiratory passages, increased susceptibility to respiratory illnesses, and early death in sensitive groups such as the elderly and asthmatics. Despite extensive regulations, many of these factors (such as ground level ozone and particulate matter) continue to cause large numbers of premature deaths (Figure 2.6).
- ***Indoor air pollution:*** radon, environmental tobacco smoke, chemicals, mould and humidity place indoor air pollution among the top five environmental risks for public health (USEPA, 2003). This reflects the large share of time that people spend indoors (around 90% of the total) and the high concentrations of these pollutants (often ten times higher than outdoor levels) in some low-quality dwellings where people work and live.[64]
- ***Industrial chemicals*** are dispersed in the environment worldwide, with potential effects on reproductive capacity, birth defects, cancer, acute illness and skin reactions. Of the 70 000 to 100 000 chemicals in use, 4 800 "high volume" chemicals represent the bulk of the total production. The assessment of the hazards associated with these high-volume chemicals has been completed for less than 700, as progress requires large amounts of data and the co-operation of chemical manufacturers (OECD, 2008a). As significant progress has been achieved worldwide for a few substances such as lead, international attention has now shifted to persistent organic pollutants (POPs) and mercur[65]y. Residues of pesticides (*e.g.* insecticides, herbicides, fungicides) are also present in several food products, in some cases in amounts high enough to harm the nervous system, especially in children.
- ***Endocrine disruptors*** are chemicals that interfere with the functioning of the endocrine system. These substances have raised much concern in recent years, as metabolic breakdown compounds from pharmaceutical products, which enter the environment with treated wastewater, may disrupt hormonal systems of fish and amphibians and, by interfering with the normal functioning of hormones and receptors, perturb both childhood development and adult reproductive health.

64. Radon is a naturally-occurring, odourless and colourless gas: it is estimated to be responsible for around 2 500 cases of lung cancers each year in France, and for 21 000 cases in the United States. ETS includes some 200 known toxic substances, most of which are classified as carcinogenic: in the United States, it is estimated to be responsible for between 150 000 to 300 000 lower respiratory tract infections annually in infants (under 18 months), and to increase the frequency and severity of asthma episodes for 200 000 to 1 million children. Asbestos, a mineral fibre, was widely used as a building insulator and fire retardant up to the 1980s and was later recognized, in some countries, as a carcinogen; in France, it is responsible for around 2 000 deaths from lung cancer per year.

65. The Stockholm Convention on POPs is being implemented, and a Convention on mercury is being prepared. Several countries (such as Finland, Sweden and the United States) have issued repeated or permanent warnings to their population not to eat certain fish species from some marine or freshwater bodies due to PCBs. In Japan, about JPY200 billion have been allocated over decades to the victims of Minamata disease (resulting from mercury), mostly paid by the polluting company.

Figure 2.6.Premature deaths from air pollution, per million inhabitants



Source: OECD (2008b), *OECD Environmental Outlook*, OECD, Paris.

Access to environmental services

Despite being recognized as a human right, access to clean water and sanitation at an affordable price remains a major concern in many parts of the world. One billion people are estimated to lack access to clean water, and 2.5 billion to sanitation services (UNDP, 2006). An inadequate sanitation infrastructure results in diseases caused by viruses, bacteria (such as cholera and E. Coli) and parasites (*e.g.* cryptosporidiosis). To meet this challenge, the Millennium Development Goals include the goal of halving the proportion of people without access to water supply and sanitation.

Access to water and sanitation differ widely among countries, with OECD countries benefiting from much better water supplies and sanitation services than developing countries. For rural populations, access to water is critical for agricultural production: this sector represents 70% of water abstraction in the world, and in many countries rural populations rank water supply as their top concern. Water supply is also a constraint on economic development in countries such as Australia, Turkey, Mexico, and Spain as well as in California (OECD, 2006b). Governments, firms and NGOs are actively involved in solving problems in the financing and governance of the water supply. While in most OECD countries, improvements to the water supply infrastructure have led to the quasi-elimination of health disorders associated with pathogenic agents (such as cholera), concern now relates to human exposure to lead in drinking water, mainly resulting from lead pipes and solder used on copper pipes. Regulations in Europe and the United States restricting lead levels in drinking water have led to the gradual replacement of lead service lines.

Assuring access to nature and outdoor recreational areas has also gained in importance with the growing awareness of the health and economic costs of a sedentary lifestyle, particularly for urban populations. Parks in cities and natural areas in peri-urban zones provide opportunities for physical exercise, recreational activities and relaxation. Forested land and areas featuring agreeable landscapes and uncontaminated water bodies benefit QoL in various ways. Outdoor activities can contribute to good health and well-being, and help to combat obesity, which is a public health issue in most OECD countries.

Access to environmental information has been established as a right over the last ten years in a number of countries (building on the Aarhus Convention and a number of other international agreements). This right of access to environmental information was also included in the preamble to the French Constitution. In most OECD countries, much progress has been made in terms of providing access to such information as an essential dimension of "environmental democracy". Progress has been made in implementing this right, as citizens are now more able to obtain, at their request, much information, at low or limited cost, on matters of concern such as environmental risks and the facilities, institutions and services available in their country, region or neighbourhood. The full implementation of laws on integrated prevention and control, as well as the existence of facilitators and mediators, has also enhanced the public's right to know about environmental conditions and hazards. Extending the right to access environmental information to more countries in the world is part of the extension of people's right to access general administrative information.

Environmental amenities and disamenities

In most developed countries, noise is ranked high by people in surveys about their own living conditions. Peace and quiet are valued particularly by urban populations. Noise has extended over time and space, and the total acoustic energy produced has increased over time. To enhance the QoL of residents and to improve the quality of tourism, authorities in some countries have designated quiet areas in nature, rural, mountainous or other specific areas (such as lakes), while others have implemented noise curfews during certain time periods (*e.g.* at night).

Exposure to noise is of concern for QoL because of its effects on physical and psychological health, behaviours and social activities. The effects on human health include annoyance, hearing impairment (for extreme exposure) and risk of cardiovascular disease from chronic exposure. Noise affects sleep, cognitive performance (especially for children) and speech. The cost of noise damage has been estimated at between 0.2% and 2% of GDP annually (EU, 2001). People are mainly exposed to noise from street and road traffic, railways and air traffic, as well as factories and construction work. Children and youth are more sensitive to high noise levels, which can lead to slower learning, long-term hearing damage or tinnitus.

Low environmental quality also raises safety concerns in risk-prone areas. "External safety" in a strict sense refers to the hazards facing people living near industrial installations and hazardous substances transport routes, for example, near large chemical plants (*e.g.* Toulouse accident), or around airports (*e.g.* Amsterdam accident) and railway yards. External safety is distinguished from (but not unrelated to) safety internal to the facilities or vehicles. While there may be important gaps between actual risks and the risks perceived by populations, both have a claim to matter for QoL. Also, large differences in perception may occur between those working within (or familiar with) a facility and those less familiar with it. While disaster prevention and control concerns everyone, progress in access to environmental information and corporate environmental reporting is needed to allay these fears.

People value environmental amenities such as green landscape, shorelines, mountains, light and tranquillity. People also express concern about the disamenities associated with noise or living near environmental infrastructures such as wastewater treatment plants, waste

landfills, and incinerators. Such concerns often lead local populations to resist the establishment of such facilities, even those of the highest standards, near the areas where they live. Some of these concerns relate to the effect of pollution on health, some to the effect on land and housing values; both can be captured through data on market housing transactions and estimates of hedonic prices.[66]

Climatic variations and "natural" disasters

Most countries face seasonal or local problems with water volumes, such as seasonal droughts, shrinking groundwater reserves or falling groundwater tables. Furthermore, there is a risk that climate change will affect rainfall distribution and evaporation, resulting in the increased frequency and intensity of extreme weather episodes, including floods, droughts, high winds and rising sea levels. Potential damage would be mainly to property in developed countries, and to human life in developing countries.

Many countries are threatened by major floods, which have severe economic and social impacts. In recent years, major flooding has occurred in the Rhine basin, Central Europe, France and other countries. Some of these floods have led to economic losses equivalent to a few percent of GDP, with most of the damage uninsured. In many countries, the frequency and severity of river flooding has increased in recent years. In 2005, storm and storm-related flood damage (*e.g.* from Hurricane Katrina, Rita, Wilma) caused insured property losses estimated at over USD 80 billion (OECD, 2006c). Since 1970, out of the 40 most costly events in terms of insured losses 10 have involved floods, four of them in 2005. In most developed countries, people affected by flood damage may receive some compensation from government and insurers. If flood insurance premiums were to be raised to fully match insurance claims, people living in flood-prone areas would face a drastic rise in their housing costs or perhaps simply be unable to get insurance at all.

Recent assessments indicate that current patterns of water use are unsustainable in many countries. An increasing incidence of local and regional droughts would lead to more water crises. The effects would be felt most by regions where water stress is already high. Several developed countries, or regions thereof, are in this category. The QoL of farmers is first affected. In Australia, many of them are living on drought compensation payments which had to be maintained over the years, and they may have to leave their land (OECD, 2008). In some parts of the world, large forest fires have been associated with drought, leading to threats, damages and casualties for peri-urban populations (Australia, Greece, Portugal and California). Rising seawater is a considerable threat to populations, and is potentially exacerbated by climatic variations (*e.g.* high winds or tornados resulting from such variations can cause seawater rises). The potential damage in terms of lost lives and property is considerable, as significant population and infrastructure are concentrated in coastal areas.

66. Estimates of hedonic pricing builds on individuals' willingness to pay for residential property (or rather its characteristics) and separates the economic valuation of other characteristics of this property from the valuation of environmental amenities such as a mountain view or access to a park (or, more broadly, environmental features, including pollution and noise). Hedonic pricing has the advantage of being based on actual choices (as opposed to opinion surveys of preferences); it relies on information on property markets, and can be applied to a range of environmental amenities and disamenities. Its limits include its high demands on data and statistical competencies, as well as assumptions that the property market is not disturbed by changes in taxation, lending or mortgage conditions external to the underlying model.

Indicators: what is available and what is needed

Much progress has been achieved in the last two decades in terms of measuring environmental conditions (building on better environmental data, accounting tools, opinion surveys, and the regular monitoring of indicators[67]), understanding their impacts (*e.g.* evaluation of related morbidity and mortality, labour productivity, the economic stakes associated with climate change, biodiversity change, damage by disasters) and establishing a right of access to environmental information. However, because of the limits both of economic accounting and of composite indices of environmental quality, the most pragmatic and reliable approach to measuring environmental conditions continues to rely on physical indicators. Several environmental indicators are already available. Some of them refer to pressures on the environment from economic activities (such as emissions), others to responses from administration, enterprises or households to environmental degradation (such as environmental expenditure), and still others to the actual state of environmental quality (such as ambient air quality). The principal function of these indicators has been to support the design and implementation of environmental policies, and to track environmental changes, as part of environmental strategies, plans, programmes, budgets. At the international level, the OECD set of "key environmental indicators" has been adopted by Environment Ministers.

When reviewed from the perspective of assessing the contribution of environmental conditions to QoL, however, existing indicators remain limited in important respects: for example, emissions indicators mainly refer to the aggregate quantities of various pollutants, rather than to the share of people exposed to them. From a QoL perspective, existing indicators should be supplemented by regular monitoring of: i) the number of premature deaths from exposure to air pollution, in particular from particulate pollution; ii) the share of the population lacking access to water services, in particular water supply and sanitation; iii) the share of the population without access to nature, with a focus on daily proximity and appropriate mapping; iv) the share of the population exposed to daytime noise above 65dBA levels[68], in particular noise in dwellings, to be collected through appropriate surveys and mapping; v) information on the damage (both insured and non-insured) incurred due to environmental disasters, such as floods and droughts; vi) measures and assessment tools for emerging environmental issues (*e.g.* endocrine disruptors, pesticides, non-ionizing electromagnetic radiation) and their longer-term effects on QoL (*e.g.* from hazardous substances, climate change, biodiversity degradation, resource depletion); vii) methods for valuing people's environmental choices (*e.g.* hedonic prices, valuation of externalities and of the services provided by ecosystems) and for supporting economic decisions related to the environment and QoL (costs of inaction, environment-related jobs; energy and material intensities); and finally, viii) surveys of people's own feelings and evaluations of the environmental conditions of their country and neighbourhood.[69] As many of the effects of environmental conditions on QoL differ across various groups of people, these indicators should (when possible) refer to people grouped according to various classification criteria.

67. Suitable survey questions are being implemented both in Europe (*e.g.* Eurobarometer) and worldwide (*e.g.* the Gallup World Poll).

68. An average noise level of 65dBA (over a day time period, *e.g.* 6.00am-8.00pm in dwellings) is considered to be unacceptable (*e.g.* by EU legislation), while 55 dBA is considered to be a comfort level.

A final issue that deserves to be raised is aggregation. In recent years, the call for more integrative environmental indices has become louder, and some non-governmental organizations and official institutions have promoted various aggregated indices (see Chapter 3). However, reservations continue to be voiced about the limits of these indices, their perceived opacity, their methodological shortcomings, and the potential for misinterpretation and misuse. Although the drive to develop more integrated measures of environmental conditions is not a substitute for the type of physical indicators mentioned above, the continuing pressure for more parsimonious aggregate indices points to the importance of setting some kind of industry standard or quality label, and for greater guidance on the interpretation of these indices.

Insecurity

Insecurity is a source of fears and anxieties that negatively affects the quality of life of the people concerned. Insecurity also implies uncertainty about the future, which decreases QoL for risk-averse individuals. Because of these considerations, a long tradition of research has aimed at measuring "insecurity" (or safety and vulnerability) and to assess its effect on well-being. For the purpose of devising suitable approaches to its measurement, it may be helpful to distinguish between personal and economic insecurity.[70]

Personal insecurity

Personal insecurity includes those external factors that put at risk people's physical integrity. The most obvious of these factors are crimes and accidents. While these external factors, in their most severe manifestations, can lead to the death of the person involved, in less extreme forms they affect a much larger number of people in each country.[71]

Death from external causes accounts for between 4% and 11% of all deaths in any given year in most OECD countries. Among the various types of external causes (based on the WHO classification), deaths due to land transport account for 23% of the total, followed by those related to accidental falls (14%) and, further behind, assaults (3%).[72] Available statistics

69. Levison (2009) goes in this direction by combining survey data with air quality data to model subjective well-being as a function of the current air quality in the United States. Controlling for county, year, and month fixed effects, he finds no happiness responses to local annual pollution averages (suggesting habituation or endogeneity), but large happiness responses to local daily fluctuations (controlling for weather and other covariates). See also Luechinger (2009) and Welsch (2006).

70. Environmental insecurity is not developed here since this issue is already considered above.

71. From this same perspective, the Calvert-Henderson Public Safety Indicators measure those outcomes that result in death or injury. These indicators identify: i) several important features of individual action within the private sphere that affect the probability of injury; and ii) specific environmental and public life-saving actions (beyond individual control) that impact on people's safety and potential for harm. The indicators included relate to various causes of death, with special focus on injuries, motor vehicle crashes and accidents due to firearms (http://www.calvert-henderson.com/pubsaf.htm).

72. Suicide, which can be considered an extreme manifestation of mental illness, accounts for the largest component of all deaths from external causes (27% on average).

on the causes of death fail to classify a large fraction of all deaths from external causes, and these "unidentified" causes account for around one-third of all such deaths in OECD countries.[73] While all of these deaths are captured by mortality statistics, the rationale for having specific measures of their frequency is that they may have a different emotional effect on people compared to deaths related to medical conditions.

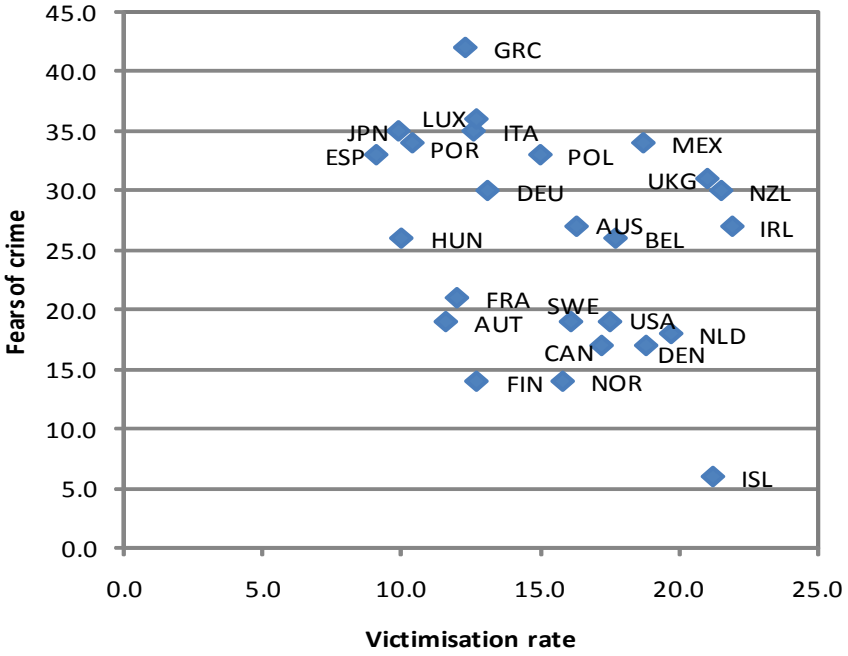
Beyond their death toll, threats to personal safety affect QoL even in their less extreme manifestations. The most obvious example is crime. Measures of crime can be derived from a variety of sources (*e.g.* administrative records and household surveys). While comparisons of crime statistics based on police records are affected by cross-country differences in reporting practices, greater comparability is achieved through household surveys specifically designed to assess people's experience with victimization. One such survey is the five-yearly *International Crime Victim Survey* (ICVS), which is run by a consortium coordinated by the United Nations Interregional Criminal Justice Research Institute (UNICRI) and the United Nation Office on Drugs and Crime (UNODC).[74] Based on this source, around 15% of all OECD citizens reported that they or other members of their family experienced one of 10 types of conventional crime in 2005, with large differences across countries (ranging between over 20% in Ireland, New Zealand, Iceland and the United Kingdom to less than 10% in Hungary, Japan and Spain). Among all these types of conventional crime, around one-third are contact crimes, while an additional 10% of respondents reported having experienced non-conventional crimes such as consumer fraud or corruption. In most OECD countries, the frequency of conventional crime has declined since 2000.

Personal insecurity with respect to crime can also be measured through data on people's *fears* of being a victim of physical aggression. The most remarkable feature of these reports on subjective fears is how little they seem to be related to measures of experienced victimization. Questions from the same survey about whether people feel unsafe when walking in the street after dark not only show that the share of people fearing crime is significantly higher than the overall victimization rate, but also that countries where a higher share of people fear crime do not record a higher frequency of victimization (Figure 2.6). Also, within countries, older and richer people feel more unsafe than younger and poorer people, despite being less likely to be victims of crime. Changes in experienced victimization and in fears of crime are also weakly related within individual countries, which suggest that the media plays a large role in amplifying concerns and focusing public opinion on specific domains.

73. Deaths from work accidents are not separately identified in OECD statistics on "causes of deaths". In some OECD countries, accidents due to commuting to and from work (which are classified as "fatal work accidents" in ILO statistics on the subject) account for up to 50% of all deaths from land transport.

74. National victimization surveys are also conducted in individual countries. For example, in the United States, the *National Crime Victimization Survey* has collected data on personal and household victimization since 1973, while in France, a similar survey started in 1996. The data included in victimization surveys generally include the type of crime; time and location of occurrence; relationship between victim and offender; characteristics of the offender and of the victim (such as age, race, gender and income); protective actions taken by the victim during the incident and results of those actions; consequences of the victimization; type of property lost; whether the crime was reported to the police and reasons for reporting or not reporting; and the offender's use of weapons. A manual aimed at increasing comparability between national victimization surveys is currently being finalized under the auspices of the Conference of European Statisticians.

Figure 2.6. Prevalence of victimization and fear of crime in OECD countries, 2005
Percentage share of respondents



Note: Victimization is measured as the share of respondents reporting that they (or some other member of their families) had been a victim of one of 10 types of conventional crime in the previous 12 months. Fear of crimes is based on questions about people feeling unsafe or very unsafe on the street after dark.

Source: 2005 International Crime Victim Survey and European Survey on Crime and Safety.

These patterns highlight the importance of more regular and objective measures of crime to orient public discussion. Victimization surveys are an essential tool to assess the frequency of crime and the fear it generates, but their practical use is often limited by small sample sizes, the low survey frequency, and the lack of a standardized classification of types of crime.

Victimization surveys also have other limits:

- They underestimate domestic violence, especially when the person is interviewed at home in the presence of other household members. Most of this violence is exerted against women, and contributes to relegating them to the home and denying them opportunities for work and leisure: this violence has been identified by the *World Health Organization* as a major health risk affecting women and children (Box 2.6).
- Answers to surveys may be affected by cultural traits that vary across countries and groups (as in the case of sexual abuse).
- Finally, and most importantly, victimization surveys are difficult to design and impossible to conduct in countries ravaged by conflicts and wars.

Other sources need to be mobilized to measure these threats to personal security.

Box 2.6. Marital violence: implications for women's and children's health and security

Violence against women, within and outside the home, in peacetime and during conflict situations, can take many forms. Each can undermine women's basic capabilities and functionings. But marital violence is the most pernicious, not least because it occurs within the family – the institution that is conventionally assumed to be driven by altruism and instrumental in enhancing human well-being. Marital violence will scar the women who suffer it, the children who witness it, and the men who perpetrate it. It has negative consequences for the well-being of individuals, their families, and the wider society (Agarwal and Panda, 2007).

The World Health Organization has identified marital violence as a major health concern (WHO, 2000, 2002). Although still largely under-reported, estimates indicate that world-wide between 10% and 50% of women ever married report having experienced physical violence from spouses (Population Reports, 1999), with this phenomenon cutting across countries and economic groups. Psychological abuse is even more common. Marital violence is found to cause serious physical and mental injury to women (Dannenberg *et al.*, 1995; Harper and Parsons, 1997; Maman *et al.*, 2000). Violence during pregnancy is associated with miscarriages, low birth weight infants, maternal morbidity, and even foetal and maternal death. Injuries caused by domestic violence can critically undermine a woman's economic freedom – such as her capability to earn a living – by making her fearful of reprisal if she goes out to work, or seeks to upgrade her skills, or explores job options. Physical or mental injury can adversely affect her job market prospects, productivity, regularity of work life, and chances of upward mobility (Brown *et al.*, 1999; Lloyd *et al.*, 1999). Marital violence can also erode a woman's social opportunities by undermining her relationships and social capital. Neighbours and friends tend to shun families where violence is common, and a woman's self-confidence can get so eroded that she withdraws from social contact. The "battered woman syndrome" implies that a woman's sense of self is so damaged that she begins to believe she deserves the abuse. It also makes her fearful to seek help when she needs it. Marital violence can similarly undermine a woman's political freedom – her ability to be an active citizen or seek her entitlements.

Domestic violence also carries intergenerational costs, such as foetal damage when the mother faces violence during pregnancy and psychological damage to children witnessing such violence. Children witnessing domestic violence tend to suffer from higher emotional and behavioural problems than other children and carry the seeds into adulthood (McCloseky *et al.*, 1995; Edleson, 1999). Girls who see their mothers being beaten by their fathers are more likely to accept spousal abuse. Boys who see their fathers beat their mothers are more likely to beat their wives. In other words, marital violence undermines the capabilities – physical and mental – of all family members, leading to high human, social and economic costs. The Australian Committee on Violence estimated that the cost of refuge accommodation for victims of domestic violence in 1986-87 was US$ 27.6 million (Carrillo 1992). Comprehensive surveys are needed to assess the extent and effects of marital violence and its correlates, and to monitor the impact of laws and policies instituted for dealing with it. Women's ownership of assets, such as housing and land, for instance, has been found in some regions to be linked with a significantly lower incidence of martial violence (Agarwal and Panda, 2007), and with higher women's security (Agarwal, 1994). Freedom from domestic violence needs to be a significant part of any exercise for evaluating the quality of life and for expanding human capabilities.

Potential victims can buy protection against the risks of crime, accident, or natural disaster from insurance companies. In this way, the negative consequences generated by the realization of the risk are partially alleviated. For example, the victims of a non-fatal car accident receive a financial compensation for their economic damages (medical bills, lost wages) and for their non-economic prejudice (injuries, temporary or permanent disability, pain, and suffering). When the accident is fatal, the dependents of the victims are entitled to compensation for their loss. In both cases, the victims or their dependents can accept the settlement proposed by the insurance company or can go to trial.

International comparisons of the compensation for a life or a non-fatal injury are difficult, as compensation scales are not official. However, some statistics are public. For example, in

France, where insurance companies have to publish the compensation obtained by the victims or the dependents, the median compensation for the moral prejudice paid to a husband or wife in the event of the death of a spouse is €15,500 (€14,000 for a child and €15,000 for a father or mother). These measures are likely to largely under-estimate the consequences following the realization of these various types of risks. Oswald and Powdthavee (2008) report evidence of a large impact of bereavement on people's subjective well-being.

Economic insecurity

The term "insecurity" encompasses a broader range of risks beyond those that bear on personal insecurity. For example, variations in income over time can be a source of insecurity for the person affected, even when it is healthy for society as a whole (*e.g.* because it reflects higher social mobility). However, assessing this variability (and how it changes over time) requires surveys that follow the same person over long periods. In the United States, where such surveys exist, the evidence points to a clear increase in income volatility over time, especially for households at the bottom of the income scal[75]e. As similar surveys are rare in other countries, this section is limited to a few types of "named" risks that can lead to economic insecurity: unemployment, illness, and old age.

Economic insecurity may be defined as uncertainty about the material conditions that may prevail in the future. This insecurity may generate stress and anxiety in the people concerned, and make it harder for families to invest in education and housing. The United Nations' *Universal Declaration of Human Rights* refers to the "right to security in the event of unemployment, sickness, disability, widowhood, old age or other losses of livelihood in circumstances that are beyond the control of each individual". This "social right" is generally enforced through the protections attached to jobs and granted through social policies.

The measurement of economic insecurity may follow different paths in terms of what is actually measured. Some approaches try to quantify the frequency of specific risks, while others look at the consequences for quality of life when a specific risk materializes, and at the means available to people to protect themselves from these risks. There have been few attempts to combine more comprehensive information about both the frequency and consequences of various risks.

The realization of each risk has negative consequences for the quality of life of the person affected, depending on the severity of the shock, its duration, the stigma associated with it (*e.g.* being unemployed) and each individual's risk aversion. Most national statistical systems and international organizations provide some measure of the financial consequences of unemployment and old age (replacement rates) or illness (out-of-pocket expenses). These consequences, however, also depend on the type of protection available and on its cost, dimensions that are rarely considered.

Economic insecurity due to unemployment

Job loss can lead to economic insecurity if unemployment is recurrent or persistent, if the replacement rate is low, or if workers have to accept major cuts in pay, hours or both to find a

75. The *Panel Survey on Income Dynamics*, managed by the University of Michigan, has been tracking a nationally representative group of households since the late 1960s.

new job. Therefore, it is not the loss of a job *per se* that causes economic insecurity, but rather the frequency and duration of the spells of unemployment and inactivity, and their consequences.

A useful distinction is that between *job instability* and *job insecurity.* The first refers to the break in the contractual relationship between the worker and the employer, where this interruption can be followed by a new hire with a different firm. The second refers to the case of a person remaining jobless for an extended period, with the definition of this length being partly conventional and partly based on available information. A possible indicator of job insecurity is the number of workers who are employed at a given date and are inactive (or unemployed) one year later (as measured through a *Labour Force Survey*).[76] Based on this definition, in 2003 job insecurity in France (as measured by the annual transition rate between employment and unemployment or inactivity) was at the same level observed twenty years earlier, despite a large increase in job instability (as measured by job turn-over, Cerc, 2005). Different patterns of job insecurity and instability are also evident across European countries (Table 2.2), with some countries combining high job stability and high job security (Belgium, Italy, and Portugal), while others feature low job stability and low job security (Germany, United-Kingdom, Spain, and Ireland), and still others exhibit either high job security and low job stability (the Netherlands, Denmark and Finland) or low job security and high job stability (France, Austria, and Greece).

Table 2.2. Transition rate from employment to non-employment in European countries

	Non-employment	Same job	Another job
Luxembourg	2,7	91,0	6,3
Belgium	2,3	86,2	11,5
Italy	4,6	85,0	10,4
Portugal	4,7	84,5	10,8
The Netherlands	3,7	79,3	17,0
Denmark	4,0	79,3	16,7
Finland	4,4	83,5	12,2
France	5,5	84,8	9,6
Austria	5,2	84,8	10,0
Greece	5,7	84,6	9,7
Germany	7,0	80,7	12,3
United-Kingdom	5,9	76,2	18,0
Ireland	7,6	73,8	18,6
Spain	8,9	70,2	20,9

Note: 5.5% of French workers employed in October 2000 were non-employed in October 2001; 84.8% were employed in the same job. Data refer to workers employed in the private sector in October 2000.

Source: Cerc (2005), based on data from the European Community Household Panel for 2000 and 2001.

The financial consequences of job instability encompass both present and future losses of income. The present loss is due to the fact that the replacement income is typically lower than the earnings on the previous job. The extent to which people are protected from the financial

76. This indicator, however, will miss those workers that have experienced recurrent spells of joblessness during that one-year period.

impact of unemployment in the short term depends on four factors: i) the probability of being entitled to unemployment insurance; ii) the level of UI benefits relative to previous earnings; iii) the probability of being eligible for social assistance when not entitled to UI; and iv) the level of social assistance. While indicators of these financial consequences are available, cross-country comparisons are difficult for several reasons (*e.g.* unemployment benefits can depend on family composition; unemployment and assistance benefits may not be easily separable; the share of people entitled to unemployment benefits may vary with the duration of unemployment), requiring special investments in this direction. In the longer term, the financial consequences of unemployment will include possible losses in wages when the individual eventually finds another job, as in the case of workers whose human capital was specific to a particular firm or industry. The risk aversion of each worker, the discount rate used, and insurance costs also influence the size of these longer-term financial losses.

Job insecurity can also be measured through self-reports of individual workers. Fears of job loss can have negative consequences for the quality of life of each worker (*e.g.* physical and mental illness, tension 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 (De Witte and Näswall, 2003). Various surveys provide information about *perceived* job insecurity, by asking workers either to evaluate their satisfaction with the security of their present job, or to rate their expectation of job loss with either probabilistic questions ("what do you think is the percent chance that you will lose your job during the next 12 months?") or qualitative questions ("thinking about the next twelve months, how likely do you think it is that you will lose your job or be laid-off?"). Data on satisfaction with job security among European countries in the early 2000s show large cross-national differences, with greater fear of job losses in southern European countries compared to northern ones (Cerc, 2005)[77].

Economic insecurity due to illness

Illness can cause economic insecurity directly, through its medical cost, 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 to go into debt, to sell their home and assets, or to forego treatment at the cost of worse health outcomes in the future. The higher health care costs recorded over the past few decades in some countries have made these risks all the more frequent.

One indicator of illness-related economic insecurity is provided by the share of individuals without health insurance. Almost all Europeans are covered by a basic health insurance plan, although with differences in the extent of coverage provided for various types of treatment (OECD, 2004). In contrast, nearly 47 million Americans (16% the population) were without health insurance in 2006 (DeNavas-Walt *et al.*, 2007), with an increase of almost 9 million people since 2000.[78] Nearly 90 million people – about one-third of the population below the age of 65 – spent a portion of either 2006 or 2007 without health coverage.

77. The data set used is the European Community Household Panel (2001), which asks people to rate their satisfaction on a 4-point scale (unsatisfied, fairly satisfied, satisfied, fully satisfied). Only 10% of Austrians are "unsatisfied" versus 48% of Greeks. At the opposite end, the response "fully satisfied" is very frequent in Austria and Ireland (37%), while it is rarely chosen in Portugal (4%), France (6%) and Greece (6%).

78. In the United States, the increase in the number of uninsured people in 2006 was concentrated among people of working age, with 1.3 million full-time workers losing their health insurance in that year.

However, this indicator underestimates economic insecurity due to illness, as even insured people generally pay something for their treatment. For example, in 2004, more than 14 million Americans paid more than 25% of their earnings on out-of-pocket medical costs and health premiums; 10 million of them were insured (Families USA, 2004). Medical costs are a factor in nearly half of all personal bankruptcies in the United States, and 80% of families who go bankrupt for medical reasons have health insurance (Himmelstein *et al.*, 2005). The financial consequences of illness can be measured by looking at out-of-pocket health expenses. This information is provided by national and international sources, but with low frequency and long delays. To these out-of-pocket health expenses should be added the loss of income that occurs if the person has to stop working and the health insurance does not provide replacement income.

Economic insecurity associated with old-age

Old-age, *per se*, cannot be considered as a risk, but it still implies a certain degree of economic insecurity due to uncertainty about the needs and resources that will be available after having withdrawn from the labour market. Two types of risk, in particular, can be identified.

- The first is the risk of falling into poverty in old-age. Most OECD countries have achieved a significant shift in the risk of poverty away from the elderly and towards younger people over the past few decades, and this shift is one of the most significant achievements of social protection in these countries. But the risk of poverty remains important in some countries (*i.e.* those with less mature pension systems) for some groups of elderly people (*e.g.* women with limited labour-force participation during their working years), and when combined with other contingencies (such as chronic health problems and disability in old age).

- The second is the risk of volatility in future pension payments. While all retirement-income systems are exposed to *some* risk (*e.g.* risk of lower pension benefits in pay-as-you-go systems due to changes in demographic conditions), the greater role of the private sector in financing old-age pensions (in the form of both employer-based occupational pensions and of personal pension savings) have allowed extending pension coverage to a larger number of workers, but at the cost of shifting risk from government and firms towards individuals, who have less capacity to hedge against these risks and whose degree of risk aversion increases with age.

The need to better measure economic insecurity in old-age will be especially critical as the current financial crisis unfolds, as many people will be approaching retirement with sharply lower annuities from accumulated financial assets, lower (or no) housing values, and a risk of losing their pensions following the default of their employer or their pension funds.

Combining information on both the frequency and consequences of each type of risk

A comprehensive measure of insecurity would ideally need to account for both the frequency of each risk and its financial consequences. While developing such comprehensive measures is a task fraught with difficulties, some attempts in this direction have been made. For example, Osberg and Sharpe (2002) measure the economic insecurity associated with unemployment as a function of the probability of becoming unemployed and the extent to which people are protected from the economic losses that follow unemployment (as estimated

by gross replacement rates for the unemployed).[79] Fleurbaey and Gaulier (2007) propose a more elaborated measure that also integrates risk aversion.[80]

The idea of combining information about both the frequency of a risk and its financial consequences (*i.e.* by taking the product of the probability that the risk materializes and the change in well-being that would result from its occurrence) can be generalized beyond unemployment to any risk. To be complete, the indicator of the change in well-being should integrate a measure of the cost of protecting oneself against the risk, which is a complex task. Research on these approaches should be further developed, so as to lead to indicators that could inform about both the scale of the various risks in each country and their distribution across the population.

Factors shaping the consequences of various risks

People may buy protection against the financial consequences of various economic risks through the accumulation of private assets, or through public or private insurance. Because of the importance of social protection in providing insurance against various types of economic risk, a tradition of research in this field has developed measures of economic insecurity based on the size and features of social protection systems in various countries (Menahem, 2007).

While comparative research on welfare state outcomes has highlighted a number of patterns that are important for assessing QoL for people in each country, these patterns are sometimes based on measures of the protection available (typically based on the size of public spending for various social programmes), which are limited in a number of respects.

- First, measures of *welfare efforts* based on gross public spending ignore the support provided through the tax system, for example by awarding tax advantages that encourage people to buy health insurance or private pensions.[81]
- Second, public programmes can substitute for a wide range of private arrangements that also provide protection against various economic risks. Some of these private arrangements also have an element of redistribution towards people in specific conditions (such as mandatory sickness payments paid by employers to their employees),[82] while others are entirely private (such as investment in housing during working age used to generate income while in retirement).

79. Osberg and Sharpe (2002) argue that this second factor should ideally be measured as the proportion of the unemployed claiming unemployment benefits times the percentage of weekly earnings replaced by benefits. Quality information about the proportion of the unemployed receiving benefits is, however, lacking. Osberg and Sharpe also define "insecurity in old age" as the share of elderly people who are poor times the average poverty gap for this group.

80. Fleurbaey and Gaullier (2007) model the income loss due to the risk of unemployment as a function of the degree of risk aversion, the probability of unemployment (computed from the rate of unemployment and its average duration), and the income loss associated with unemployment.

81. In the United States, for example, consideration of these various forms of tax support increases public social spending by around 1 point of GDP (to a level of net public social spending of 18.6% of GDP in 2003), while in France the same flows lower it by around 4 points of GDP (to a level of 29.2% of GDP).

82. In the United States, these mandatory private programmes with an element of redistribution (hence included in the OECD definition of "total social spending") account for 8.4% of GDP, as compared to less than 3% of GDP in France.

- Last, all types of protection against economic insecurity have costs. In the case of public programmes, the taxes needed to finance them will affect the strategies that individuals may follow to hedge against insecurity. These costs should also be taken into account when measuring the financial consequences of various types of risk.

Aggregation across different categories of economic risk

A general problem confronted by all attempts to derive a comprehensive measure for economic insecurity is that of aggregation. While each of these risks can be quantified according to specific indicators, they do not lend themselves to aggregation as they lack a common metric to assess their severity. Some steps in this direction have, however, been taken.

- Osberg and Sharpe (2002) scale measures for four main types of economic risk (unemployment, illness, single parenthood, old-age) through a "linear scaling technique", applied in such a way that, by convention, the increase in each scaled variable represents lower economic insecurity; the four scaled variables are then aggregated into an overall "economic security index" based on weights that represent the relative importance of four demographic groups in the total population.[83]

- Fleurbaey and Gaullier (2007) compare populations that differ in some non-income dimension by computing the "equivalent income variation" that would make each population indifferent between its current situation and a reference situation with respect to the non-income dimension. This equivalent income approach allows aggregating various risks.

- Other approaches, such as the *Personal Security Index* developed by the Canadian Council on Social Development, aggregate objective measures of various risks pertaining to both economic security (*e.g.* risks of losing one's job and financial resources, of disease and injury) and personal security (*e.g.* risk of crime and theft) using subjective weights based on specific surveys that ask people to rate three types of security (economic, health, personal). Such a methodology could be applied to other countries.

Cross-cutting issues

Most of the challenges involved in providing credible measures of QoL are specific to each of the dimensions considered. Some challenges are, however, cross-cutting, and would be missed by further research and data gathering undertaken within each of the dimensions considered above.[84] Three of the most salient challenges are addressed below.

83. These weights are: i) for unemployment, the share of the population aged 15-64 in the total; ii) for illness, the share of the population at risk of illness (100%); iii) for single parenthood, the share of the population comprised by married women with a child under 18; and iv) for old age, the share of the population aged 45-64 in the total. These shares are normalised for all years to sum to unity.

84. While insecurity is dealt with as an objective factor shaping quality of life, it can also be considered as a cross-cutting issue because of the large variety of risks the individuals are exposed to. The placement of insecurity among the objective factors has been debated at some length and is conventional.

Inequalities

Measures of QoL typically refer to *average* conditions in each country for each of the specific domains considered. What these average measures miss are the *inequalities* in people's experiences. Accounting for these inequalities is necessary to fill the gap between country-wide estimates and people's feelings about their own conditions. Developing such indicators would also allow paying special attention to the conditions of people with the lowest QoL in terms of each of its various domains.

While established methodologies and data sources allow measuring inequalities in the distribution of economic resources within countries 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 because inequalities in these non-monetary dimensions cannot always be described through information about how these features are distributed around their mean. These difficulties require developing measures of inequalities that are specific to each field.[85] It is also necessary to assess these inequalities in a comprehensive way, by looking at differences across individuals, groups and generations.

- *Inequalities among people*. Individuals are the unit of analysis for most analyses of QoL as – even when people derive their well-being from a larger unit, such as the family or community – there may be asymmetries in how resources and opportunities are distributed within these units: the distribution of family chores between men and women, and differences in expenditure patterns (*e.g.* in terms of how much is devoted to children) depending on who within the family controls the financial flows, are examples of this general pattern. The most direct way of highlighting these differences in QoL across people is to compare the outcomes for those at the bottom of the QoL scale and those at the top. In the case of education, for example, in 2006 the test scores (in science) of students aged 15 in France were around 146 points higher at the top quarter of the achievement scale than at the bottom quarter, a difference equivalent to almost 4 years of schooling. These differences in students' performance can have lasting implications for youth as they move into adult life: poor learning outcomes at the end of compulsory schooling lead to a higher probability of dropping out of school before completing secondary education, worse earnings and career prospects when entering the labour market, a lower probability of benefiting from on-the job training and a higher probability of poverty in adult life.

- *Inequalities between groups with different individual characteristics.* A second way in which these disparities in QoL outcomes can manifest themselves is among homogeneous groups of the population, *e.g.* by age, gender, socio-economic status or other criteria. These between-group differences are important, as groups help to structure the identity of individuals and to define the ways in which people express their voice and act collectively. Groups can also be a source of inequality when they exclude people from the benefits and guarantees that are available to insiders. While these

85. This difficulty is best illustrated by the case of health, where attempts to create indices of inequalities that parallel those used to describe income inequalities have run into a series of methodological difficulties because of the way various health outcomes are distributed among the population. For example, distributions of mortality have most of their mass in the tails, which makes standard measures of inequality (such as Gini) behave in non-standard ways. Other aspects of QoL, such as voting, can be measured only through dichotomous variables (either participating in a ballot or not), which make inequality measures conceived for continuous variables unsuitable.

between-group differences manifest themselves in all dimensions of QoL, they are especially important in the case of health. For example, inequalities in average life expectancy between ethnic groups range from 6 ½ years between Afro-American and white men in the United States to 18 years between aboriginals and others in Australia.

- *Persistence of advantage and disadvantage across generations.* In general, most respondents in various OECD countries declare that inequalities are acceptable when they do not trap people in situations of disadvantage that persist from generation to generation; and yet there is much evidence that a range of QoL outcomes are transmitted from one generation to the next. In particular, the parents' socio-economic conditions are among the most important determinants of their children's opportunities. For example, babies born to a mother who is malnourished and in poor-health have a higher probability of being underweight and of experiencing poor health in adulthood. Similarly, the socio-economic status of parents is among the most important determinants of children's educational results.

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 types of inequality, such as between ethnic groups, are more recent (at least in countries that have experienced large waves of immigration) and are set to become politically more salient in the future as immigration continues.

This discussion points to two main conclusions. The first is that there are as many inequalities as there are features of QoL described above. Each of these inequalities is significant in itself, which underscores the importance of avoiding the presumption that one of them will always encompass all the others. The second conclusion is that, because of the links among the dimensions described above, 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: their combined effect is often to exclude young women from poor households from attending school and getting rewarding jobs, denying them possibilities for self-expression and political voice, and exposing them to hazards that put their health at risk. 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 lives, 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.

Links between dimensions

Measuring QoL in one dimension at a time, however unavoidable that may be given people's expertise and specializations, runs into substantive problems. These can be illustrated by looking at the relation between health, on the one hand, and income, on the other. Taking them separately runs into two problems:

- The first, conceptual, is that income and health are not separable in any reasonable formulation of human well-being: because of this, measuring QoL in one space at a time ignores the interrelations between them, and will lead to incorrect judgments. Broome (2002), for example, argues that people's lives consist of a set of activities that require

both commodities (*e.g.* income) and health (in the same way as the act of reading requires both the eye and the brain). This argues against the parsing of QoL into additive components such as health, education and income.

- The second, empirical, is that measuring QoL by looking at one feature at a time misses the double disadvantage of those who are, simultaneously, sick and poor, as well as the double advantage of those who are, simultaneously, healthy and rich. For example, the standard argument of economists about "Pareto improvements" (*i.e.* the notion that improving the situation of one person, without hurting that of any other, is an unambiguous gain for society as a whole) is powerful and convincing in many situations, but it goes badly wrong when it is applied to one component of QoL while ignoring others. Making the rich even richer in terms of income may be regarded as a Pareto improvement when the income of the poor is unchanged, but this ignores the potentially negative effects of greater income inequalities on the health of the poor – as in the case where the rich become uninterested in supporting universal health care – or on their participation to the political process – as in the case where rich people gain greater political influence. These are real concerns, and casting Pareto arguments in terms of income alone at best misses them, while at worst it rules them out of order.

Both arguments point in favour of taking a more "holistic" approach to QoL, where links between dimensions are systematically assessed. Examples of such inter-relations among QoL dimensions abound. For example, the consequences for subjective well-being of being both poor and sick far exceed the sum of the two effects. Much research is currently devoted to disentangling these relations between different dimensions of QoL, and in distinguishing between associations and causality. The scope for progress, however, is hampered by lack of information on how the main dimensions of QoL are "jointly distributed" across people. In practice, most existing surveys do not allow considering these links, as measurement in one area is typically performed within the remit of existing disciplinary boundaries.

The most obvious way of estimating the joint distribution of the various attributes of QoL would be through a survey in which comprehensive data on all dimensions of QoL were collected for the same sample of people. A less ambitious (but sub-optimal) technique would consist in using different samples for different dimensions, but with enough variables common to the various surveys to allow estimating the joint distribution. This could be achieved by including questions that allow classifying respondents by socio-economic status, education, ethnicity or migrant status within the surveys used in specialized domains. Whatever technique is used, developing information on the joint distribution of various QoL dimensions would constitute real progress.

The importance of better assessing the inter-relations between QoL dimensions goes well beyond measurement and extends to *policy design*. While it may prove impossible to define a complete ordering that covers all QoL dimensions, in which case no summary index of QoL would be feasible, indicators pertaining to different QoL dimensions (or perhaps a selected subset) should be considered jointly when designing policies in specific fields. This would be an improvement in itself, as health, education and various other policies are typically delegated to different agencies, thereby missing their interactions. Moving beyond this "silo thinking" in policy making is one of the greatest appeals of the broad notion of QoL.

Aggregation across dimensions

The search for an aggregate measure of QoL that combines information across all its dimensions is often perceived as the "holy grail" of all efforts to go beyond conventional economic measures. This perspective is, however, both limited and deceptive. Limited, because establishing a comprehensive measurement system for QoL that is capable of producing high-quality information in its various fields is a task which is more difficult and longer-term than that of combining the available information in a single summary measure. Deceptive, as aggregating the various aspects of QoL cannot be accomplished without value judgments that are necessarily controversial: the point already illustrated with regard to "combined" measures of health status has a more general validity.

Despite this, the challenge of providing a more parsimonious description of QoL than the one provided by scores of non-monetary indicators is real. According to one view, only sub-domain indicators should be produced, leaving it to users to form their opinion about QoL. Others argue that a synthetic indicator has unique political clout, and cite the influence of GDP as proof that such indices are essential. Each stance has its strong and weak points. On the one hand, leaving it to users to make the synthesis supposes that users (the media, politicians, the general public) are equipped to do this consistently, which is sometimes dubious: the most likely consequences of producing only sub-domain indicators could well be that GDP will remain the dominant indicator, and that statistics produced with no concern for a synthesis will fail to supply appropriate data. On the other hand, the search for a synthetic indicator risks embroiling statistical institutions in political debates that put their neutrality at risk.

Various methods of aggregation across QoL dimensions exist, and they relate to the approaches described earlier in this chapter. These methods can be described as providing answers to different questions, such as the ones detailed below. These questions and answers closely map into the theoretical approaches to QoL discussed earlier in this chapter: while the first and second questions in this list relate to the capability approach, the third and fourth questions relate to the subjective well-being approach, and the last question relates to the approach based on the theory of fair allocations:

1. *Is society doing well?* Answering this question requires aggregating information on *average scores* of various indicators across QoL domains.

2. *Are people living well?* This question put the emphasis on the *conditions of each individual* in society. Answering this question requires aggregating indicators of QoL at the individual level, and then summarizing this information at the level of entire countries.

3 *Are people happy in their life?* In this question, the emphasis is on people's hedonic experiences. Answering this question requires *aggregating the various hedonic experiences for each person*, and then finding a suitable summary for the country as a whole.

4 *Are people satisfied with their lives?* This question focuses on *evaluative judgments made by individuals over their lives as a whole*. Answering this question requires aggregating measures of satisfaction across people.

5 *Do people have the quality of life they want?* This question retains the individual as the focus of analysis, and requires information on how much they would be willing to

sacrifice in one dimension of their QoL in order to obtain some reference level in other dimensions. One approach to answering this question relies on the concept of "equivalent income".

These are separate questions, and which of them is most relevant depends on the context and on ethical choices between the various approaches. Because they are all related to approaches that are defended and considered legitimate by respectable contributors to the debate about social progress, one way out of the dilemma of proposing a single summary measure of QoL is that statistical institutions construct *several* synthetic indicators, responding to the different questions above. This would preserve their neutrality while providing sufficient and consistent data for political debate. A related possibility would be to provide indicators in which some parameters, in particular those reflecting value judgments, can be modified by the users (as already done on some websites that propose alternative synthetic indicators).

This next section reviews the strengths and weaknesses of various approaches to aggregation; it omits the life satisfaction method, as this was extensively discussed in a preceding section of this chapter.

Aggregating averages across domains

The first method of aggregation is based on the idea of deriving a single composite index by combining indicators of the *average* conditions of each country in several domains. This method is related to the capability approach but also, more broadly, to the social indicator movement. [86] The weights of the various components of this composite indicator can be chosen based on a perfectionist approach that considers the various aspects of life as having an objective importance, or based on a more subjective approach that takes account of some typical or average preferences of the population. Researchers working with this latter perspective often advocate participatory procedures for identifying the domains that matter more and for weighting across them.

The main advantage of this method of aggregation is its simplicity, and its small data requirements. Indicators like the UNPD's Human Development Index have attracted huge media attention, and generated country rankings that depart significantly from those based on conventional economic measures. Most applications of this approach rely on objective data for various domains, but they could easily be extended to incorporate subjective well-being as one domain. Because of the simplicity of this approach, it is easily communicated and understood by the general public, and many grass-root movements have favoured various applications of this approach. Nevertheless, this approach has a number of limits:

- The first limit is that, by retaining the notion of a "representative agent", it cannot track the accumulation of disadvantages by certain subgroups. The combined index will not improve if the correlation of inequalities across domains falls while average performance in each domain remains the same. In practice these composite indicators try to compensate for this limit by including measures of inequality or poverty as a

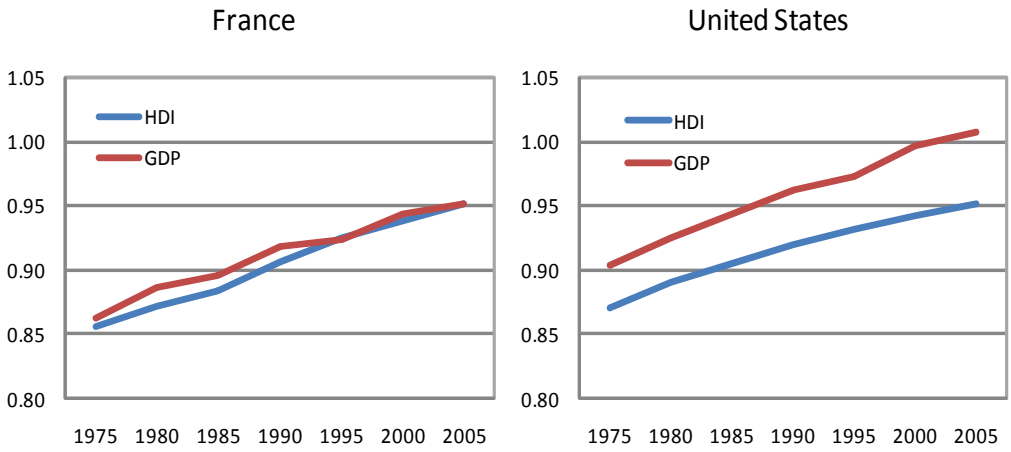
86. The social indicator movement was particularly active in the 1960s and 1970s. Key contributions from this research are gathered in the journal, *Social Indicators Research,* founded in 1974, which publishes research dealing with measurement of the quality of life.

specific component. However, this does not overcome the methodological problem of neglecting individual conditions.

- A second limit is related to the choice of weights for various domains. The weights used to aggregate averages for various domains are conventional, and even the choice of using unweighted data is a value judgement with important implications. For example, the HDI is a simple average of life expectancy – the ratio life expectancy minus 20, divided by 85 (maximal life expectancy) minus 20 – education – two-thirds of the adult literacy rate plus one-third of the enrolment rate for primary, secondary and tertiary – and income (the log of per capita GDP minus the log of 100, divided by the log of $40,000 minus the log of 100). However, adding the *logarithm* of GDP to the *level* of life expectancy implicitly values an additional year of life in each country by its GDP per capita (Ravallion, 1997), thus treating an additional year of life expectancy in the US as worth 20 times a year of life expectancy in India and nearly 50 times a year of life expectancy in Tanzania.
- A third limit has to do with the interpretation of *changes* in these aggregate indicators. For example, levels of the HDI give a fresh look at the world, since the country-rankings it generates are quite different from those based on GDP per capita. However, as time passes and the HDI is updated year to year, its movements have tended to be dominated by changes in the GDP component, at least for those developed countries (such as France and the United States) whose performance in the health and education domains is close to the top (Figure 2.7)[87].
- A final drawback is that this approach does not allow for a diversity of viewpoints about the relative importance of various dimensions of QoL. A single set of weights is applied to the whole society, independently of the heterogeneity of people's attitudes towards income, leisure, health, education, etc. While, in principle, different weights could be used for different countries, this would make the choice of weights even more delicate and effectively prevent any comparisons across countries.

87. However, when extended to the full range of countries at different levels of economic development, the correlation between GDP growth and changes in the HDI over the period 1990-2006 is only 0.44. In particular, some poor countries, such as Egypt, Tunisia and Bangladesh, have recorded significant improvements in their HDI with only moderate economic growth, while others have experienced significant economic growth and seen their HDI drop. This highlights the fact that, over reasonably long time periods, GDP growth can correlate poorly with changes in the non-income dimensions of well-being.

Figure 2.7. Trends in the HDI and in GDP per capita in France and the United States



Values of the HDI as reported in the UNDP Human Development Report 2007-2008. Trends in the GDP series are based on OECD values of GDP per capita at constant prices, based on the same transformation underlying the HDI calculation (*e.g.* the difference between the natural logarithm of GDP per capita in a given year and the log of 100 USD, divided by the difference between the log of 40000 USD and the log of 100).

Source: OECD and UNDP data.

Despite these limits, the websites of most statistical institutes and international organizations already provide scores of average indicators for various domains of QoL. These websites could then also provide options to users (in terms of the scaling of the domain indices and the choice of their weights) for constructing composite measures based on aggregating average measures for domains. Some of these options could be earmarked as having been chosen by authors or institutions, but the official websites would make clear that the institutes are not "endorsing" the indices and the corresponding weights. This solution is technically easy to implement, with the only hard choice being the choice of domains and individual indicators.

Aggregating indicators at the individual level

The second method aggregates over various QoL dimensions at the level of each person, and then computes a mean (or a similar summary measure) over all people in the sample for each country. This approach, for example, would allow combining a measure of income, health and education for each person, before computing measures of "average" QoL of the entire population. The ethical foundations of this approach are similar to those for the indicator based on aggregating averages across domains, but it treats the person as the unit of moral concern. This, in turn, allows recording the correlations across sub-domains. While this approach is most commonly applied to objective dimensions, it can be extended to include subjective aspects of QoL as well.

This approach is already used to construct measures of multidimensional poverty, *i.e.* measures that count people who are deprived in several dimensions. The same approach can be extended to QoL based on a "top sufficiency" cut-off, which would consider a person as having sufficient QoL if he or she achieves sufficiency in some range of dimensions (Alkire, 2008). While this type of measure is generally computed with respect to people having

insufficient achievements, it could be converted to represent QoL by subtracting the insufficiency achievements from 100%.

This approach shares the strengths of the previous one (*i.e.* it relies on objective data and it is simple to compute) but also represents a big improvement relative to a composite index based on average indicators by domain. It has, however, been little used in practice, as it requires information on the "joint" distribution of various attributes across individuals. Beyond this difficulty, however, this approach shares similar weaknesses with the approach based on aggregating average indicators by domain.

- The first limit is that of choosing the weights for the construction of an individual index of QoL. One path often used in studies of multidimensional poverty is that of *counting* the number of items in which the QoL of an individual is below a cut-off level, and then considering the sum of these items. This sum, in turn, can be based on either unweighted items (*i.e.* all have the same importance) or on weighting them based on the statistical properties of the sample (for example, with greater weight attributed to deprivations in those items that are most common in each country). Such weights, however, carry no ethical authority. While different schools of thought have suggested different approaches to setting weights within this approach, they all ignore the diversity of preferences in the population.[88]

- A second limit is that most indices based on this approach focus on people at the bottom of the QoL scale, rather than on the entire distribution. These approaches then rely on (arbitrary) cut-offs that are used to classify people in terms of dichotomous variables (poor or not-poor), which are then aggregated at the individual level. While the focus on those in the lower part of the QoL scale may be justified for policy purposes, it is not ideal in other respects, as inequalities in QoL, rather than insufficient achievements, may matter *per se*. To develop aggregate measures of QoL that are informative about the entire distribution would require developing scales that can be compared across domains.

Aggregation based on people's hedonic experiences

Beyond people's own evaluations of their lives as a whole, which – as argued above – may be regarded as an aggregate indicator of a certain sort, approaches to subjective well-being based on hedonic experiences suggest a different path to aggregating across some of the factors that shape QoL. While this approach may still be considered as putting a domain – hedonic experiences – at the forefront, it may also be regarded as providing a way of weighting different experiences through a common yardstick: the intensity of the hedonic experiences that they generate.

88. Rawls (1982) suggests that weights should reflect the typical ("representative") preferences of individuals from the worst-off group. Sen (1985, 1992) suggests seeking a partial ranking of individual situations based on a consensus between conflicting preferences. Sen also notes that, whatever one's views about the relative importance of the various QoL domains, being better-off in each domain should imply being better-off overall. This "dominance" principle, however, would also imply that individual attitudes about the importance of the different domains play no role in interpersonal comparisons: for example, the dominance principle would consider Jones as better off than Smith if the two have the same low level of health but the former is slightly richer than the latter – even if Jones may care much more than Smith about being deprived in health.

Measures of the flow of hedonic experiences are intimately connected to the use of time. Experiences occur over time, and time use is an objective feature of experience. Moreover, positive and negative affect are systematically related to different uses of time. Time is the ultimate finite resource for people and society, and traditional GDP measures do not reflect the quantity of individuals' non-work time – indeed, an increase in leisure time will generally *lower* GDP, despite its positive contribution to QoL. Moreover, attempts to value leisure time in GDP reflect only the marginal valuation of leisure time, while major changes in society that influence the amount and use of leisure time cannot be valued in this perspective.

The U-index is one way of blending objective and subjective measures of how people spend and experience their time. This index measures the proportion of one's time in which the strongest feeling is a negative one (Kahneman and Krueger, 2006). Positive feelings include happiness and enjoyment while negative feelings include feeling frustrated, depressed, hassled, angry, worried, or criticized. More precisely, using the *Day Reconstruction Method* – in which respondents describe particular episodes in their daily experience – an episode is classified as unpleasant if the most intense feeling reported for that episode is a negative one, *i.e.* if the highest rating on any of the negative affects is strictly greater than the maximum rating of any of the positive affects.

While the U-index shares some of the weaknesses described earlier with respect to all subjective measures of well-being, it has properties that make it attractive from the perspective of measuring QoL:

- First, it addresses the issue of non-comparability in the use of scales by different people. To the extent that interpersonal differences in the interpretation of scales apply equally to positive and negative emotions (*e.g.* people who report a more positive evaluation of some states will also provide less negative evaluations of other states), then the U-index will be robust to some barriers to interpersonal and international comparisons. In this sense, the U-index provides an ordinal measure at the level of feelings.

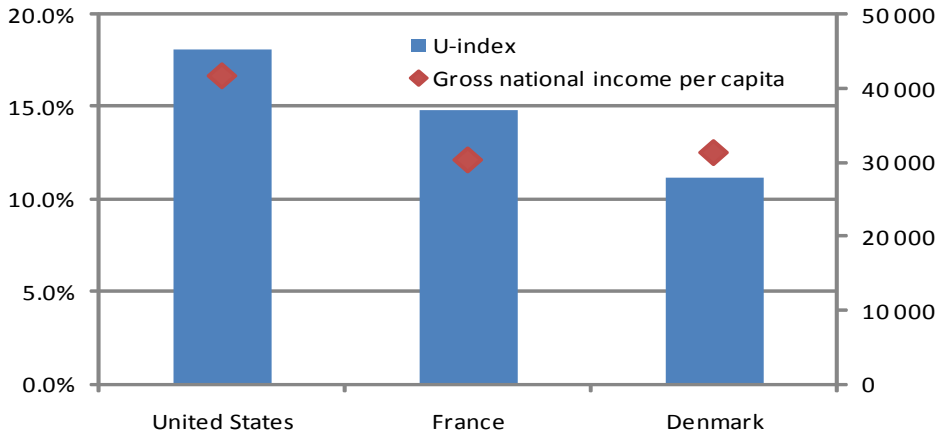
- Second, the U-index naturally focuses attention on the most miserable in society, in the same way as the conventional poverty rate does with respect to income. Evidence from available experimental measures of this type does suggest that a very large share of the negative feelings in society is experienced by a small number of people.[89]

The experimental measures of the U-index that are currently available (limited to specific cities and demographic groups) highlight significant differences across countries, with a ranking that differs significantly from the one obtained based on GDP per capita (Figure 2.8) or life satisfaction.[90] Implementing this measure at the national level would require collecting data on "evaluated" time use, *i.e.* adding questions about emotional experience during specific episodes in ongoing time-use surveys. Steps in this direction are envisaged in some OECD countries (*e.g.* the United States) and could be generalized to others (*e.g.* in the context of the next collection of time-use data for European countries).

89. An alternative approach to aggregation based on people's hedonic experiences would be to measure the proportion of one's time in which the strongest feeling is a positive one. This approach, however, would focus attention on the most fortunate in society.

90. Blanchflower (2008) argues, however, that the structure of U-index equations is fairly similar to those based on life-evaluations.

Figure 2.8. Measures of the U-index in three cities



Note: Estimates for the U-index refer to women aged 18-68 who are not full-time students in the three cities of Columbus, Ohio (United States), Rennes (France) and Odense (Denmark), as provided by Alan Krueger. The U-index is the proportion of time that the maximum rating of being tense, depressed, or angry exceeds the rating of being happy. Measures of GDP per capita in 2008 are OECD estimates.

Aggregation by the equivalence approach

While various indexing methods pertaining to the equivalence approach exist, one that is easily applied is based on the notion of equivalent income, *i.e.* the income that would make each person indifferent between their current situation and a situation based on their current income and some reference levels for the non-monetary dimensions of QoL.[91]

The main strengths of this method are the following:

- First, it partly avoids the "hedonic treadmill" and the "aspiration treadmill" problems, *i.e.* the risk of underestimating the disadvantages of deprived individuals who adapt to their situation and report high life evaluations and hedonic experiences. As this method only relies on the ordinal rankings of situations defined by individual preferences, it does not consider that people are better-off when they obtain high satisfaction by lowering their aspirations.[92] They are declared well-off only when they obtain what they want.

Second, it satisfies the property that, if two individuals have the same preferences in life and agree about which of them is better-off, the method will lead to the same ranking.[93] The

91. To illustrate, imagine that Jones, with income of €20,000 and reduced mobility, would consider himself equally well-off with good health and an income of €15,000. Taking good health as the reference for health, €15,000 is his equivalent income. So Jones is considered worse off than Smith who is healthy and whose income is €17,000.

92. This method, however, does not correct the evaluation when the *direction* of individual preferences is shaped by adaptation (for instance, when individuals come to like cheap but unhealthy food). Correcting for this problem would require some kind of paternalistic definition of what is good for people.

93. To illustrate, suppose that Jones and Smith above have identical preferences and therefore agree that Jones is worse off, because they consider that his reduced mobility is worth more than his extra €3,000 income. In this situation, Jones is considered worse-off because his equivalent income is lower.

idea of respecting individual preferences is usually limited in welfare economics to the respect of preferences in one's own life, but it seems sensible also to respect interpersonal comparisons when they are based on common preferences.

This approach, however, also has weaknesses:

- First, it requires that people have well-defined preferences about the various aspects of life. In practice, this condition does not always apply, as inconsistency and instability plague the choices of people in the real world. In these conditions, this approach requires that people have some consistent values at a deeper level, although these may be hard to observe.

- Second, the choice of reference values for the non-monetary dimensions of QoL opens many possibilities and involves difficult ethical choices. As this choice may not always be consensual (although it may be for some dimensions such as health), it is one area where users could be left some choice.

It should be noted that the equivalent income should not be interpreted as giving a special value to money over other aspects of life. The weights given to non-monetary aspects of QoL reflect individual preferences. For instance, if people give great importance to health in their life, this will give a great weight to health in the computation of their own equivalent income. In other terms, the incorporation of non-monetary aspects of QoL in this measure does not require a non-monetary unit of measurement.

Assuming that people have consistent and stable preferences, these can be retrieved from three sources: observed choices[94]; stated preference surveys (*e.g.* contingent valuation surveys[95] or discrete choice experiments); and satisfaction surveys (if it is possible to filter the adaptation phenomenon). These techniques are standard in economics, and have in some cases been applied to compute equivalent incomes.[96] If one does not trust individuals' immediate preferences and seeks to track their deeper preferences, a more refined methodology might be needed, and this has not yet been developed.

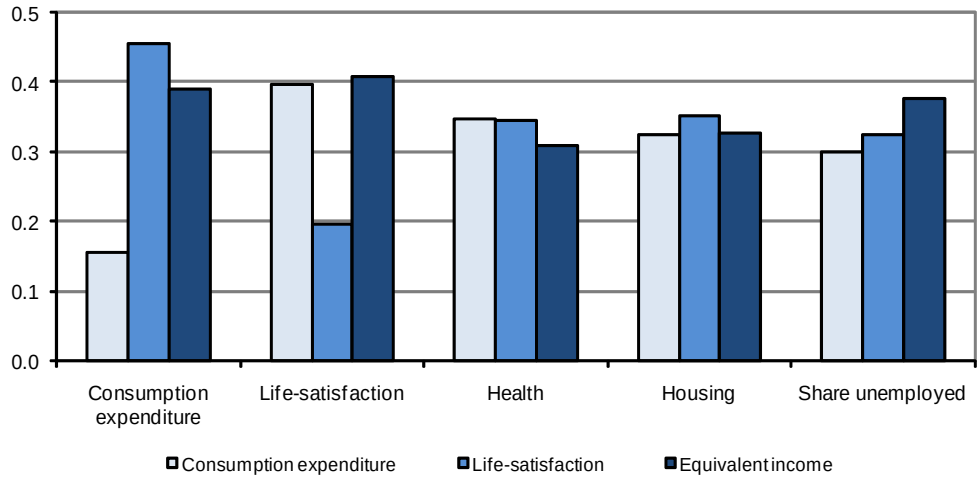
94. Revealed preferences are fruitful only for aspects of life that can be traded off by individuals and cannot be used to estimate preferences about aspects that are imposed by circumstances.

95. Contingent valuation surveys have been criticized for being unreliable when individuals are asked how much they would be willing to contribute to more or less remote aspects of the environment (see in particular Kahneman *et al.*, 1999; and Diamond and Hausman, 1994). As far as the equivalent income is concerned, the questions would be directly about their own situation, which may partly alleviate some of these problems. Fleurbaey *et al.* (2009a) have asked respondents what sacrifice in their consumption they would have accepted over the last 12 months if they could have avoided the health problems they endured during the same period, and their results indicate that this willingness-to-pay increases with income and with the severity of health problems. But there is also substantial evidence of the systematic biases that you get in asking people to value their health, as reviewed by Dolan and Kahneman, 2008.

96. Muellbauer (1974a,b) and King (1983) have computed money-metric utilities on the basis of household demand data. Browning *et al.* (2006) use similar data to compute individual equivalent incomes that correct for household composition (*i.e.* they estimate what income would be sufficient for an individual to reach his current indifference set on consumption, if he were single). Fleurbaey *et al.* (2009b) rely on a satisfaction survey to compute equivalent incomes in Russia by correcting for various aspects of QoL such as health, housing, unemployment and wage arrears. Fleurbaey and Gaullier (2007) present aggregate estimates of equivalent incomes (taking account of hours worked, longevity, unemployment, household composition) and show that the ranking of OECD countries is significantly different compared to the one based on GDP per capita.

Despite these difficulties, experimental applications of this equivalent income approach provide insights into the characteristics of the worse-off in society that differ from those based on other approaches. Figure 2.9 shows the average characteristics of people classified as "worse-off" (*i.e.* in the bottom quintile of the distribution) based on three measures of their quality of life (household expenditure per consumption unit, reported life-evaluations, and an estimate of their equivalent income, where the weights applied to various non-monetary dimensions are estimated from an equation referring to people's life-satisfaction) in a sample of Russian respondents. People in the bottom quintile of the distribution of equivalent income report lower expenditure, health and housing quality and a higher incidence of unemployment compared to people in the bottom quintile of life-satisfaction. In other terms, the equivalent income approach seems to identify a sub-population that cumulates disadvantages in many more dimensions of their lives compared to approaches that rely on either monetary expenditure or life-evaluations (Fleurbaey *et al.*, 2009).

Figure 2.9. Characteristics of the most deprived people according to different measures of quality of life, Russia in 2000



Note. Data refer to people considered as "worse-off" (*i.e.* in the bottom quintile of the distribution) according to three different measures of their quality of life: i) household expenditure per consumption unit (*i.e.* a standard measure of living standard); ii) life-satisfaction (based on the question "*To what extent are you satisfied with your life in general at the present time?"* with answers on a five-point scale); and iii) a measure of equivalent income, based on four "functionings", *i.e.* self-reported health, employment status, quality of housing, and having incurred wage arrears. For each of these three measures of quality of life, the figure plots the average levels of various factors shaping quality of life among the "worse-off" based on one measure, relative to its average level across the three measures. Estimates of the equivalent income of people in the sample are based on preferences estimated from an equation on the determinants of life-satisfaction**.** Reference values for the various non-monetary dimensions correspond to "good health", "not being unemployed", "not having wage arrears", and "having a median value of housing quality".

Source: Fleurbaey at al. (2009), based on data from the Russia Longitudinal Monitoring Survey.

Summing-up

These different approaches to aggregation across domains of QoL may be viewed as providing answers to different questions. Despite these differences, research in these various approaches appears to have progressed sufficiently to make them worthy of measurement by government statistical agencies. Developing more systematic data about the various features of the QoL holds the promise of improving policy-making and delivering a more comprehensive description of the progress of societies.

Key messages and recommendations

The argument in this report can be summarized through the following points:

- Quality of life includes the full range of factors that make life worth living, including those that are not traded in markets and not captured by monetary measures. While some extensions of economic accounting allow including some additional elements that shape quality of life in conventional money-based measures of economic well-being, there are limits to the extent to which this can be achieved. Non-monetary indicators have an important role to play in measuring social progress, and recent advances in research have led to new and credible measures for at least some aspects of quality of life. These measures, while not replacing conventional economic indicators, provide an opportunity to enrich policy discussion and to inform people's view of the conditions of the communities where they live; today, they have the potential to move from research to standard statistical practice.

- Recent research has shown that it is possible to collect meaningful and reliable data on subjective well-being. Subjective well-being encompasses three different aspects: cognitive evaluations of one's life, positive emotions (joy, pride), and negative ones (pain, worry, anger). While these different aspects of subjective well-being have different determinants, in all cases these determinants go well beyond people's income and material conditions. For example, all these measures agree in highlighting that unemployment has major damaging effects on the subjective well-being of individuals and nations. All these aspects of subjective well-being should be measured separately, to derive a more comprehensive measure of people's quality of life and to allow a better understanding of its determinants (including people's objective conditions). National statistical agencies should incorporate questions on subjective well-being in their standard surveys to capture people's life evaluations, hedonic experiences and life priorities.

- Quality of life also depends on the objective conditions and opportunities available to people. These "capabilities" are considered by some analysts to matter more than just through their effects on people's subjective states, and should be seen rather as basic conditions of people's autonomy. How societies are organized makes a difference to people's lives, as can be seen in measures of people's heath and education; their daily work and leisure activities; citizens' political voice and the responsiveness of institutions; people's social connections and their environmental conditions; and the physical and economic insecurity that shapes their lives. The challenge in these fields is to improve upon what has been achieved already, to develop recognized statistical standards in various fields, and to invest in statistical capacity in areas where available indicators remain deficient (*e.g.* insecurity). This is particularly the case for data on how

people spend their time (and on their enjoyment in performing these activities): developing such data at regular intervals and based on standards that allow comparisons across countries and over time is an important priority.

- Indicators of quality of life should inform about the inequalities in individual experiences. This is important, as social progress depend not only on the *average* conditions in each country but also on the *inequalities* in people's conditions. Accounting for the diversity of experience (by gender, across groups and generations) is necessary to fill the gap between country-wide estimates and people's feelings about their own conditions. Inequality in each of the dimensions of quality of life is significant in itself, and this underscores the importance of avoiding the presumption that any single dimension will always encompass all the others. At the same time, because of the links between the dimensions of quality of life, various inequalities may also strengthen each other.

- Some of the most important policy questions for quality of life relate to how developments in one area affect those in others, and how developments in various fields are related to those in income. The consequences for quality of life of having multiple disadvantages far exceed the sum of their individual effects. Developing measures of these cumulative effects requires information on the "joint distribution" of the most salient features of quality of life (such as affect, health, education, political voice) across all people in a country. While the full development of this information could probably be achieved only in the distant future, concrete steps in this direction could be taken by including in all surveys some standard questions that allow classifying respondents based on a limited set of characteristics.

- The search for a scalar measure of quality of life is often perceived as the most important challenge for quality-of-life research. While this emphasis is partly (or largely, according to some) misplaced, this report recognizes the strong demand in this field and believes that statistical offices have a role to play in answering it. Several scalar measures of quality of life are possible, depending on the question addressed and the approach taken. Some of these measures are already being used (*e.g.* average levels of life-satisfaction for a country as a whole, or composite indices that aggregate averages across domains, such as the *Human Development Index*). Others could be implemented if national statistical systems made the necessary investment to provide the 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 existing time-use surveys. Similarly, the approach based on counting the occurrences and severity of various objective features of people's lives requires information on the joint distribution of these features, while the approach based on the notion of "equivalent income" also requires information on individuals' preferences with respect to these items. Statistical systems should provide the information needed to allow the computation of several aggregate measures of quality of life.

REFERENCES

Abraham, K., and C. Mackie, eds. (2005), *Beyond The Market: Designing Nonmarket Accounts for the United States*, National Academies Press, Washington DC.

Agarwal, B. (1994), *A Field of One's Own: Gender and Land Rights in South Asia*, Cambridge University Press, Cambridge.

Agarwal, B. and P. Panda (2007), "Toward Freedom from Domestic Violence", *Journal of Human Development*, 8(3).

Alkire, S. (2008), "The Capability Approach to the Quality of Life", background report prepared for the Commission on the Measurement of Economic Performance and Social Progress, Paris.

Alkire, S. (2008), "The Capability Approach as a Development Paradigm?", mimeo.

Alkire, S. (2002), *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction*, Oxford University Press, Oxford.

Alonso, J. *et al.* (2004), "Prevalence of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMED) project", *Acta Psychiatrica Scandinavica*, Volume 109, Supplement 420.

Anand, S. and K. Hanson (1997), "Disability adjusted life years: a critical review", *Journal of Health Economics*, 16(6).

Arendt, J.N. (2005), "Does education cause better health? A panel data analysis using school reforms for identification", *Economics of Education Review*, 24.

Arrow, K.J. (1951), *Social Choice and Individual Values*, New York: Wiley.

Banks, J., M. Marmot, Z. Oldfield, and J. Smith (2006), "Disease and disadvantage in the United States and England", *JAMA*, 295.

Becker, G., T. Philipson and R. Soares (2005), "The quantity and quality of life and the evolution of world inequality", *American Economic Review*, 95.

Berkman, L. F. and T. Glass T. (2000), "Social integration, social networks, social support, and health", in Berkman and I. Kawachi (eds.), *Social Epidemiology*, Oxford University Press, Oxford.

Boadway, R. and N. Bruce (1984), *Welfare Economics*, Basil Blackwell, Oxford.

Broome, J. (2002), "Measuring the burden of disease by aggregating well-being", in C. J. L. Murray, J. A. Salomon, C. D. Mathers, and A. D. Lopez (eds.), *Summary measures of population health*, World Health Organization, Geneva.

Browne, A., A. Salomon and S. Bussuk (1999), "The Impact of Recent Partner Violence on Poor Women's Capacity to Maintain Work", *Violence Against Women*, 5 (4).

Browning, M., P.A. Chiappori and A. Lewbel (2006), "Estimating consumption economies of scale, adult equivalence scales, and household bargaining power", Boston College Working Paper No. 588, Boston.

Burchardt, T. (2005), "One man's rags are another man's riches: Identifying adaptive preferences using panel data", *Social Indicators Research*, 74.

Carleton University, Country Indicators for Foreign Policy, http://www.carleton.ca/cifp/gdp_indicator_descriptions.htm

Carrillo, R. (1992), Battered Dreams: Violence against Women as an Obstacle to Development, United Nations Fund for Women, New York.

Case, A. C. and C. H. Paxson (2005), "Sex Differences in Morbidity and Mortality", *Demography*, 42(2).

CERC (2005), *La sécurité de l'emploi face aux défis des transformations économiques*, Conseil de l'emploi, des revenus et de la cohésion sociale, Rapport No. 5, La Documentation Française, Paris.

CERC (2006), *La France en transition, 1993-2005*, Conseil de l'emploi, des revenus et de la cohésion sociale, Rapport No. 7, La Documentation Française, Paris.

Chaeyoon, Lim and Robert D. Putnam, "Praying Alone is No Fun: Religion, Social Networks, and Subjective Well-being" (unpub. ms, 2009).

Clark, A.E., P. Frijters and M. Shields (2007), "Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles", *Journal of Economic Literature*, vol. 46, no. 1, March.

Clark, A.E. and A.J. Oswald (1994), "Unhappiness and unemployment", *Economic Journal*, 104.

Cohen, S., C. M. Alper, W. J. Doyle, *et al.* (2006), "Positive emotional style predicts resistance to illness after experimental exposure to rhinovirus or influenza A virus", *Psychosomatic Medicine* 68 (6).

Cutler, D., A. Lleras-Muney and T. Vogl, T. (2008), "Socioeconomic status and health: Dimensions and mechanisms", *National Bureau of Economic Research*, WP 14333, Boston.

Dannenberg, A.L. D.M. Carter, H.W. Lawson, D.M. Ashton, S.F. Dorfman and E.H. Graham (1995). "Homicide and Other Injuries as Causes of Maternal Death in New York City, 1987 through 1991". *American Journal of Obstetrics and Gynecology*, 172.

Deaton, A. (2008), "Income, Health and Well-Being around the World: Evidence from the Gallup World Poll." *Journal of Economic Perspectives* 22 (2): 53–72.

Deaton, A. (2006), "Global Patterns of Income and Health: Facts, Interpretations, and Policies", WIDER Annual Lecture 10, United Nations University, World Institute for Development Economics Research, Helsinki.

Deaton, A., J. Fortson and R. Tortora (2008), "Life (evaluation), death and HIV/AIDS in Africa", Research Program in Development Studies, Princeton.

Deaton, A. and J. Muellbauer (1980), *Economics and consumer behaviour*, Cambridge University Press, Cambridge.

Dee, T.S. (2004), "Are there civic returns to education?", *Journal of Public Economics*, No. 88.

De Walque, D. (2007), "Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education", *Journal of Health Economics*, no. 26.

De Witte, H. and K. Näswal, (2003) "'Objective' vs. 'Subjective' job insecurity: consequences of temporary work for job satisfaction and organizational commitment in four European countries". *Economic and Industrial Democracy*, 24.

van Doorslaer, E., C. Masseria and the OECD Health Equity Research Group Members (2004), "Income-related inequality in the use of medical care in 21 OECD countries", OECD Health Working Papers No. 14, OECD, Paris.

DeNavas-Walt, C., B. D. Proctor and J. Smith (2007), "Income, Poverty, and Health Insurance Coverage in the United States: 2006", U.S. Census Bureau, August.

Diamond, P. A. and J. A. Hausman (1994), "Contingent valuation: IS some number better than no number?", *Journal of Economic Perspectives*, 8.

Di Tella, R., R.J. MacCulloch, and A.J. Oswald (2003), "The macroeconomics of happiness", *Review of Economics and Statistics*, 85(4).

Diener, E. (1984), "Subjective Well-Being", *Psychological Bulletin*, 93: 542-575.

Dolan, P and D. Kahneman (2008), "Interpretations of utility and their implications for the valuation of health", *Economic Journal*, 118 (525)

Dowrick, S., Y. Dunlop, J. Quiggin (2003), "Social indicators and comparisons of living standards", *Journal of Development Economics* 70.

Dreze, J. and A.K. Sen (2002), *India. Development and Participation*, Oxford University Press, Oxford.

Edleson, J.L. (1999), "Children's Witnessing of Adult Domestic Violence" *Journal of Interpersonal Violence*, 14 (8).

Edwards, R D. and S. Tuljapurkar (2005), "Inequality in Life Spans and a New Perspective on Mortality Convergence across Industrialised Countries", *Population and Development Review*, Vol. 34, No. 4, December.

EU (2001), *White Paper on Noise*, ECC, Brussels.

Families USA (2004), "Health issues 2004", available at http://www.familiesusa.org/health-issues-2004.

Fleurbaey, M. (2008a), "Individual well-being and social welfare: Notes on the theory", background report prepared for the Commission on the Measurement of Economic Performance and Social Progress, Paris.

Fleurbaey, M. (2008b), "Ethics and Economics", in S. N. Durlauf and L. E. Blume (eds.), *New Palgrave Dictionary of Economics*, Second Edition, Palgrave Macmillan.

Fleurbaey, M. and G. Gaullier (2007), "International Comparisons of Living Standards by Equivalent Incomes", Centre d'Etudes Perspectives et d'Informations Internationales, Working Paper No. 03, Paris.

Fleurbaey, M., S. Luchini, C. Muller and E. Schokkaert (2009a), "Equivalent income and economic evaluation of health care", mimeo.

Fleurbaey, M., E. Schokkaert and K. Decancq (2009b), "What good is happiness?", CORE Discussion Paper 2009/17, Universitè catholique de Louvain, Belgium.

Fowler, J. H. and N. A. Christakis (2008), "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study", *British Medical Journal*, vol. 337, December.

Frank, R. H. (2008), "Should public policy respond to positional externalities?", *Journal of Public Economics*, No. 92.

Frey, B. S. and A. Stutzer (2002a), *Happiness and economics*, Princeton University Press, Princeton and Oxford.

Frey, B.S. and A. Stutzer(2002b), "What can economists learn from happiness research?", *Journal of Economic Literature*, 40(2), June.

Fuchs, V. R. and P. Farrell (1982), "Schooling and Health: The Cigarette Connection", *Journal of. Health Economics*, 1(3), December.

Gilbert, D. (2005), *Stumbling on Happiness*, Vintage Books, New York.

Gimenez Nadal, J. I. and A. Sevilla Sanz (2007), "A Note on Leisure Inequality in the US: 1965-2003", Economics Series Working Papers No. 374, Department of Economics, University of Oxford, Oxford.

Grimard, F. and D. Parent (2007), "Education and smoking: were Vietnam war draft avoiders also more likely to avoid smoking?", *Journal of Health Economics*, 26.

Grossman, M. (2008), "Education and non-market outcomes", in E. Hanushek and F. Welch (eds), *Handbook of the Economics of Education*, North-Holland, Amsterdam.

Groot, W., H. Maasen and H. van den Brink (2007), "The health effects of education", *Economics of Education Review*, 26.

Harper, M. and L. Parsons (1997), "Maternal Deaths due to Homicide and Other Injuries in North Carolina: 1992-1994", *Obstetrics and Gynecology*, 90 (6).

Helliwell, J. F. (2008), "Life Satisfaction and Quality of Development", working paper 14507, National Bureau of Economic Research, Cambridge.

Helliwell, J.F. (2005), "Well-Being, Social Capital and Public Policy: What's New?" NBER Working Paper No. 11807, (2005), National Bureau of Economic Research, Cambridge, United States.

Helliwell, J. F. (2001), "Social Capital, the Economy and Wellbeing", in *The Review of Economic Performance: The Longest Decade: Canada in the 1990s*, Vol. 1, Centre for the Study of Living Standards, Ottawa, Canada.

Helliwell, J. F. and Huang, H. (2005), “How's the job? Well-being and social capital in the workplace”, paper presented at the Annual Meeting of the Canadian Economics Association, McMaster University, May.

Helliwell, J. F. and R. D. Putnam (2004), “The social context of well-being”, Philosophical Transactions of the Royal Society of London: Biological Sciences, London.

Hetzel, A. M. (1997), *History and Organization of the Vital Statistics System*, National Center for Health Statistics, US Department for Health and Human Services, Hyattsville, Maryland.

Himmelstein, D. U., E. Warren, D. Thorne and S. Woolhandler (2005), “Illness and Injury as Contributors to Bankruptcy”, *Health Affairs*, February.

Hodiamont, P.P.G., Rijnders, C.A.T., Mulder, J. and Furer, J.W. (2005). “Psychiatric disorders in a Dutch Health Area: a repeated cross-sectional survey.” *Journal of Affective Disorders*, vol. 84.

Janicki-Deverts, D., S. Cohen, W. J. Doyle *et al.* (2007), “Infection-induced proinflammatory cytokines are associated with decreases in positive affect, but not increases in negative affect”, *Brain Behavior and Immunity*, 21 **(**3**).**

Kahneman, D., I. Ritov and D. Schkade (1999), “Economic preferences or attitude expressions? An analysis of dollar responses to public issues”, *Journal of Risk and Uncertainty*, 19.

Kahneman, D. and A. Krueger (2006), "Developments in the Measurement of Subjective Well-Being”, *Journal of Economic Perspectives*, 20 (1): 3-24.

Kahneman, D., E. Diener, E. and N. Schwarz. eds. (1999), *Well-being: The foundations of hedonic psychology*, Russell Sage Foundation, New York.

Kaufmann, D., A. Kraay and M. Mastruzzi (2008), “Governance Matters VII: Aggregate and Individual Governance Indicators, 1996-2007”, Policy research working paper no. 4654, World Bank Institute, Washington DC.

Kenkel, D., D. Lillard and A. Mathios (2006), “The roles of high school completion and GED receipt in smoking and obesity”, *Journal of Labour Economics*, 24.

Kessler, R.C. *et al.* (2007), “Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organisation World Mental Health Survey Initiative”, *World Psychiatry*, vol. 6, October.

King, M (1983), "Welfare analysis of tax reforms using household data", *Journal of Public Economics*, 21.

Krueger, A.B., D. Kahneman, D. Schkade, N. Schwarz and A. Stone (2008), “National Time Accounting: The Currency of Life”, NBER, forthcoming in A. B. Kruger (ed.), *Measuring the Subjective Well-being of Nations: National Accounts of Time Use and Well-Being*, University of Chicago Press, Chicago.

Krueger, A.B. and D. Schkade (2008), "The Reliability of Subjective Well-Being Measures", *Journal of Public Economics* 92(808): 1833-45.

Krueger, A.B. and A. Mueller (2008), “The Lot of the Unemployed: A Time Use Perspective”, paper prepared for the LoWER conference “Institutions, market and European Unemployment Revisited”, Oxford.

Krueger, A. B. (1999), "Measuring Labor's Share", *American Economic Review*, 89(2), pp. 45-51.

Lafortune, G., Gaëlle Balestat, and the Disability Study Expert Group Members (2007), “Trends in severe disability among elderly people: Assessing the evidence in 12 OECD countries and the future implications”, OECD Health Working Papers No. 26, OECD, Paris.

Layard, R., G. Mayraz and S. Nickell (2008), “Does Relative Income Matter? Are the Critics Right?”, paper prepared for the conference on well-being at Princeton University, October.

Levinson, A. (2009), “Valuing Public Goods Using Happiness Data: The Case of Air Quality », NBER wp15516.

Lim, C. and R. D. Putnam (2009), “Religion and Happiness”, forthcoming.

Lloyd, S. and N. Taluc (1999), “The Effects of Male Violence on Female Employment” *Violence Against Women*, 5.

Luechinger, S. (2009), "Valuing Air Quality Using the Life Satisfaction Approach", *Economic Journal*, forthcoming.

Lundborg, P. and H. Andersson (2008), “Gender, risk perceptions and smoking behaviour”, *Journal of Health Economics*, 27.

Luttmer, E. (2005), “Neighbors as negatives; relative earnings and well-being”, *Quarterly Journal of Economics*, August, 120(3).

Mackenbach, J. P. (2006), “Health Inequalities: Europe in Profile”, UK Presidency of the EU, February.

Maman, S., J. Campbell, M.D.Sweat and A.C.Gielen (2000), “The Intersection of HIV and Violence: Directions for Future Research and Interventions”, *Social Science and Medicine*, 50.

Maniquet, F. (2007), "Social orderings and the evaluation of public policy", *Revue d'Economie Politique*, 117.

McCloskey, L.A. *et al.* (1995), “The Effects of Systematic Family Violence on Children’s Mental Health”, *Child Development*, 66.

Menahem, G. (2007), ”Prestations sociales, sécurité économique et croissance en Europe”, *Revue de l’OFCE,* Centre de Receherche en Economie de SciencePo, Paris.

Milligan, K., E. Moretti and P. Oreopoulos (2004), “Does education improve citizenship? Evidence from the United States and the United Kingdom”, *Journal of Public Economics*, 88.

Moulin, H. and W. Thomson (1997), "Axiomatic analysis of resource allocation problems", in K.J. Arrow, A.K. Sen, K. Suzumura (eds.), *Social Choice Re-examined*, vol. 1, International Economic Association, New York: St Martin's Press and London: Macmillan.

Muellbauer, J. (1974a), "Inequality measures, prices and household composition", *Review of Economic Studies*, 41.

Muellbauer, J. (1974b), "Household composition, Engel curves and welfare comparisons between households", *European Economic Review*, 5.

Murphy, J.M., N. Laird, R. Monson, A. Sobol and A. Leighton (2000), "A 40-year perspective on the prevalence of depression - The Stirling County study", *Archives of General Psychiatry*, Volume 57, Issue 3.

Murphy, K.M. and R.H. Topel (2006), "The value of health and longevity", *Journal of Political Economy*, 114.

Nussbaum, M.C. (2001), "Adaptive preferences and women's options", *Economics and Philosophy*, 17.

Nussbaum, M.C. (2000), *Women and Human Development: The Capabilities Approach*, Cambridge University Press, Cambridge.

Nussbaum, M.C. (2006), "Capabilities as Fundamental Entitlements: Sen and Social Justice", in B. Agarwal, J. Humphries and I. Robeyns (eds.), *Capabilities, Freedom and Equali*ty: *Amartya Sen's Work from a Gender Perspective*, Oxford University Press, Delhi.

Nussbaum, M.C. (2007), "Foreword: Constitutions and Capabilities: 'Perception' against Lofty Formalism", *Harvard Law Review* Vol. 121: 4.

OECD (2009), *Society at a Glance*, forthcoming, OECD, Paris.

OECD (2008a), *Measuring Human Rights and Democratic Governance – Experiences and Lessons from Metagora*, OECD, Paris.

OECD (2008b), *Environmental Outlook*, OECD, Paris.

OECD (2007), *Understanding the Social Outcomes of Learning*, Centre for Educational Research and Innovation, OECD, Paris.

OECD (2006a), *Society at a Glance*, OECD, Paris.

OECD (2006b), *Water, the Experience of OECD Countries*, OECD, Paris.

Okun, M. A. (1975), *Equality and Efficiency – The Big Tradeoff*, The Brookings Institution, Washington D.C.

Office for National Statistics (2006), *Trends in life expectancy by social class*, National Statistics website http://nswebcopy/StatBase/Product.asp?vlnk=8460&Pos=1&ColRank=1&Rank=272

Oreopoulos, P., (2007), ""Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling", *Journal of Public Economics*, vol. 91(11-12), December.

Osberg, L. (1998), "Economic insecurity", SPRC Discussion Paper No. 88, Social Policy Research Centre, University of New South Wales, Australia.

Osberg, L. and A. Sharpe (2002), "An Index of Economic Well-Being for Selected OECD Countries", *The Review of Income and Wealth*, Series 48, Number 3.

Oswald, A. J. (1997), "Happiness and economic performance", *Economic Journal*, 107.

Oswald, A.J. and Powdthavee, N. (2008). "Death, happiness and the calculation of compensatory damages", *Journal of Legal Studies*, December.

Oswald, A.J. and Powdthavee, N. (2007). "Obesity, unhappiness and the challenge of affluence: Theory and evidence". *Economic Journal*, 117.

Owen, A.L, J. Videras, C. Willemsen (2008), "Democracy, Participation and Life Satisfaction", *Social Science Quarterly*, 89(4).

Population Reports (1999), "Ending Violence Against Women", *Issues in World Health*, 27(4).

Powdthavee, N. (2008), "Putting a Price Tag on Friends, Relatives, and Neighbours: Using Surveys of Life Satisfaction to Value Social Relationships", *Journal of Socio-Economics*, volume 37, issue 4.

Putnam, R. D. (2000), *Bowling Alone*, Simon & Schuster.

Putnam, R. D. (1993), *Making Democracy Work – Civic traditions in Modern Italy*, Princeton University Press, Princeton.

Ravallion, M. (1997), "Good and bad growth: the human development reports", *World Development*, 25(5).

Rawls J. (1982), "Social unity and primary goods", in A. Sen, B. Williams (eds.), *Utilitarianism and Beyond*, Cambridge University Press, Cambridge.

Rychen, D.S. and L.H. Salganik (2003), *Key Competencies: For a Successful Life and a Well-Functioning Society*, Hogefe and Huber, Cambridge, MA.

Riley, J. C. (1997), *Sick, Not Dead: The Health of British Workingmen during the Mortality Decline*, Johns Hopkins University Press, Baltimore.

Ringer, S. (2007), "How Good are the Good Democracies?", in S. Ringer, *What Democracy is for?*, Princeton University Press, Princeton and Oxford.

Sacker, A. and Wiggins, R.D. (2002). "Age-period-cohort effects on inequalities in psychological distress". *Psychological Medicine*, vol. 32.

Sampson, R. J. (2003), "Networks and Neighbourhoods – The Implications of Connectivity for Thinking about Crime in the Modern City", in edited by McCarthy H., P. Miller, and P. Skidmore (eds.), *Network Logic: Who Governs in an Interconnected World?*, Demos, London.

SDSA (2008), *State of Democracy in South Asia: A Report*, Oxford University Press, New Delhi.

Sen, A. (1999), *Development as Freedom*, Oxford University Press, Delhi.

Sen, A. (1992), *Inequality Re-examined*, Clarendon Press, Oxford.

Sen, A. (1987a), *The Standard of Living*, Cambridge University Press, Cambridge.

Sen, A. (1987b), *Commodities and Capabilities*, Oxford University Press, Oxford.

Siedler, T. (2007), "Schooling and citizenship: evidence from compulsory schooling reforms", University of Essex, ISER Working Paper 2007-2.

Spasojevic, J. (2003), "Effect of education on adult health in Sweden: Result from a natural experiment", PhD dissertation, City University of New York, New York.

Smith, J. (2003), "Guide to the Construction and Methodology of the Index of Economic Well-being", mimeo, Centre for the Study of Living Standards, Ottawa, Ontario, Canada.

Sunstein, C. R. (1991), "Preferences and Politics", *Philosophy and Public Affairs*, 20(1).

Stevenson, B. and J. Wolfers (2008), "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox", *Brookings Papers on Economic Activity*, Spring.

Stone, A. A., .A. B. Krueger, A. Steptoe and J. Harter (2008), "Exploring the Socio-Economic Gradient in Daily Colds and Flu, Headaches, and Pain", mimeo.

Thorn, W. (2009), "International Adult Literacy and Basic Skills Surveys in the OECD Area", OECD, Education Working Paper No. 22, OECD, Paris.

Touya, D. M. (2006), "Can we teach civic attitudes?", *Estudios Sobre la Economía Española*, mimeo.

UNDP (2009), "A Brief Update on the Human Development Index", background report prepared for the Commission on the Measurement of Economic Performance and Social Progress, Paris.

UNDP (2007), *Governance Indicators: A Users' Guide*, second edition, United Nations Development Program, New York.

UNDP (2006), *Human Development Report, Beyond Scarcity: Power Poverty and the Global Water Crisis*, United Nations, Development Programme, New York.

Van Djik, J., J. Van Kesteren and P. Smit Paul (2008), *Criminal Victimisation in International Perspective - Key Findings from the 2004-2005 International Crime Victims Survey and European Survey on Crime and Safety*, WODC Publication no. 257, January.

Verhaak, P.F.M., Hoeymans, N. and Westert, G.P. (2005). "Mental health in the Dutch population and in general practice: 1987-2001", *British Journal of General Practice*, vol. 55.

Wauterickx, N. and P. Bracke (2005), "Unipolar depression in the Belgian population - Trends and sex differences in an eight-wave sample", *Social Psychiatry and Psychiatric Epidemiology*, Volume 40, Issue 9.

Weissman, M.M. *et al.* (1992), "The Changing Rate of Major Depression - Cross-national Comparisons", *Journal of the American Medical Association*, volume 268, Issue 21.

Welsch, H. (2006), "Environment and Happiness: Valuation of Air Pollution Using Life Satisfaction Data", *Ecological Economics* 58, 801-813.

WHO (2008), T*he Global Burden of Disease: 2004 update*, World Health Organization, Geneva.

WHO (2000), *Women's Mental Health: An Evidence-Based Review*, World Health Organization, Geneva.

WHO (2002), *World Report on Violence and Health*, World Health Organization, Geneva.

Winkelmann, L., and R. Winkelmann (1998), "Why are the unemployed so unhappy? Evidence from panel data", *Economica*, 65.

Wolfers, J. (2003), "Is business cycle volatility costly? Evidence from surveys of subjective wellbeing", *International Finance*, 6(1).

Wolfson, M. C. (1999), "Measuring health – visions and practicalities", *Statistical Journal of the United Nations*, ECE, IOS Press.

Young, I. M. (1993), "Justice and Communicative Democracy", in Gottlieb, R. (ed.), *Radical Philosophy: Tradition, Counter-Tradition, Politics*, Temple University Press, Philadelphia.

APPENDICES

Appendix 2.1. Some major sources of indicators on political voice and democratic governance

Source	*Nature of indicators*	*Country coverage*	*Period*	*Data basis*	*Potential bias*
Freedom House	Annual survey of freedom	192	1955- ongoing	Based on answers of "experts" to a set of questions. Indicators fall under 6 heads: Political voice (electoral process, pluralism, participation); Government functioning (freedom from corruption, transparency); Media independence; Civil society freedoms of demonstration & public discussion; Rule of law (independently chosen judiciary; protection from political terror; equal treatment of all); Personal autonomy (right to employment, education, gender equality, etc.)	Bias insofar as the experts may be biased or represent partial understanding. Citizens not asked.
World Bank Institute	Governance Matters VII	213	1996-2007	Collates 276 variables for perceptions of governance drawn from 31 data sources constructed by 25 organizations. Gives indicators under 6 heads: Voice & accountability, Political stability & absence of violence, Government effectiveness, Regulatory Quality, Rule of Law, Control of Corruption (see also appendix 1)	Data quality only as good as the organization from which obtained: e.g. information from Freedom House above would have the bias noted. The WBI recognizes that margin of error in assessment can be high.
George Mason Univ & Univ. of Maryland	Polity IV country reports	161	1800-2004 annual update	Based on the assessment of literature by academics	
Carleton University	Summaries all the above indicators	192		Compiles indicators for Democratic participation, Accountability, Human Rights, Political Stability & Violence, Rule of Law (see also appendix 2)	
Reporters without Borders	Press freedom Index	168	2002,annual update	Questionnaire survey in each country	Assumes that state owned media limits freedom
Danish Center for Human Rights	Human Rights Indicators	150+	2006, start date unclear	UN data, Amnesty International	Might underestimate non-state violations of h. rights
UN Inter-Parliamentary Union	Women in Parliaments	189	1997, updates	Compiles information from national parliaments	
University of Michigan	World Values survey	65	1981; 2001-06	Questionnaire survey, individuals sampled	Change in questionnaire over time, affecting comparisons

Appendix 2.2. Questions on social capital and civic engagement in the US Current Population Survey, September and November modules

- Not counting family members, about how many close friends do you currently have, if any? These are people you feel at ease with, can talk to about private matters, or call on for hel[97]p.
- I am going to read a list of things some people have done to express their views. Please tell me whether or not you have done any of the following in the last 12 months, that is between xxxx and now:
  - (Have you in the last 12 months) Contacted or visited a public official -- at any level of government -- to express your opinion?
  - (Have you in the last 12 months) Served as an officer or served on a committee of any local club or organization?
  - (Have you in the last 12 months) Taken part in a march, rally, protest or demonstration?
  - (Have you in the last 12 months) Bought or boycotted a certain product or service because of the social or political values of the company that provides it?
  - (Have you in the last 12 months) Showed support for a particular political candidate or party by distributing campaign materials, fundraising, making a donation or in some other way?
  - Attended a meeting where political issues are discussed?
- The next questions are about the groups or organizations in which people sometimes participate. I will read a list of types of groups and organizations. Please tell me whether or not you participated in any of these groups during the last 12 months:
  - A school group, neighborhood, or community association such as PTA or neighborhood watch groups?
  - A service or civic organization such as American Legion or Lions Club?
  - A sports or recreation organization such as a soccer club or tennis club?
  - A church, synagogue, mosque or other religious institutions or organizations, not counting your attendance at religious services.
  - Any other type of organization that I have not mentioned?
  - In the last 12 months have you attended a meeting of any group or organization?

97. Advisory committee proposed adding the following probes of bridging social capital: “Of these close friends you just described, how many of them, if any: a. Are White; b. Are Latino or Hispanic; c. Are Asian; d. Are African American or Black; e. Have a college degree; f. Didn’t complete high school”.

- In the last 12 months have you attended any public meetings in which there was discussion of community affairs?
- In the last 12 months have you worked with other people from your neighborhood to fix a problem or improve a condition in your community or elsewhere?
- During a typical month in the past year, when communicating with family or friends, how often were politics discussed – basically every day, a few times a week, a few times a month, once a month, or not at all?
- During a typical month in the past year, how often did you and your neighbors do favors for each other? By favors we mean such things as watching each other's children, helping with shopping, house sitting, lending garden or house tools, and other small acts of kindness-- basically every day, a few times a week, a few times a month, once a month, or not at all?
- During a typical month in the past year, how often did you talk with any of your neighbors— basically every day, a few times a week, a few times a month, once a month, or not at all?
- During a typical month in the past year, how often, if at all, did you communicate with friends or family by Email or on the Internet - basically every day, a few times a week, a few times a month, once a month, or not at all?
- During a typical month in the past year, how often did you eat dinner with any of the other members of your household - basically every day, a few times a week, a few times a month, once a month, or not at all?
- I am going to read some ways that people get news and information. Please tell me how often you did each of the following during a typical month in the past year:

  – Read a newspaper in print or on the Internet - basically every day, a few times a week, a few times a month, once a month, or not at all?

  – Read news magazines such as Newsweek or Time, in print or on the Internet - basically every day, a few times a week, a few times a month, once a month, or not at all?

  – Watch the news on television or get news from television internet sites - basically every day, a few times a week, a few times a month, once a month, or not at all?

  – Listen to the news on the radio or get news from radio internet sites - basically every day, a few times a week, a few times a month, once a month, or not at all?

  – Obtain news from any other Internet sources that we have not previously asked about such as blogs, chat rooms, or independent news services - basically every day, a few times a week, a few times a month, once a month, or not at all?

- In any election, some people are not able to vote because they are sick or busy or have some other reason, and others do not want to vote. On November 4, 2008, there were general national elections for political offices such as senators, congressman, and governors. Did you vote in the election held on Tuesday, November 4, 2008?
- Were you registered to vote in the November 4, 2008 election? <Several probes about reasons for non-voting are also included.>

- This month, we are interested in volunteer activities, that is activities for which people *are not paid*, except perhaps expenses. We only want you to include volunteer activities that (you/NAME) did *through or for an organization*, even if (you/he/she) only did them once in a while.
- In the last 12 months have you done any volunteer activities through or for an organization?
- Sometimes people don't think of activities they do infrequently or activities they do for children's schools or youth organizations as volunteer activities. In the last 12 months, have you done any of these types of volunteer activities?
- How many different organizations have you volunteered through or for in the last year, that is, since September 1, 2007? What organization is it/are they? < Follow-up probes are then asked about the type of activity and the amount of time for each organization.>
- Now I'd like to ask a question about donations to charitable and religious organizations. Charitable organizations focus on areas such as poverty and disaster relief, health care and medical research, education, arts, and the environment. During the past 12 months, did you donate money, assets, or property with a combined value of more than $25 to charitable or religious organizations?[98]
- People nowadays have a lot to keep up with and no one can keep up with everything. I have two questions about federal laws. If you don't happen to know the answers, just tell me and we'll move on.

  – What individual or group of individuals has the responsibility to make the final decision on whether a law is constitutional or not - is it the President of the United States, the Supreme Court, or the Congress?

  – Do you know how much of a majority is required for the U.S. Senate and House to override a presidential veto: Is it 80 percent, 67 percent, 60 percent or 51 percent?

98. The Advisory committee had proposed, "Some people contribute money for a wide variety of causes while others don't. During the past 12 months, did you or your household happen to give any money to any charitable or religious cause? (IF YES) a. How much money, if any, did you and your family give to all non-religious charities, organizations, or causes in the past 12 months? (IF NECESSARY: $0, Less than $100, $100 to less than $500, $500 to less than $1000, $1000 to less than $5000, $5000) (IF YES) b. How much money, if any, did you and your family give to all religious causes, including your local religious congregation in the past 12 months? (IF NECESSARY: $0, Less than $100, $100 to less than $500, $500 to less than $1000, $1000 to less than $5000, $5000)".

Appendix 2.3. Additional questions on social connections not included in CPS

1 We'd like to ask you a question about how you view other people. Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

2. If you lost a wallet or a purse that contained two hundred dollars, and it was found by a neighbor, how likely is it to be returned with the money in it? Would you say very likely, somewhat likely, somewhat unlikely or not at all likely?

3. How often do you attend religious services? (If necessary: Several times a week; every week; nearly every week; 2-3 times a month; about once a month; several times a year; about once or twice a year; less than once a year; never.)

4. In any election, some people are not able to vote for various reasons. Thinking about elections for LOCAL government officials since November 2006, have you voted in none of them, some of them, most of them, or all of them?

5. If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?

6. Generally speaking, would you say that you can trust people you work with a lot, some, only a little, or not at all?

CHAPTER 3 - SUSTAINABLE DEVELOPMENT AND ENVIRONMENT

Introduction

The mandate of subgroup 3 was to examine how environmental and sustainability issues can be better incorporated in some revisited indexes of economic performance and social progress. The issue is clearly separated from the issue of measuring current well-being whose monetary and non-monetary aspects were the topic of subgroups 1 and 2.

In short, the global question we had to look at was to know whether some well-chosen statistics can tell us whether our children or grandchildren will face opportunities at least equivalent to the ones that we have had. In doing so, we have to take account of the large body of literature that has been already devoted to this topic. The problem in this field is not the lack of ideas. The problem is rather to understand why it seems so difficult to propose some federative indexes allowing a shared perception of whether our economies are sustainable or not. We shall see that there are actually some deep reasons for such difficulties. We shall nevertheless try to examine which routes could be followed to accommodate and/or progressively overcome these difficulties.

More precisely, the organization of this technical paper will be the following.

Section 2 will be devoted to the general review of the state of the art. We shall discuss five main classes of approaches. Section 2.1 will consider Dashboards, i.e. extensive sets of indicators gathering information on several aspects of development and its expected durability. In their current versions, these dashboards are extremely rich, and this richness is their limit. Their complexity does not allow them to compete efficiently with single headline indicators such as GDP. Even if the principle of dashboards is to be retained in the end, it is a sound objective to try to keep them as parsimonious as possible. This invites to explore all proposals that have been made to push this parsimony at the extreme, i.e. all proposals that pretend to characterize sustainability with one single number.

One first class of such indicators are composite indices reviewed in section 2.2.. The problem of this second class is that these indicators are generally built without any firm analytical framework. Subsection 2.3 will then move on to attempts at integrating environmental considerations in standard National Accounts. This will consist in shortly presenting the System of Eco-Environmental Accounting (SEEA), and in particular the concept of green GDP or green NNP. We shall argue that green GDP, contrarily to some expectations, does not exactly fill the objective of measuring sustainability. It measures at best what could be produced without endangering environmental sustainability: it does not tell us how far we are below or above such a production level, and it is mute on other aspects of sustainability.

This leaves us with the last two classes of indicators that seem to be the most serious candidates for a global assessment of sustainability, because they clearly adopt this

perspective of measuring overconsumption of resources albeit in very different ways. Subsection 2.4 will be devoted to the concept of genuine savings or adjusted net savings (ANS), popularized in particular by the World Bank. This approach consists in assessing sustainability in terms of preservation of a very comprehensive concept of wealth, including physical, human and environmental capital. The other indicator is the ecological footprint (EF), that measures how far we live above or below the environmental carrying capacity of our territories or of the whole planet. A variant of this indicator, the Carbon Footprint focuses on those damages to the environment that matter more specifically for global warming. These two indicators will be presented in section 2.5.

From this general survey proposed in section 2, it will indeed appear that no indicator emerges as consensual, even among those that try to rely on a well-defined concept of global sustainability. Such a situation is obviously a source of perplexity for the statistician. Why are there such different views about sustainability? Is there some scope for a larger approach that would subsume all these different views in a comprehensive way?

The aim of section 3 of the paper will be to shed more light on this central question. This will be done by restarting from fundamentals. The basic question is to build an indicator that would warn us, sufficiently in advance, whether we are or not on a non-sustainable growth path. Some theoretical answers exists to this question that can be viewed as a further generalization of the concept of adjusted net savings reviewed in section 2.4. Such a generalization can accommodate many settings: in particular, it is devised to accommodate the cases of efficient as well as inefficient functionnings of markets or institutions –and this property is of course crucial in the environmental domain-, or situations where possibilities to substitute man-made assets to natural ones are strongly limited.

But this is no way sufficient to conclude that we have at hand the solution to our problem of sustainability measurement. Quite the contrary. This theoretical perspective will rather be used to show what are the major obstacles that we face in practical implementations. The main message from this theoretical presentation will be that, as soon as current prices loose their informative power, the quantification of sustainability cannot proceed without explicit predictions of future economic and environmental trajectories, and without some explicit normative choices about values to be attached to such or such kind of trajectories, i.e. what is to be sustained, and for whom. There can be very different views about all of these points.

All these difficulties are particularly important for the environmental component of sustainability. This will argue in favor of relatively eclectic final recommendations, mixing the ambition of pursuing efforts for building comprehensive indexes, more focussed on the *economic* side of sustainability, but accompanied by a set of specific "physical" indexes focussing more specifically on *environmental* dimensions of sustainability. These tentative recommendations will be exposed in section 4.

Four appendixes have been added to this technical report. The first one just gives the exhaustive contents of one of the large scale dashboards reviewed in section 2.1, the European Dashbord on Sustainable Development. The three other develop some particular aspects of the ANS or extended wealth approaches to sustainability. Appendix 2 discusses the measurement of one of the components of this ANS, change in human capital. Appendix 3 presents some variants of the ANS for France that emphasize the difficulty to get a version of this index that sends strong messages concerning the issue of global warming. Appendix 4 tries to elaborate further upon this difficulty, looking at connexions with debates that have

followed the publication of the Stern Report: it shows that the difficulty to get strong messages on climate change based on ANS indicators cannot be an argument in favor of inaction. This is one more argument in favor of monitoring this specific issue separately from global sustainability assessment provided by the either the ANS or some of its extended versions.

2. Taking stock

2.1. Dashboards or sets of indicators

Dashboards or sets of indicators are a widespread approach to the general question of sustainable or durable development. They consist in gathering and ordering series of indicators that bear direct or indirect relationship to socio-economic progress and its durability. First examples of this approach go back to the 60s and the 70s. At that time, the concept of sustainability did not have the importance that it has gained since and these early dashboards were essentially focusing on the measurement of social progress (the so-called «social indicators » movement) without any reference to environmental issues. Interest for such dashboards then declined somewhat, but reincreased strongly since the 1990s. Notably, the Rio Summit in 1992 framed the definition of sustainable development as resting on the three pillars of economic efficiency, social equity and environmental sustainability and led to the adoption of Agenda 21 whose 40th chapter invites signatory countries to develop quantitative information on their actions and realizations along these three pillars. As a consequence, the second generation of dashboards that are now available generally devote a large room to sustainability and environmental issues (see Bovar et al. (2008).

To be a little more explicit about the typical contents of such dashboards, we have reproduced in Appendix table 1 one example of such an extensive dashboard, the European Dashboard of sustainability indicators. This dashboard covers 10 different themes, with 11 level 1 indicators, 33 indicators for level 2 and 78 indicators at level 3, these level 2 and 3 indicators covering 29 sub-themes. The most striking point for an external user is the extreme variety of these indicators. Some are very global ones -GDP growth itself keeps its place, and is even the first indicator on the list -, some more specific such as the percentage of smokers in the population. Some pertain to outcomes, some other 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? Probably 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 upon what we consider as “high” or “low” in terms of fertility?

The main criticism to these dashboards is therefore their heterogeneity. One can also be worried about the lack or harmonization of these dashboards, the frequent changes that occur in their compositions, or the fact that they deliver too much information to be efficient communication tools, even when their main messages are summed up in a limited set of headline indicators. More fundamentally, these dashboards may appear as lacking a clear

definition of what is exactly required for sustainability, and even a clear definition of what sustainability consists in. The most striking feature in this respect is the apparent confusion that is conveyed by these dashboards between indicators of *current* levels or trends in well-being, and on the effective sustainability of these levels and/or trends.

In defense of these dashboard, one may first recall that their hybrid character is in fact consubstantial to the initial Brundtland program of simultaneously caring about development *and* its sustainability. Development can be rapid but non sustainable in the long run. Symmetrically, sustainability can be warranted but with very low levels of development. The originality of DD strategies is precisely to orient policy making in directions combining both aspects, i.e. seeking the highest level of current development that is compatible with long run sustainability. In this context, it must come as no surprise to have a cohabitation of indicators applying to both domains, even if this is at the expense of readability.

Box 1: Weak versus strong sustainability

Weak and strong sustainability are two concurrent concepts that are very frequently used to classify empirical approaches to durable development (e.g. Dietz and Neumayer, 2004).

The term of *weak sustainability* has been coined to characterize economic approaches to sustainability that emerged during the 1970s. These approaches were extensions of standard neo-classical growth theories. Standard growth models generally consider that output is only determined by technology and the available quantities of two production factors, labor and capital. The main innovation of this literature has been to introduce natural resources as an additional production factor in these models, and to specify the laws for the evolution of this natural factor, for instance a modeling of extraction behavior in the case of an exhaustible mineral resource - this literature developed after the first oil shock.

These models generally assumed large substitution possibilities between natural resources, capital and labor. Combined with exogenous technical progress, this offered one solution to the finiteness of resources, at least from a theoretical point of view: as oil resources decline, production is expected to use less and less of them but without any decline in standard of living, either thanks to pure technological progress, or by replacing oil by some alternative fossil energy or any other man-made production factor.

Promoters of *strong sustainability* rather consider that substitution possibilities necessarily face physical limits. Critical levels must be maintained for most of natural resources. These critical levels must be at least equal to those necessary for basic-life support functions, and more probably higher if we want to keep reasonable levels of environmental *resilience*, i.e. the capacity of eco-systems to regenerate and return to equilibrium after shocks. The concept of strong sustainability is often considered as irreducible to monetary approaches. All environmental variables of interest have to be followed in physical terms.

We shall follow the tradition to separate these two concepts in most of this text even if, as will be seen in section 3, the distinction is in some sense less clearcut than usually said. This is allowed by the fact that the economist's toolbox also includes representations of production processes where substitution possibilities are limited *a priori.* It is also possible to combine the economic approach with models of eco-system dynamics to try to put monetary values on features such as resilience or irreversibilities, at least at the theoretical level. The term of inclusive wealth theories is sometimes used for these models that push as far as possible this integration of the economic and environmental viewpoints. Practical applications of such eco-environmental models remain generally limited to specific eco-systems (see for instance Mäler, Aniyar and Jansson, 2008 for a series of illustrations) but, to follow the terms of Weber, "if inclusive wealth could be computed at the global level, it would constitute a normative indicator combining weak durability (service flows) and strong durability (resilience of ecosystems)" (Weber, 2008, our translation).

Second, their development has spurred many efforts to improve the accuracy and international comparability of existing indicators and it has also spurred many efforts for producing new ones. This movement is an important stimulus to statistical production that, *in fine*, should have positive consequences for all other possible approaches to DD measurement.

Third, whatever the implied cost, one must admit that complexity is often unavoidable. It is especially the case if we adhere to a principle of "strong" versus "weak" sustainability (see box 1), according to which preserving the well-being of future generations requires simultaneously maintaining critical levels for many environmental assets, rather than a global combination of such assets: this unavoidably implies a parallel follow-up of all these assets. Even if we accept weak sustainability, its management generally relies on the combination of several instruments, each of which needs to be followed separately. And there is no reason why the detailed lists of all these objectives and instruments should be exactly the same in every place and at any period of time. The requisites of sustainability are inevitably area and period specific. Made-to-measure dashboards are the right instruments for capturing this complexity. Aiming at extreme simplification and at a forced harmonization would mean abandoning this comparative advantage of the method.

In short, dashboards are instruments that are useful and cannot be bypassed. At one moment or another, the follow-up of a strategy necessary implies the delivery of quantitative information both on the instruments of this policy and on how far we are from the realization of its targets. Dashboards are inescapable when we come to that phase. Even at a more global level, the final conclusion of this report will be that we cannot avoid some form of multidimensional view of sust***ai***n***a***bility. But parsimony remains an objective. If a dashboard of sustainability needs to be built, we should try to keep it as small, focussed and structured as possible. This requires a good analytical framework of what sustainability means. To this aim, one can gain in exploring the several efforts that have been devoted to building more comprehensive mono-dimensional sustainability indicators.

2.2. Composite indexes.

Composite indexes are one way to circumvent the problem raised by the extreme richness of dashboards and to synthesize their abundant and purportedly relevant information into one single number. Here also initiatives have been numerous (see Afsa et al. (2008) or Gadrey and Jany-Catrice (2007) for more detailed surveys), with a larger role played by academics and non governmental organizations, while dashboards are usually built with a strong involvement of official statistical institutes.

The general idea of those indexes is to re-scale elementary components in order to ensure comparability and to aggregate them, with possibly unequal weights, to produce one figure. At the simplest level, there exist for instance indexes that try to 'green' the Human Development Index by combining it with informations on emissions of pollutants (Desai, 1994 or Lasso de la Vega and Urrutia, 2001). These indexes are among those reviewed for France by Nourry (2007), with the conclusion that results are difficult to interpret.

A much more elaborated and relatively well-known index is Osberg and Sharpe's Index of Economic Well-Being (Osberg and Sharpe, 2002). This IEWB covers simultaneously 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. Among the countries considered, Norway has the highest level of economic well-being, followed by Italy, Germany, Sweden and France. The four Anglo-Saxon countries trail back, with Canada in eighth place and the United States in last place. Jany-Catrice and Kampelmann (2007) revised the Index of Economic Well-Being (IEWB) for France with enhanced data over a larger time period. Their results confirmed the divergence between GDP and IEWB since the end of the 1980's, but largely due to the lack of improvement in reducing inequalities and improving economic insecurity. The "green" dimension of this index remains secondary at this stage.

More strictly focussed on environmental issues, researchers from Yale and Columbia applied the methodology of composite indicators for building both an "Environmental Sustainability Index" (ESI) and an "Environmental Performance Index" (EPI) (Estes et al., 2005). The ESI covers 5 domains: environmental systems (Air, Land, Water, Biodiversity), environmental stresses reductions (air pollution, waste pressure, natural resource management), 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. SO_2 emissions per habitant or phosphorus concentration), health related indicators (e.g. infant death rate from respiratory diseases) or environmental governance (Local Agenda 21 initiatives per million people). These 76 indicators are aggregated in 21 intermediate indicators ultimately aggregated for producing the global ESI, after equal weighting. The EPI is a reduced form of the 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. According to such an index, Finland ranked highest in 2005 (global index of 75). The global ranking of countries has some 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 CO_2 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.

Globally, all these composite indicators should probably be better regarded 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 *raisons d'être*. A similar comment can be made concerning attempts to synthesize other dashboards with statistical procedures more elaborated that simple weighting, such as principal components analysis (Jollands et al. 2003, David, 2008). These approaches are interesting ways to summarize a great deal of information. Once we have the global view, we can return to detailed elements: a country that is badly ranked can look at the variables that are predominantly responsible for

its situation and try to improve its score along these variables. Such an incentive to policy change is not to be neglected at all.

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 for dissatisfaction is twofold. The first is the same as for large dashboards, i.e. the lack of a 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 economic indexes, because they are not linked to any form of market valuation. Indeed, 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 merits of this literature. The problem is rather than their normative implications are seldom made explicit or justified.

2.3. Adjusted GDPs: greening the national accounts

Subgroup 1 has examined for which reasons GDP or even total final consumption can be only very partial indicators of welfare. In their seminal paper (1973), Nordhaus and Tobin followed this line of criticism and proposed to build 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).

Additionally, based upon their MEW, they built a *Sustainable* Measure of Economic Welfare (SMEW) taking into account changes in total wealth. 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. The initial MEW was further enhanced into an Indicator of Sustainability of Economic Welfare (ISEW) by Daly and Cobb (1989) and refined by Cobb and Cobb (1994) in order to incorporate natural resources, as did also the Genuine Progress Indicator (GPI)[1]. Those indicators deduce the costs of water, of air and of noise pollutions from consumption and accounts mainly for losses of wetlands, farmland, primary forests, and for CO_2 damages and ozone depletion too.

1. The Genuine Progress Indicator (GPI) is very similar to the ISEW. It has been promoted since 1995 by the non-governmental organization “redefining progress” (see for instance Talberth, Cobb and Slattery, 2006).

In what sense do these adjustments towards a better measure of welfare provide us with something related to "sustainability"?

Indeed, since Samuelson (1961) and Weitzman (1976), economic theory has argued that a correctly adjusted Net National Product should equal the maximum level sustainable –i.e. non declining- level of consumption that can be reached for today and the future. This is the "hicksian" view of income, that defines income as what can be consumed this year without being poorer at the end of the year, i.e. ending the year with the consumption perspectives equivalent to current year consumption. The SMEW and its successors follow this track. Closer to our concern here, especially in environmental matters, Hamilton (1996) proposed several theoretical models, featuring depletion of renewable or non-renewable resources, pollution or environmental amenities and derived the way final consumption should be adjusted for to provide a sound measure of welfare and thus a level of sustainable consumption. Under this theoretical framework, empirical investigations to compute environmentally-adjusted net domestic product (eaNDP) taking into account the consumption of natural capital, often known as "Green GDP", have flourished since the late 80's and the first System of Environmental Economic Accounting (SEEA) established in 1993 (see early empirical studies by Repetto et al. (1989) or Alfsen et al. (2006) for a survey, and SEEA Handbook Chapter 11 (2003) for additional references).

Nevertheless, these accounting adjustments remain rather controversial[2].

One first problem is that methods of valuation needed for those adjustments are generally indirect and often 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 from many accountants.

More precisely, experience from the field suggests two main options for valuing degradation. The first relies on damage-based estimates and the second 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 these degradations on human health and thus on human capital.

Cost-based estimates instead answer the question, 'how much would it cost to avoid environmental degradation?', and in turn can be divided into two types. The first type relies on *maintenance costs*, that is, the value of the costs that 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 (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 production technology).

2. On these difficulties and some proposals, see in particular Vanoli (1995).

The second type of cost-based estimate attempts to overcome those limitations and answers 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. Interest then focuses less on the new 'greened' aggregates themselves than on the gap between the existing economy and the 'greened' version (and on the possible transition paths between the two).

This need to rely on some joint physical and economic modelling at some stage or another will systematically reappear later on, whatever the approach preferred.

But there is a more fundamental problem with green GDP, which also applies to Nordhaus and Tobin's SMEW and to the ISEW/GNI indexes. None of these measures characterize sustainability *per se*. Green GDP just charges GDP for the depletion of or damage to environmental resources. 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 we need are measures of **overconsumption** or, to put in dual terms, of **underinvestment**. This is precisely what our last two indicators purports to do.

2.4. Adjusted net savings (ANS) or changes in extended wealth

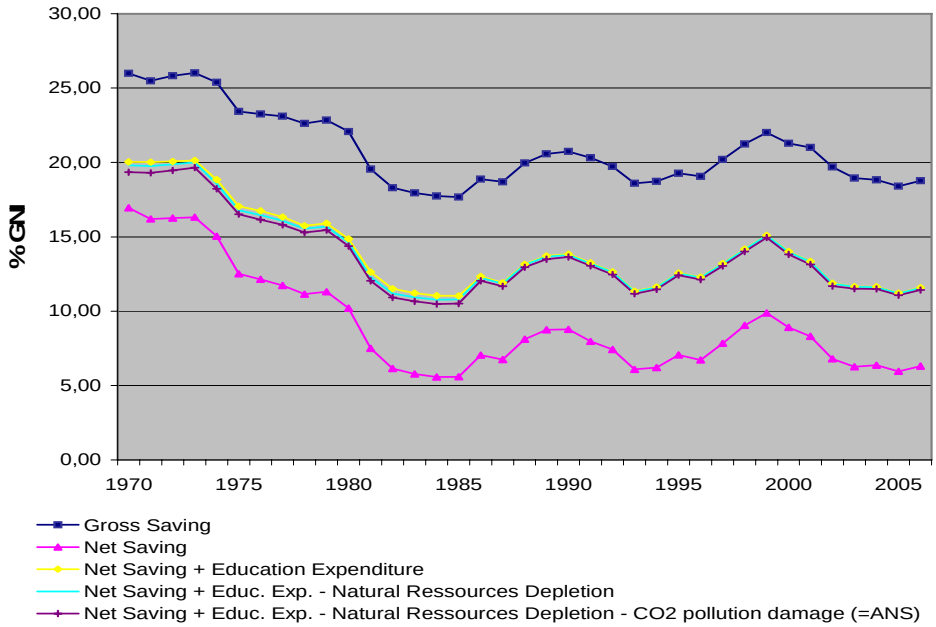
Adjusted net savings (also known as genuine savings) is a sustainability indicator building on the concepts of green national accounts and whose theoretical background is again Hicksian interpretation of income and wealth, but the focus is now on savings, i.e. the variation of wealth. It is this variation that, if negative, will send the signal that current consumption levels will not be possibly maintained over the future. Alternatively, following Arrow *et al*. (2004), one may prefer to label this indicator as "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". The analogy with the case of individual wealth is straightforward: if I have disinvested or dissaved this year to finance my consumption, this implies that I am poorer at the end of the year. Eventually, I will have the possibility to dissave again next year to maintain this level of overconsumption. But I know that I will not be able to do so indefinitely: one day or later, I will have to adjust my consumption downwards. 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 that may provide future generations an opportunity set that is at least as large as that being faced by currently living generations.

Equipped with these definitions, researchers at the World Bank (see, in particular, Hamilton, Pearce and Atkinson, 1996; Hamilton and Clemens, 1999) took the lead in computing "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. 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 deducted[3]. Negative adjusted net saving rates imply that total wealth is in decline, and as such provide a warning of unsustainability.

How does this indicator compares to standard measures of saving and investment in national accounts. As exemplified on Figures 2.1 and 2.2 for France and the USA, the level of ANS is mainly driven by gross saving adjusted for capital consumption and education expenditure, and appears to be very marginally affected by other corrections (even if adjustments for natural resources depletion were not that marginal in the USA during the 80's).

Figure 2.1: From GNS to ANS. What really counts (France)



3. 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).

Figure 2.2: From GNS to ANS. What really counts (USA)

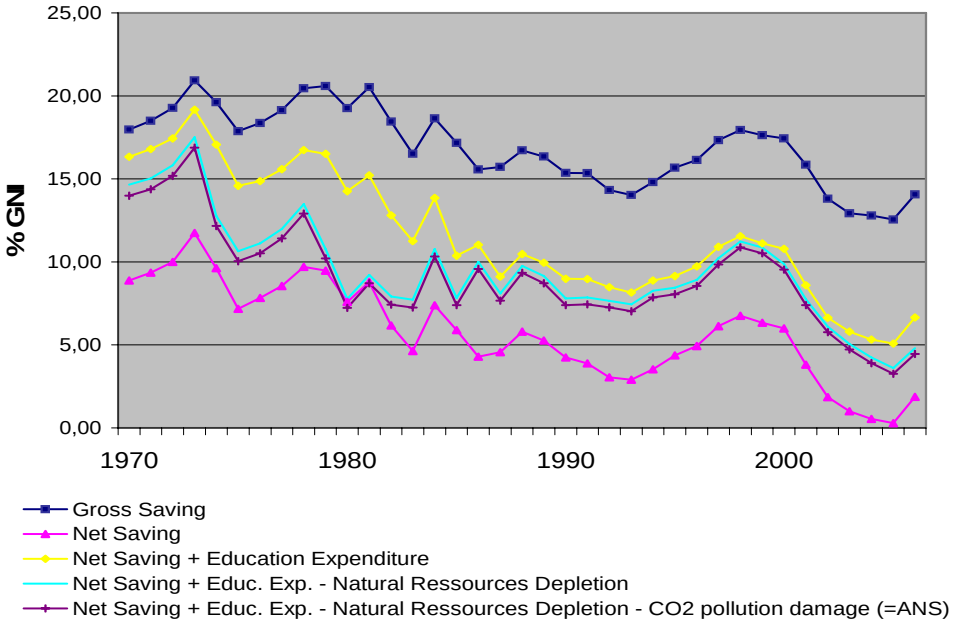
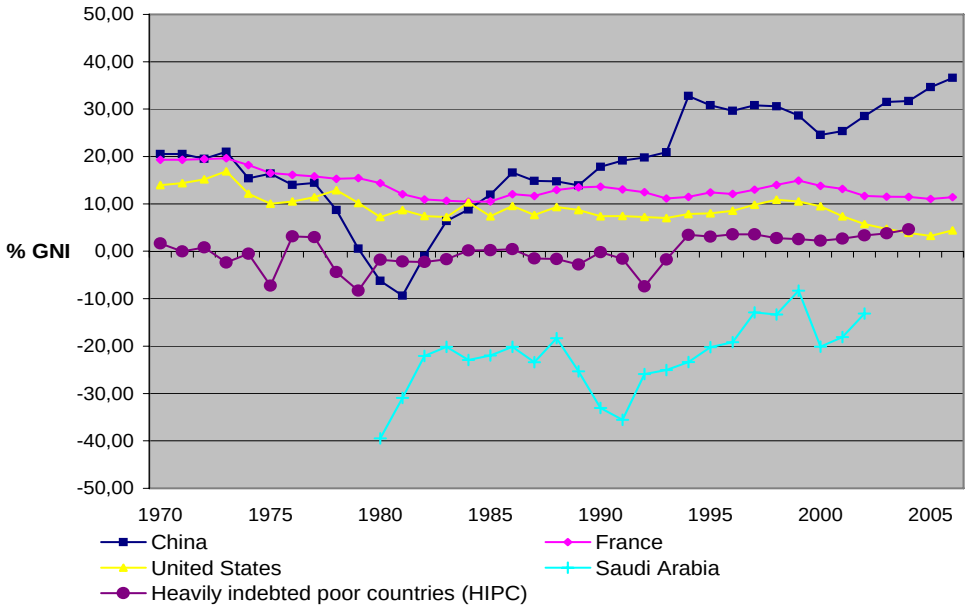


Figure 2.3: ANS for selected countries, 1970-2006



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 (figure 2.3).

Those numbers have the advantage of benefiting from a consistent conceptual framework defining sustainability. The seminal and regularly updated work by the World Bank to collect data for those estimations of ANS is a significant contribution towards true wealth accounting.

Nevertheless, current methodology underlying empirical calculations of the ANS per country has well-known shortcomings, articulated by the authors themselves[4]. Some of them could be easily handled out. For instance, it is rather straightforward to provide further adjustments taking into account technical progress and population dynamics (considering adjusted net savings per capita). But, more generally, the relevance of the ANS approach crucially depends on **what** is counted (the different forms of capital passed through to future generations), namely what is included in the "extended wealth", and on **the price** used to count and aggregate (what is counted) in a context of imperfect valuation by markets.

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" (Where is the wealth of Nations, 2006, pp. 154), and *a fortiori* biodiversity loss. As for pricing environmental degradations, in the absence of any market valuation, accounting prices must be evaluated by modeling long term consequences of given changes in environmental capital and how they impact prospective well-being. We have some established pieces of theory that tell us how this can be done but practical implementation raises considerable problems. Appendix 3 provides in-depth analysis of such shortcomings and feasible improvements. These implementation problems will be at the center of the arguments developed in section 3.

2.5. Footprints

The *Ecological Footprint* was initially proposed by Wackernagel and Rees (1995). Since then, it has been sponsored by Redefining Progress and the WWF, which provides regularly updates of the index in its annual *Living Planet Report*, and has gained increasing popularity among green NGOs and public opinion. Methodology and updates are currently maintained by the Global Footprint Network.

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

4. Many other authors have tried to improve the original indicator: Arrow et al. (2004) for major world areas, Hanley et al. (1999) and Pezzey et al. (2006) for Scotland, Lange and Wright (2004) for Bostwana, Arrow et al. (2008) for the US and China, Nourry (2008) for France and Randall (2008) for Australia….

biological resources and services useful to humanity", as Moran, Wackernagel and co-authors (2008) phrased it.

Results are well-known and rather striking (figure 2.4): 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.

Undoubtedly, the phrasing, the concepts and the results are very appealing and seem very easy to understand. Those reasons explain the success of this indicator, among citizens and international influential bodies. But these pedagogical qualities do not mean that the indicator is without weaknesses, and, just as the ANS, but for very different reasons, it has been criticized on many of its aspects[5].

Some very general criticisms are the fact that the index gives no weight to possibilities offered by technical progress, or the fact that it is mute on unsustainability problems that result from the extraction of fossil resources, or on items such as biodiversity or water quality.

Even on what it is supposed to do well, the indicator raises some problems. To look at them a bit more closely, let's consider the 6 different types of land use separately.

For cropland, the Ecological Footprint accounts do not rely on the definition of a "sustainable" agricultural yield, a yield leaving soils with sufficient quality to offer the same yield next year for instance. On the contrary, the biocapacity is indeed the observed capacity obtained from real yield. As a consequence, at the country level, the footprint of production will always equal the biocapacity of the country for cropland. This has two consequences. First, at the country level, ecological deficit for cropland will only mirror trade deficit. Second, at the worldwide level, exports and imports clear up and thus Ecological Footprint and Biocapacity for cropland will always be equal.

Those points are not concealed by the authors[6] but are not really put forward when interpreting and analysing the results. Van den Bergh and Verbruggen (1999) had already pointed out what they call the substantial anti-trade bias inherent of the Ecological Footprint methodology[7]. 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 most recent update of the methodology (*Calculation methodology v1.0*, 2008)

5. A recent and more extensive review of the index has been provided by the French CESE (Le Clezio, 2009) and by the French Ministry of Ecology (CGDD, 2009)

6. See *Calculation Methodology for the National Footprint Accounts, 2008 Edition, version 1.0.*

7. One should notice that, unlike adjusted net savings, EF is based on final consumption (or final demand), that is production + imports - exports. To put it differently, the required equivalent land area for a given consumption is attributed to the resource *consumer* rather than the *producer*. For instance, non-renewables extracted in a developing country and exported to a developed country (e.g oil) count towards the developed country's EF.

acknowledged that a country's ecological credit is not in itself a sufficient criterion for sustainability. Moran, Wackernagel et al. (2008) moved 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, they acknowledge that EFs are not intended as measures of a country's own sustainability, but more of its contribution to global unsustainability. Yet the fact that no ecological deficit can exist for cropland at the worldwide level remains. At this worldwide level, this indicator is not informative at all on the excess pressure that men, on the average, exert on the regenerative capacity of this cropland. The same reasoning applies to grazing land. Once again, at the country level, ecological deficit will only mirror the trade deficit on livestocks products and ecological deficit at the worldwide level is impossible.

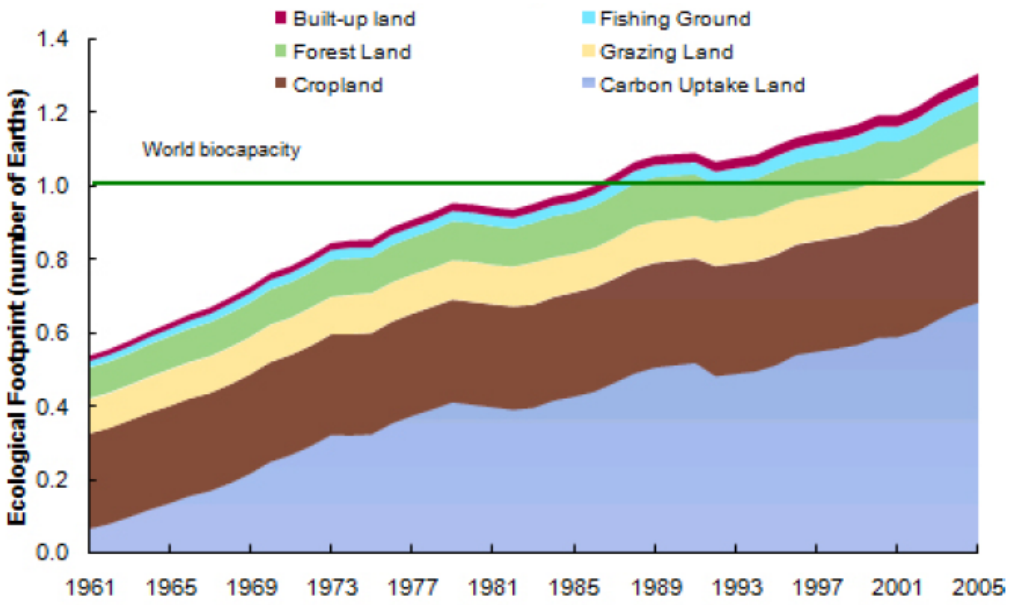
Regarding built-up lands, demands placed on this land use type always equal the biocapacity, as they both represent the cropland area lost for providing those services. Hence, the contribution of that type of land to the overall ecological deficit is null.

Concerning demands placed on fishing grounds and forest lands, the EF provides a treatment that seems more completely in line with a full sustainability perspective, that is the flow of consumption available leaving the stock unchanged for next year. Nevertheless, at the worldwide level, one observes that biocapacity exceeds footprint, suggesting that there is no sustainability problems regarding the use of those resources. Moreover, shall they turn into a deficit, one must keep in mind that they represent quite a small part of the overall EF (respectively 9% and 3% for forest land and fishing grounds).

Last, but not least, carbon uptake land is the amount of forest land required to uptake carbon dioxide anthropogenic emissions. In 2005, it is the largest contributor (more than 50 %) to humanity's current total Ecological Footprint and increased more than tenfold from 1961 to 2005. In the standard presentation of the Footprint, there is no formal biocapacity in front of the demand for waste (CO_2) assimilation. Note that an alternative view could be to aggregate the footprint (area of forest) needed to supply wood for construction, fuel and paper and the footprint needed to absorb CO_2 emissions, and to compare this aggregate with the biocapacity of forest, as forest is indeed the counterpart used to monitor carbon dioxide absorption and each use (wood products and CO_2 absorption) puts a competitive demand on the same unit of forest. In doing so, we would be able to display a huge global deficit in forest land, mirroring the overaccumulation of CO_2 emissions in the atmosphere.

Till now, we have reviewed potential unbalances for each kind of land separately. To aggregate those different areas and come up with one single figure, both biocapacity and footprints originally measured in hectares are expressed in terms of a standardized land area unit, termed global hectare. It is a way to re-weight land areas according to their relative potential *agricultural* productivity. This potential is assessed through the spatial GAEZ model of potential agricultural yields framed by IIASA and FAO. As an example, consider one hectare of cropland and one hectare of forest. The GAEZ model would tell you that the potential yield of the cropland area is twice the potential yield for forest land. Then, while aggregating cropland and forest land to compute overall footprint, cropland areas will be weighted twice forest areas. This is not neutral for policy recommendations as forest will generally have lower equivalence factor (weights), and as such, a heavy substitution from forest to cultured lands will directly increase the overall biocapacity available and will alleviate consequently the ecological deficit.

Figure 2.4: Humanity's ecological footprint by component 1961-2005



Source: Ewing et al. (2008)

In short, the apparent simplicity of the EF hides several difficulties and this often argues in favour of better focussing its real contribution and of maybe less extensive but more transparent approaches. Comparisons of EFs between countries should be used, shall it be used, as an indicator of inequality of consumption and interdependencies between geographical areas (e.g. Mac Donald and Patterson, 2004). But the main message conveyed by the Ecological Footprint, and it is indeed a message, is that humanity, and notably western developed countries, has increased dramatically the level of CO_2 emissions in the atmosphere during the last 40 years (figure 2.4). Those annual emissions are far above the level that could be assimilated by nature, with the well-known consequences on temperature and climate. The methodological approach in the Footprint calculations consist in expressing those emissions in terms of the required equivalent surface (forest) to absorb them. It gives the flavour and the sense of this tremendous problem to the man on the street, and it is its most valuable virtue. But its value-added compared to true carbon accounts is not striking otherwise.

Indeed, the Ecological Footprint methodology provides interesting elements for a full carbon account of consumption, as the carbon footprint at the country level takes into account not only direct CO_2 emissions, but also indirect emissions embodied in imported goods. More generally, carbon accountancy (see the "*Bilan Carbone*" methodology in France for instance) is probably a more promising attempt to monitor the demand we (a country, a local community, a citizen…) place on available waste absorptive capacity of our ecosystem.

3. Single index approaches to sustainability: where are the obstacles?

Let's try to sum up the general impression that emerges from the rapid overview proposed in section 2. This section has shown the large number of existing attempts to quantify sustainability. Up to a certain point, such an abundance is normal. Sustainability is not a one-dimensional issue and, at some moment or another, we have to control it through relatively large sets of indicators: this is precisely the case we made in favour of dashboards.

But from another point of view, abundance is often considered as a serious drawback. If we want to draw attention of public opinion or policy makers on sustainability issues, one would also like to have some synthetic headline figures able to compete with the popularity of GDP. Such was precisely the ambition of many of the indicators that we have reviewed in the first section and, in this respect, it becomes a real problem to see all these so-called summary indexes providing such different views on the sustainabilities of different countries.

Just to insist upon how far these discrepancies can go, we have plotted against each other the values of three indicators available in a systematic way for almost all countries in the world (figures 3.1 and 3.2): the Environmental Sustainability Index, the Adjusted Net Savings as computed by the World Bank (in % of GNI) and the Ecological Footprint.

In both cases, the relationships confirm that the links are weak and that messages can go in opposite directions. The ESI and the ANS display a small positive correlation due to the fact that they both have a bias in favor of more developed countries. Concerning the ESI, this is due among other things to the weight it gives to active environmental policies and the quality of institutions that support these policies. Concerning the ANS, this is due to the fact that developed countries are more able to accumulate physical and human capital while the exploitation of exhaustible resources is more often concentrated in southern countries. The correlation is of the same sign for the ANS and the EF, but must now be read in the other direction : countries sustainable from the point of view of the ANS have higher EF's and are therefore less sustainable or, more precisely, contribute more to global unsustainability than countries with low ANS's.

Is it possible to go beyond such inconsistencies and to ultimately agree on one shared representation of global sustainability? If yes, this would provide a major breakthrough. But we shall see that the answer is unfortunately not positive, for reasons that are deep ones and not likely to be solved easily. These reasons need to be properly understood, if we want to orient our efforts in the appropriate directions. This is the main purpose of this section.

Our analysis will have two steps. The first step will answer the following question: if we were asked to produce only one global indicator of sustainability, what is the methodology that one should apply? The crucial element here is the aggregation procedure required for combining all the very heterogeneous elements that matter for future well-being (subsection 3.1). Some recent economic literature proposes one answer to this question, that neither retains the naïve assumptions that market prices are good guides for this aggregation, nor considers that such a relative valuations can be only left to arbitrariness. Simple examples will show how such a framework would perform in stylized settings, including a potential capacity to subsume the traditional opposition between weak and strong approaches to sustainability (subsection 3.2).

But this framework also highlights the very strong requirements of such an ideal index and it is therefore a good point of departure for systematically analyzing all the problems one that

has encounters when trying to quantify sustainability. We do not pretend to explore systematically all of them, but we shall raise some points that have appeared to be particularly crucial. We shall first argue that this framework shows how important are the technical and normative uncertainties that make univocal assessments of sustainability very difficult to provide (subsection 3.3 and 3.4). We will then use it to clarify the difficulties raised by the international dimension of the problem (subsection 3.5). All this will pave the way for the more eclectic recommendations that will be proposed in the concluding section 4.

Figure 3.1: Relationship between the ESI and ANS indexes

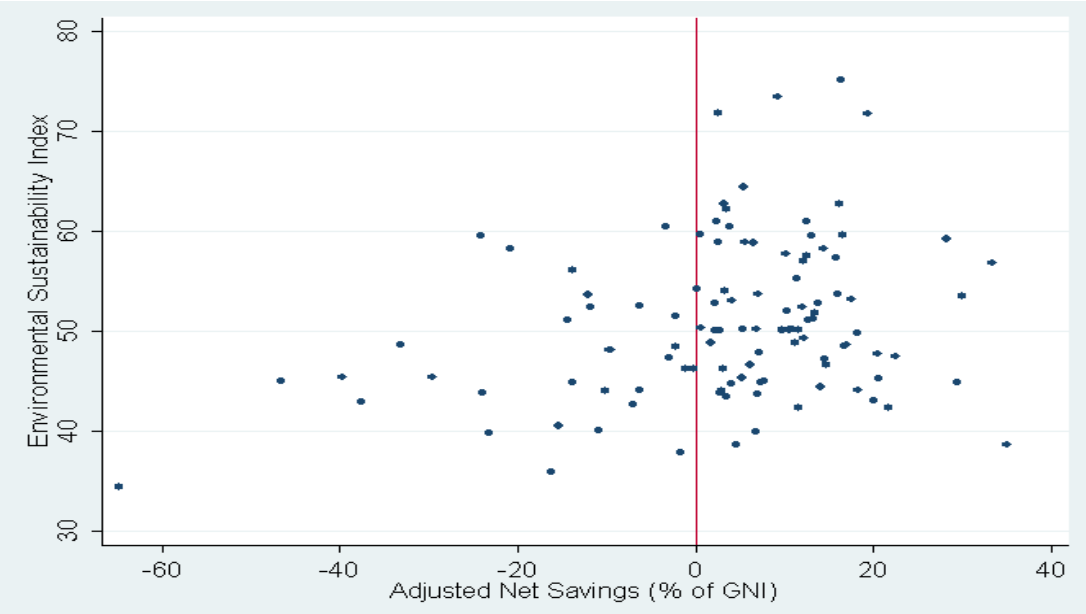
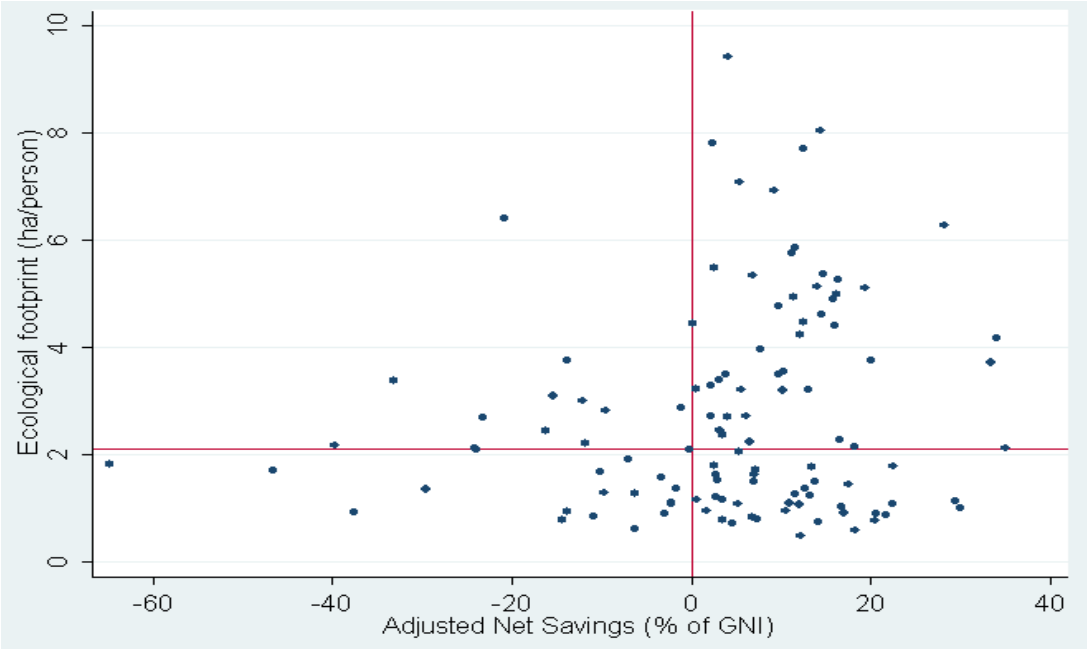


Figure 3.2: Relationship between the ANS and the EF indexes



3.1. If we had to build one single sustainability indicator, how should we try to proceed?

From what has been seen in section 2, at least one consensual point should emerge, which is that the sustainability issue is about what we leave to future generations and whether we leave enough resources of all kinds to provide them with opportunity sets at least as large as the ones we have had for ourselves. This is the basis of what one is generally presented as a "stock", "wealth", "asset" or "capital based" approach to sustainability. The assets to be considered are numerous: fossil resources, other renewables environmental resources, but also physical, human and social capital, or general knowldege. To avoid any misunderstanding, presenting them as "assets" or "capital goods" does not mean at all that we consider that these assets should all be privately owned or submitted to market forces. Many of them are collective assets that cannot be managed efficiently by market mechanisms.

This "wealth" or "stock based" view of sustainability has, for instance, provided the basic paradigm for the Millenium Ecosystem Assessment monitored by the United Nations between 2001 and 2005 and which, at this stage, constitutes the reference inventory for environmental trends all over the planet. Such a conceptual framework is, of course, fully in line with economists' traditional views on dynamic aspects of well-being (see box 2). It is also interesting to observe that reference to this notion of wealth is shared by several heterodox authors (see for France Méda, 1999 or Viveret, 2002).: All this suggests that we hold here a common language that can help convergence between a wide variety of points of view.

Just to take one example, let's recall one of the frequent 'heterodox' criticisms addressed to standard GDP, the fact that, from the point of view of this index, destructions of natural or physical capital can be counted positively, because of the positive impact that repairing can have on activity. To be exact, it is not always the case that this repairing activity will increase GDP: it does so only if it generates a net increase in activity, not if there is simply a reallocation of work between normal activities and these repairing activities. Yet, when an increase in activity actually takes place, it is correct to say that GDP sends an inappropriate message if interpreted in terms of well-being. The stock based approach to sustainability clearly avoids such a paradoxical result, in a way that has close links to the question of defensive expenditures raised at length in the contribution of subgroup 1. In this stock-based approach, a catastrophe is registered as a capital loss, i.e. an impoverishment and an immediate threat on sustainability. Repairing activities will be counted as investments but whose function will only be to compensate for the initial loss. They do not make us wealthier if the initial capital loss has been adequately registered. And since this is investment, this should not enter into the evaluation of current well-being: without the catastrophe, people could have enjoyed more positive activities than just restoring the conditions that prevailed before the catastrophe.

This being said, all the problems and possible divergences arise when one comes to quantification. The basic element of quantification is to measure changes in quantities or qualities of these different assets or resources, what is called "physical measures". Measurement problems are already considerable at this stage, but let's assume they have been overcome. At a given point in time, we observe by how much we accumulate some stocks, in a way that should contribute positively to future well-being. We also observe how we deplete or depreciate others, in a way that contributes negatively to future well-being. The question is how much sustainability or unsustainability this represents on the whole? If we impose to

ourselves the constraint of answering this question with only one number then we must find a key for aggregating all these physical variations.

For assets or resources that are traded in markets, one way to value their contributions to future well-being can be to use prices currently observed on markets. But, already for this simple case, we know that there are many reasons why such an approach is problematic. The first one is that, even when prices exist, there is a risk that they misrepresent the true long term contributions of corresponding assets to well-being because of their incapacity to fully incorporate all the externalities -positive or negative- that can be associated to the accumulation of these assets. Even if such was not the case, there is the additional problem that such prices also reflect myopic or irrational behaviors of investors or resource-holders, often leading to erratic changes around fundamentals. For instance, one just has to think of the huge variations that have been observed for financial assets or oil prices over the recent years.

Box 2: Sustainability, wealth and intertemporal approaches to well-being

Measures of sustainability that we discuss in the text have links with the more global question of measuring social well-being in an intertemporal perspective. Yet We try here to sum up the main ideas in literary terms, the reader being referred to section 3 of M. Fleurbaey's contribution to the commission for a much more rigourous argumentation (Fleurbaey, 2008).

Wealth and intertemporal well-being: two twin concepts

The intertemporal dimension of well-being can be introduced by starting from the individual point of view. The idea is that an individual's current well-being does not only depend upon what she consumes or enjoys right now, it also depends on what she expects to consume or enjoy in the more or less distant future. She can be satisfied with a relatively low level of immediate resources if she expects that it will raise overtime. Conversely, she will remain poorly satisfied with a currently high standard of living if she knows that this will be only a short-lived episode. The natural treatment of this intertemporal dimension is to consider that this individual's *global* well-being *V* is better measured as a weighted combination of her current and future levels of *instantaneous* well-being *U*. In practice, this weighted combination generally takes the form of a discounted sum of *U*'s.

How does this notion relate to wealth? The link is direct. We know that the rigourous definition for the value of an asset is the discounted sum of dividends that it will generate at all future periods. In the same way, the right definition of an individual's wealth is to measure it as the discounted value of future consumptions or enjoyments that she can plan to derive from her current endowments and this is nothing else than the definition of intertemporal well-being. Of course, it must be clear that such a concept of wealth goes far beyond the standard notion of financial of physical wealth. Owning financial or material assets increases for sure a person's possibility set in terms of future consumption. But having a high level of schooling, being in good health or benefiting from a large social network are also some forms of current wealth that enlarge one's perspectives of earning money, consuming and/or enjoying life at later periods.

What if we move to the social point of view? One possibility could be to aggregate these prospective *V*'s for all individuals alive at the current period. But, from a collective point of view, this is only one aspect of the problem. Collectivities are expected to survive beyond the life horizons of their current members, these members being continuously replaced by new ones. This means that intertemporal well-being of a given collectivity rather needs to be measured as a discounted sum of consumptions or enjoyments by all its current or future members at all future dates. If we do so, we will have the same correspondance between the notion of intertemporal social welfare and a concept of current "extended wealth" aggregating the utility-generating potential of all our current endowments, not only for ourselves, but also for all future generations. We shall use the letter *W* to characterize this extended concept of wealth.

The difficulties of such an evaluation are however considerable, as highlighted long ago by Samuelson (1961). According to his own terms, this puts quite a large amount of "futurity" in our

evaluation of social well-being, and makes it quite impossible in practice. Even if we were able to produce very long term projections of future instantaneous consumption or well-being, we would remain confronted to the issue of discounting, i.e. of deciding how to weigh the relative well-beings of people living today and of people leaving 100 years ahead from now or more. This remains the matter of many debates. Discounting is anavoidable from a practical point of view (to avoid infinite sums), but is ethically problematic: in principle, all people should be treated equally, irrespective of their date of birth. Anyway, whatever we do, practical indexes of welfare requiring an intertemporal aggregation until the end of times are both hard to build, and clearly hard to communicate upon.

Two ways of measuring sustainability: sustainable well-being or changes in global wealth

Because of these difficulties, the commission did not choose to emphasize such a long run intertemporal view of social well-being. Subgroups 1 and 2 on standard GDP issues and Quality of Life essentially concentrate on elements that pertain to the current period.

The intertemporal dimension is considered in this subgroup 3 through the narrower but not less important dimension of *sustainability*, i.e. the question of knowing whether the society will be able, at all future dates, to have a level of current well-being as high as the one that we have today. The connection with wealth and intertemporal well-being can be stated as follows. Measuring wealth is measuring the size of the cake to be divided between successive generations, while measuring sustainability is raising a question about how this cake will be shared between generations or periods, the objective being to have no future generation less well-off than the current one. This metaphor of the cake is not fully satisfactory in an intertemporal context. In such a dynamic framework, the size of the cake is not fixed but also depends on the behavior of successive generations. Yet this imperfect metaphor helps understanding the difference and the connection between the two notions of global wealth and sustainability.

Now, there are two traditions in this assessment of sustainability.

The first ones tries to define a concept of sustainable consumption or well-being. The idea is to compute the maximum constant level of well-being that can be ensured at all successive periods with current levels of resources as a starting point. A large amount of literature has been devoted to the characterization of this U_{sust}. This was the initial idea of Nordhaus and Tobin's index of "sustainable economic welfare" and its has been taken over by many successors of this index. This notion has been further conceptualized by several theoretical papers starting from Weitzman (1976). Back to the empirical side, the concept of green GDP can also be viewed as one partial attempt to compute this sustainable well-being.

There are however two limits in this approach. First, concerning green GDP, we must emphasize that it only corresponds to a partial version of this U_{sust} concept. It measures production net of environmental degradations, but does not take into account forms of accumulation or decumulation of many forms of assets, specifically of intangible assets. Second, even if we were able to measure U_{sust} in a satisfactory way, this is by no means a sufficient statistic for characterizing sustainability. To put it simply, fully replacing GDP by a measure of green GDP does not help us assessing whether we are on a sustainable path or not. Measuring sustainability requires *comparing* this green GDP to our actual consumption of resources. Rephrasing it in terms of utilities or well-being, what is needed to assess sustainability is the (U, U_{sust}) couple. We are on a sustainable path as long as $U \leq U_{sust}$. We are in an unsustainable situation as soon as U becomes higher that U_{sust}.

The second tradition consists in building an indicator that directly measures this gap between what we consume or enjoy and what would be sustainably enjoyable. This gap can be captured by the change in extended wealth dW and computations of adjusted net savings are one attempt to quantify this change. The intuition is simple. An increase in wealth between t and t+1 means that we start at period t+1 with resources that should allow maintaining the level of well-being that we had at time t without compromising the future. A decrease in wealth means that we start at time t+1 with a lower endowment that the one we started with at time t. This does not necessarily forbid consuming as much in t+1 than we did in t. But this clearly means that we will not be able to do it indefinitely. One day or later, society will have to adjust its standard of living downwards, and the adjustement will be all the more violent that it will come late.

As it was the case with the computation of W, the message will be sensitive to the choice of the discount factor ρ, but with a somewhat different meaning. When we compute W, the choice of ρ is an ethical problem, i.e. knowing which relative weights we give to well-being today or well-being tomorrow. When we compute dW and use it as an indicator of sustainability, the choice of ρ rather affects the

anticipatory power of the index. Choosing a value of ρ is like the choice of a focal distance for binoculars. A high value for ρ will allow focussing on risks of a downturn in U in a relatively near future. But, with such values, the indicator will be silent on downturns of U that are expected for a very distant future. Warning messages on long term unsustainability will only appear if a low value is adopted for ρ.

This connection between wealth and sustainability also means that it can make sense to look at them simultaneously. This is precisely what did the World Bank in its most recent evaluations of Adjusted Net Savings (World Bank, 2006). Not unsurprisingly, these evaluations show that most of the "wealth of nations" lies in intangible resources such as skills, human capital, in conformity with an intuition already put forward by Adam Smith more than two centuries ago. Yet this predominance of human capital is essentially true for developed countries, a much larger share of the wealth of poor countries consisting in their natural assets. And, unfortunately, Adjusted Net Savings for these countries often come out as negative, because the exploitation of these natural resources is not compensated by sufficient accumulation of other physical or human assets. This does not necessarily dismiss the responsibility of developed countries in such a state of affairs. Neither does it mean that these less-developed countries should consume less. The message can rather be that they have to invest more efficiently. All this just emphasize the exposition of these poorer resource-dependent countries to medium term problems of unsustainability, and the associated risk of an increasing worldwide inequality.

Anyway, even if we were ready to believe in the validity of market prices when they exist, there remains the problem of assets that are not traded as such on any market, for which no direct prices are observable. Excluding them from computations is the kind of answer we have to avoid. We have to find alternative procedures.

In some cases, indirect monetisation still seems possible in a relatively straightforward manner. Let's take the example of human capital, presented with more detail in appendix 3. For this form of capital, the strategy used by initial versions of the ANS was to value its accumulation by the monetary cost of producing new human capital, i.e. education expenditures. This solution is obviously too crude: it ignores depreciation of this human capital and the fact that the same level of expenditures can produce human capital of very different qualities.

An alternative way is to use the partial information brought in by the labour market: human capital is valued by the discounted flow of labor incomes that it is expected to generate. This assumes that we can extrapolate in the future the current rates of return on this human capital. Of course, this is once again a disputable hypothesis. We cannot warrant that returns on education will be the same tomorrow as they are today. But, on such an issue, we have the feeling that some reasonable valuation methods are not fully out of hand and that they can help us comparing intensities of net human capital accumulation across countries or over time.

The problem appears much more complicated for environmental assets, for which the information provided by markets becomes very limited or even non existent.

In some cases, a frequent solution is to use estimates of costs of avoiding the degradation of these assets, for instance the cost of installations that would fully avoid the emission of such or such pollutant in the atmosphere, or the prices than one would have to fix to maintain levels of emission or environmental degradation below some given targets, as it is for instance done for the valuation of CO_2 emissions. Such a solution remains unfortunately unsatisfactory for what we want to do. It can fit for indexes of the green GDP style. Green GDP measures

what we would be able to produce without deterioration of the environment: in such a perspective, it is an appropriate strategy to subtract from standard GDP or NNP the potential cost that we would have to incur in order to maintain the environment in its current state. But we have seen that green GDP is not exactly the right answer to the specific issue of sustainability.

What sustainability indices have to measure is something else: we want them to tell us whether damages that are currently made to the environment are going to drive the well-being of future generations below the one that we enjoy today. It is only under strong conditions that there can be equivalence between this long run impact and the current marginal cost of pollution abatement. This equivalence occurs only if we currently optimize between the present costs of these abatements and the benefits that will result from them in the future. It is excessively restrictive to assume that such is the case in the real world.

In fact, what a true measure of sustainability requires is a direct evaluation of what these future damages are going to be, and how they are going to affect future well-being. And, to be exhaustive, the same kind of imputation is in fact required for all other assets: even when they have some apparent market values, their true valuations should be based on the net expected amounts of services that there are expected to deliver tomorrow, in interaction with other assets, and we know that the absence of markets for only a subset of goods leads to biased on prices for all goods, not only the ones that cannot be traded. For instance, if capital accumulation is, everything else equal, good news for the well-being of tomorrow, but if it simultaneously puts a strongly negative pressure on then environment, then this external effect should be reflected negatively in the present value of these capital goods and it is quite likely that current market prices will not do that.

3.2. How would the indicator behave in some typical cases?

How should all these problems be treated? The informational requirements are considerable. As soon as market prices are not a reference, we need to rely on imputed or shadow prices and these imputations require no less than a full projection model for the economy, the environment and their interactions, and a perfect anticipation of how these changes are going to affect tomorrow's well-being (Arrow, Dasgupta and Mäler, 2003).

Let's try to get an idea of what this means in a very stylized example. This report is not the place for a thorough exploration of these analytical properties of sustainability indexes, but a few illustrations are nevertheless interesting to examine[8]. The illustrative setting that we shall use assumes that there are only two "assets": produced capital that can consist either in physical or human capital or a mix of them, and a natural resource. Produced capital plays the usual role it has in economic growth models: it determines output, and it is accumulated by reinvesting part of this output, the other part consisting in consumption which is one of the components of well-being. Produced capital is also subject to depreciation. This specification implies that, in terms of economic production, we ultimately converge towards a stationary

8. Technical details are not provided but are available upon request. As a general rule, it is an interesting way of testing indexes to try to examine their behavior in theoretical settings, in parallel to their implementation on real data. If they fail to deliver interpretable messages in such theoretical settings, it is a sufficient condition for having them discarded. If they pass the test but only in some cases, it helps us identifying for which cases they are useful or misleading.

state for both its capital stock and its output, which is determined by the investment rate and the depreciation rate. This occurs after a period of continuous growth, if the economy starts from a capital stock that is lower than its equilibrium value, or eventually after a period of continuous "degrowth", if the economy adopts an investment rate that is not sufficient to maintain its initial capital stock.

The environmental dimension is then introduced through the existence of a renewable resource that is essential to well-being and that stops regenerating sufficiently once a certain production threshold is reached, through the kind of mechanism that footprint indexes aim at capturing. Passed this point, the natural asset starts depreciating in an irreversible way, and, under our assumptions, this ultimately drives well-being to zero, even if economic production goes on increasing.

It is in such framework that we have tested what would be the behavior of a generalized savings rate indice built according to the principles outlined above. The idea is that, at each period, we observe net variations for produced capital and for environmental capital and that we aggregate them according to their relative contributions to the future sequence of current well-beings, discounted at a certain rate. We shall say that we have a negative sustainability gap when the indice is below zero: it warns us that well-being is going to decline sooner or later below its current level.

Figure 3.3 starts by a situation where this environmental constraint is never binding but where unsustainability occurs because of unsufficient renewal of produced capital. This is what we can call a case of economic unsustainability: it comes as no surprise that the index properly points at a negative sustainability gap all over the simulated period. In that case, the concept is close to a standard net savings rate, and it appropriately tells us that savings are not sufficient to maintain current well-being. This society lives above its means. Indexes focussing on the environmental component would send no message concerning this form of unsustainability. This gives at least one good reason for giving some place for such an index in any approach to sustainability.

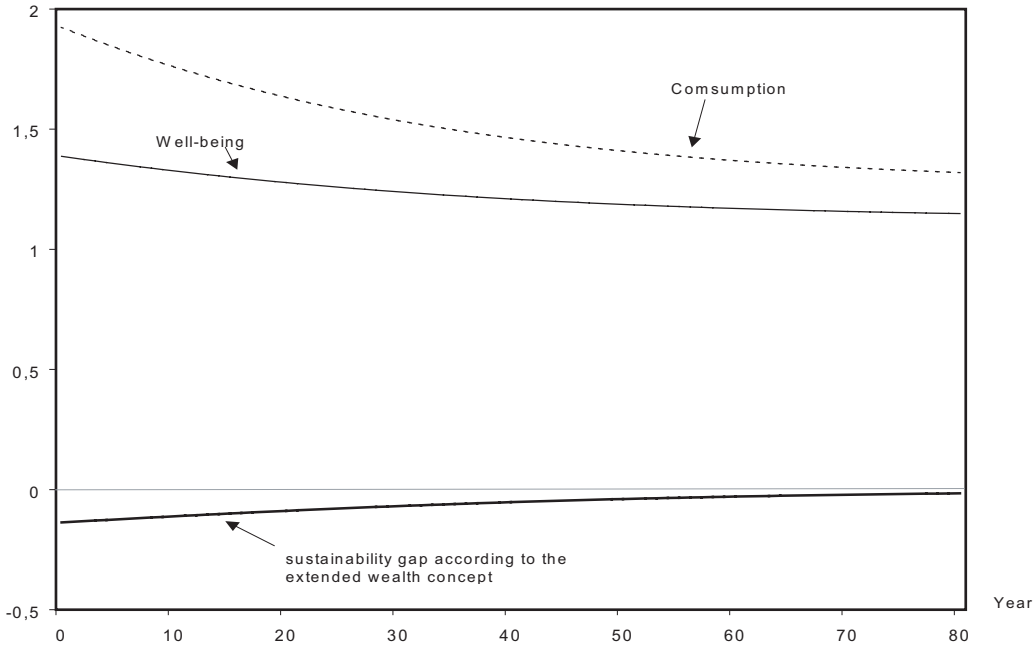
What if unsustainability rather comes from the environmental side? This is the situation represented on figure 3.4 where capital accumulates and consumption are growing up to the point where the stock of the environnemental good starts declining, leading to a downturn in global well-being. The indicator built according to the theoretical prescriptions appears here again able to anticipate this situation. If it is built with a high discount rate (5% in our example), it will do it in a relatively myopic way: a negative sustainability gap appears only a few years before well- being starts to decline. But it does send the message however, and it does it in a much more forward-looking manner when built with a low discount rate of 1%[9].

Now, where does this capacity of anticipation come from? It simply comes from the assumption that the index builder is in position to base his assessment on the relevant imputed prices that are displayed on figure 3.5, and which appears to have two properties. The first one is a strong upward trend in the imputed value of the environmental good that reflects the fact that it becomes increasingly determinant for the evolution of well-being. But this is not

9. On this influence of the discount rate of the indicator on the messages sent by the index, see the technical paper prepared for the commission by M. Fleurbaey (Fleurbaey, 2009). The role of the discount rate in this context is somewhat different from the role it has played in controversies surrounding the Stern report: more on this is presented further in appendix 4.

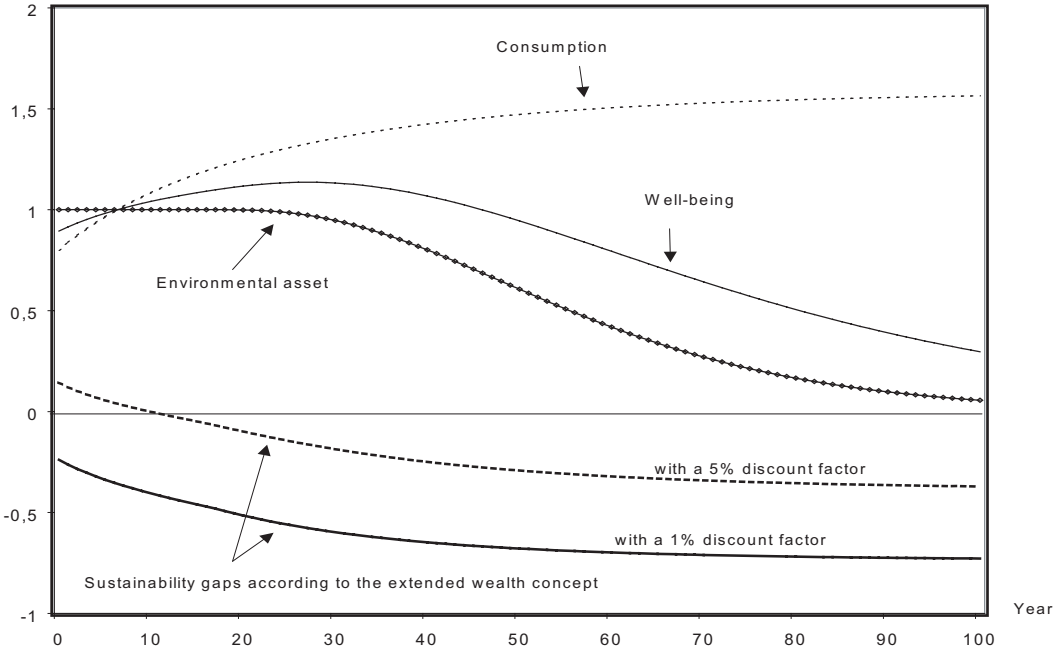
enough. The second requirement, in this particular case, is the fact of immediately attributing a negative value to the accumulation of physical capital, in anticipation of the fact that its continuation will ultimately lead us above the critical environmental threshold.

Figure 3.3: Sustainability gap in a stylized case of economic unsustainability



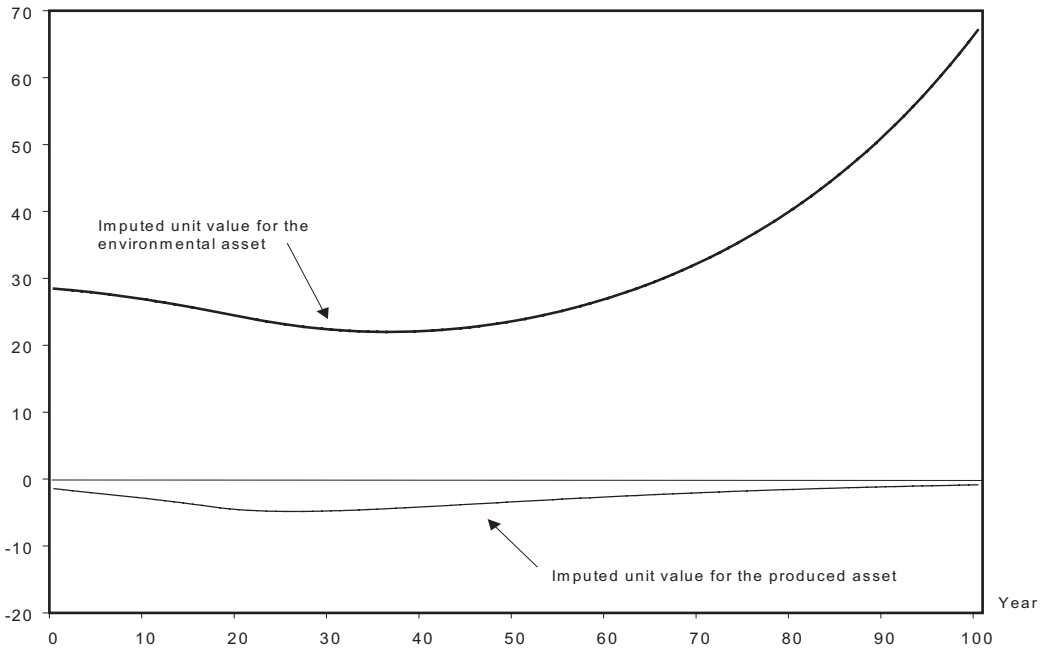
Reading: In a scenario where consumption and well-being continuously decline due to an sufficient physical investement, the extended wealth indicator permanently sends the signal that current consumption is not sustainable.

Figure 3.4: Sustainability gap in a stylized case of environmental unsustainability



***Reading*:** If excessive consumption leads to environmental degradations that ultimately reduces well-being, the extended wealth index warns that we are on a non sustainable path, with a capacity of anticipation that depends on the discount rate.

Figure 3.5: Imputed values for the environmental and produced asssets for the scenario of environmental unsustainability (for a 1% discount rate)



Reading: The results of figure 3.4 rely upon specific properties of accounting prices: a strongly increasing positive value for natural capital, and a negative one for produced capital.

This example has no pretention to realism, although some may consider that it actually depicts one of our possible futures, e.g. one of the scenarios of "collapse" analyzed with lots of details by Diamond (2008) or of bifurcation scenarios described by climatologists concening long term consequences of climatic change. Its interest is before all pedagogical. It shows what are the requisites of a proper functioning of this accounting approach. It is as able as other approaches to capture situations of "strong" unsustainability. But this is true only under requirements that are themselves quite strong:

– It requires a full modelling of the future joint development for the economy and the environnement, and their impact on the sequence of future levels of well-being. This means that building the index is in fact no less that a full projection exercise. In fact, the index itself does not bring any additional information to this projection exercise. It is only one way among others to sum up with one single number the whole results of this projection exercise[10].

– It eventually requires being ready to accept strong messages from these models in terms of valuation of the different assets. These may eventually differ in a huge way from signals that can be sent by markets.

All this provides a good framework for clarifying the difficulties of the exercise. From such a starting point, it becomes relatively easy to unfold the various reasons that make

10. The same is true for "sustainability" indexes used in public finance, such as the "tax gap" or indexes of implicit debt for social security administrations: all these indexes are only some ways to sum-up long-run projections of expected surpluses or deficits.

sustainability assessment difficult, and specifically when we pretend to do so in monodimensional terms.

3.3. One first difficulty: behavioral and technical uncertainties.

One first obvious difficulty with sustainability measurement is that, done proprely, it is no less than a full projection exercise. Hence, it should come as no surprise to be confronted to the problem that encounters any prospectivist, the fact that the future is uncertain.

In the most extreme form, the argument can be that the future will depend on what we shall make of it. Indeed, making a projection also implies projecting behaviors, including the behavior of policy makers. There are as many assessments of sustainability as there are possible paths for future policies, and shadow prices are actually one way to evaluate and compare these policies (Dreze and Stern, 1990; DasGupta, 2003).

Let's assume however that we refrain from going up to this point. What we are primarily required to do is to assess sustainability in "business as usual" scenarios, i.e. scenarios where current behaviors or policies continue to hold forever. However, even in such cases, sources of uncertainties remain considerable:

– Even if we have the right conceptual model for describing future eco-environmental interactions, there is uncertainty concerning the parametrization of this model, or concerning the magnitude of future external shocks that could cause deviations from the baseline path predicted by this model.

– There is uncertainty about the model itself: are we really sure that we are using the right one? Aren't we neglecting such or such aspect of the environment that may prove of crucial importance tomorrow? This was the case some decades ago concerning climate change or biodiversity. What are the next environmental variables that may come in the forefront in the decades to come?

– Last but no least, there is uncertainty about one fundamental component of the index which is the choice of well-being indicator that we use for assessing future states of the nature and of the economy. In theory, this strongly intertwines conclusions of this group with conclusions by the two other ones. There as many evaluations of the sustainability of well-being as there can be definitions of what current well-being is.

Without running into details, let's elaborate a little bit on these issues and possible answers to them, starting with the first two ones.

For components of technological uncertainty that are amenable to probabilistic formalization, one possibility could be think in terms of confidence interval. We know that no indicator can tell us for sure that we are sustainable or not: it can say we are while we are not, it can say we are not while we are. One could consider assorting indicators of measures of these two opposite risks. Alternatively, one could consider submitting the indicator to stress tests, or present alternative evaluations in worst cases scenarios, in accordance with the precaution principle. These ideas can be a topic of future research, but they already appear a bit difficult to apply in practice.

The problem is more crucial concerning more radical uncertainties concerning the proper modelling of interactions between the environmental and economic spheres. They leave room for more radically divergent applications of the general framework of section 3.2. The choice

of one index against another will reflect divergences in a priori beliefs concerning the model that best describe reality, or also different sensibilities to the risk of using a mistaken model, with some preferring models that are highly conservative on the environmental side, while other will accept the risk of underestimating these environmental risks. 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 probability distributions of future eco-environmental scenarios. There is no reason why sustainability assessment could have escaped such difficulties.

Such a problem brings us well beyond the problems usually faced by statisticians whose ordinary job is essentially that of measuring the *current* state of the world. As far as this current state of is concerned, there can be measurement problems and/or divergences over how to aggregate the different characteristics of this state of the world in summary indexes. These problems are already large, but there is in principle no room for heterogeneity of beliefs or expectations. Trying to quantify sustainability adds this dimension of assessing the future, and heterogeneity of beliefs concerning this future come into play as an additional source of complexity.

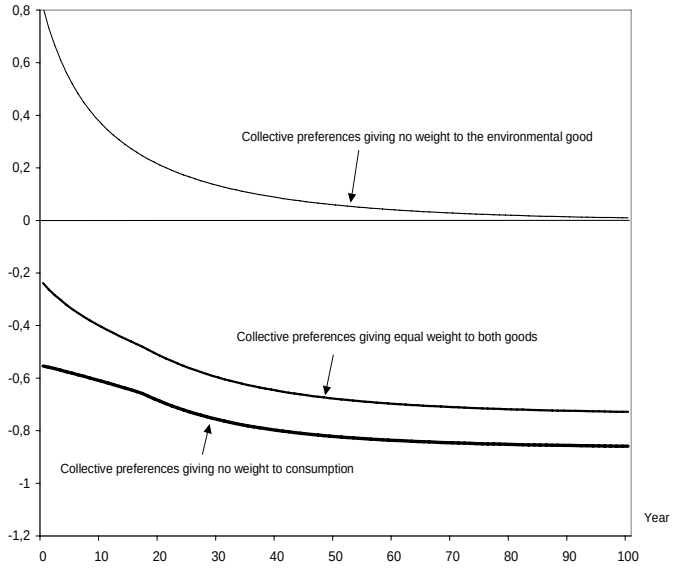
One answer to this difficulty could be to provided indexes built from several competing models. But the pedagogical character of such exercises is not certain, and this encourages to take the problem in another way. If there are some natural items whose interaction with the economic sphere can be large with a significant probability but in a way that is difficult to model, there is a strong case in favor of separate follow up of these environmental items, leaving aside the ambition of incorporating them into one single indicator.

3.4. Normative uncertainties: sustainability of what?

On the normative side also, one can say that there are as many indices of sustainability as there are definitions of what we want to sustain. This remark may seem trivial, but it is paradoxically not that frequent in the literature. It deserves a few comments. 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. if apples are cheaper than oranges, it reflects among other things something about peoples relatives tastes for apples and oranges. No normative choice is therefore to be made by the statistician. This would still hold true for sustainability measurement if all the assets we have to consider were traded on perfect markets by individuals fully aware of the importance of these assets not only for their own future well-being but also for the future well-being of their descendants.

As soon as we consider that market prices cannot be trusted, we lose this indirect expression of revealed preferences. Direct specifications have to be injected in our valuation instruments, and results will be naturally very sensitive to these specifications. Just to illustrate that, we have rerun the simulation of figure 3.4. for alternative specifications of well-being: the original specification of this figure 3.4 that gave equal weight to the environnemtal quality and consumption, and the well-being functions of the pure environmentalist and of the pure consumerist that give symetric maximal weights to the environment and to consumption, and no weight at all to the other component. The result goes without saying but nevertheless deserved being recalled: these two polar assumptions lead to entirely different evaluations of sustainability.

Figure 3.6: Sustainability gap for the same scenario as figure 3.2 but different specifications of well-being



***Reading*:** Messages from the sustainability indicator given on figure 3.4 are strongly modified if we adopt other representations of collective preferences concerning consumption and the quality of the environnement.

Are there some solutions to this indeterminacy? One could try to solve this problem empirically, trying to infer the definition of the adequate well-being function from current observations of how people value environmental factors compared to economic ones. Since this cannot be done from observed prices, we must rely on other means, such as contingent valuations, or direct measures of the impact of environmental amenities on indexes of subjective well-being such as those considered in subgroup 2. But the limits remain numerous. In particular, 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? The good properties of the index depend on the capacity of the well-being function to capture the relative valuation of environmental and non-environmental goods over the full range of variation of their relative quantities.

Such a global profile may prove quite hard to infer simply from current observations in a narrow interval of variation of eco-environmental variables. Some may for instance 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. On the other hand, some anti-environmentalists may argue on the opposite: that these future generations may be completely indifferent to the disappearance of some environmental amenities that we currently value only because we are used to them. This brings into the debate the additional complexity that can stem from changes in well-being over time, changes that may be themselves path dependent.

Pushing further, there is also the problem of knowing how the chosen indicator of well-being must aggregate individual preferences, i.e. the distributive dimensions of well-being. If

for instance we assume that a headline indicator of current well-being must be the total disposable income of the bottom x% of the population, rather than global disposable income, then indicators of sustainability must clearly 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 can imply some specific investment in institutions that help protecting efficiently this population from poverty. In principle, our theoretical framework tells us how we could ideally put some value on such "institutional" investment. Arrow et al. (2003) actually mention institutions as some of the assets that should be ideally included in a really comprehensive measure of wealth. But needless to say that the prospect of being able to do so is still more remote than for other assets.

In short, all this question of properly predefining well-being (a) emphasizes the necessary link between conclusions by this subgroup and conclusions from the two other subgroups and (b), may be again argument in favor of multiple sustainability indices, corresponding to different definitions of what we are trying to sustain. In a sense, this shows that there is no intrinsic opposition between the "extended wealth" approach to sustainability and the idea often put forward by partisans of composite indexes that the weighting of the different ingredients of sustainability should be a matter of democratic debate. The advantages of the stock-based approach is that it provides us with a framework that helps identifying what are those items that we have to weight and that its makes clearer on what basis these weights should be computed, compared to arbitrary weighting procedures that are often adopted in composite indexes. But, as soon as we agree that market prices of assets cannot be the reference for the relative valuation, we are brought back to the problem of knowing on which bases such valuations can be established.

3.5. The cross-national problem: sustainability of whom?

Let's move on to the properties of indicators in a multinational context. We have seen that this aspect of the problem is another source of major cleavage between the ANS and many other approaches to sustainability. According to the index chosen, it is less developed or more developed countries that are the most severely concerned by unsustainability, the first ones because of underinvestment in physical and natural capital and/or because of a poor management of their natural resources, the second one because of their high standard of living that puts a lot of pressure on the world's ecosystem or natural assets.

One possible defense of the ANS view on this issue is to argue 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 net savings, 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. This is the so called "Hartwick rule" (Hartwick, 1977),

according to which sustainability for a resource exporting country requires a full reinvestment of all its rent. A country that sells a non renewable asset necessarily gets poorer if it does not convert all the resulting income in another asset.

In reality, this Hartwick rule needs some to be refined. If we take into account the fact that the price of an exhaustible resources must be on an ascending trend (the Hotelling rule phrased in Hotelling (1931)), then the value of a given stock of this resource is expected to increase autonomously over time, and this allows a country to be sustainable even if it does not reinvest the entirety of the income currently derived from this resource. But, once this correction is made, ANS computations are correct again.

Yet they are so only if an additional assumption is valid, 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. This problem will be all the more crucial when there are no markets at all, or in the presence of strong externalities.

Can an extended wealth approach handle this difficulty? The answer is not straightforward. Let's for instance assume a very simple two-country setting with a natural resource that is now a global public good with free access. We assume that the two countries produce and consume at each period but with different technologies. 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 the natural resource. Last, we push further the asymmetry by assuming that it is only country 2 that is affected by this degredation 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. With such a setting it is natural to redefine countries 1 and 2 as being respectively "the polluter" and "the polluted".

How can the mechanics of imputed prices work in such a context? One possibility consists in computing country-specific accounting prices for the natural resource and apply these two different prices to the global degradation of the natural asset when computing changes in extended wealth for each of the two countries. The difference between the two prices will reflect the fact that the two countries suffer differently from environmental changes. If we do so, it is easy to guess that the accounting price for the polluter will be zero: this polluter is not impacted 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 to this asset. If there is a degradation of this natural asset, this will imply characterizing the polluter as sustainable, and the polluted as unsustainable.

From a certain point of view, this result reflects some reality. It is true that it is the polluted's well-being that is going to decline. But from another point of view, the message to policy makers is misleading. Country 2 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. From this second point of view, what matters is not each country's own sustainability, but each country's contribution to global unsustainability. This can eventually be done in an extended wealth framework if we are able to compute a global extended wealth and impute changes of this global wealth to each of the countries taken separately. But this was more spontaneously the way footprint indicators work. This calls again for a

diversification of approaches, using for each kind of asset the kind of instrument that seems the most relevant.

4. Conclusion: main messages and tentative recommendations

To sum up, what have we learnt, and what can we conclude? This trip through the world of sustainability indicators has been a bit lengthy a technical. The issue is indeed complex, more complex than the already complicated issue of measuring current well-being or performance. For this reason, the conclusions of this subgroup will remain somewhat more open than those of the other two subgroups. But we shall nevertheless try to articulate a limited set of three messages followed by four recommendations, which we shall also try to keep as pragmatic as possible.

Message 1: Measuring sustainability differs from standard statistical practice in a fundamental way: to do it adequately, we need projections, not only observations.

The usual business of statisticians is to try to measure what happens or what has happened in the more or less remote past. When it comes to sustainability, the question turns to producing numbers about the future, which is not yet observed by nature. Admittedly, some could argue that, in a world of perfect capital markets, all the relevant information on this future path of the economy is conveyed in the current valuation of assets or of the services that they currently provide. If one good is going to be scarce in the future, this should be already reflected in its current price. Such is the implicit view of some existing implementations of the ANS index. But this is clearly a view of the mind. Recent events have shown how far well-established capital markets can be mistaken in their implicit predictions of future economic developments. This is all the more true in domains where markets are notoriously underdeveloped or non-existent, and such is of course the case in most environmental domains.

Neither can we contemplate the idea of measuring sustainability by just questioning people about it -as one is sometimes inclined to do for the measurement of current well-being. Such questions about individual or global perspectives are frequently asked, and the results are of course interesting. For instance, according to the 2006 edition of the Eurobarometer Survey conducted for the European Commission, it was 76% of French respondents that foresaw a more difficult life for their children than for themselves, with only 8% foreseeing the opposite. Such messages are interesting because of their sharp contrast with standard long-term projections of GDP/capita based on the extrapolation of current productivity trends. They reinforce the conviction that measuring sustainbility is a real issue. But they clearly do not provide a measure of sustainability. They simply measure feelings or beliefs concerning sustainability. We clearly need to go further than that. What we expect from statistics is to go beyond such feelings or day-to-day subjective perceptions.

All this means that it is definitely impossible to answer the question in the same way as usually done with accounts or social statistics. What is needed are projections, and not only projections of technological or environmental trends, but also projections of how they will interact with socio-economic or even political forces. Stated as such, the challenge is extreme. In practice, ambitions will remain much lower, i.e. just providing numbers identifying a risk of unsustainability under the continuation of current trends or behaviours. But even that task

remains a considerable one that goes much beyond the normal job of statisticians and/or of economists. It requires a much broader set of expertise than it is the case for usual accounting activities.

Message 2: Measuring sustainability also entails prior responses to normative questions. In this respect too, it strongly differs from standard statistical activity.

The coexistence of different appreciations of sustainability may reflect not only different predictions of what the future can be, but also different views about what will really matter tomorrow for us or our descendants. Let's put it in another way. Everybody should in principle converge on the idea that sustainability means the preservation of future well-being. But the question remains to know what well-being we wish to sustain exactly. Some may argue that we just have to warrant the constancy of GDP per capita. Some will accept to keep such a focus on monetary income, but they will put more emphasis on the intragenerational distribution of resources -as the Brundtland report actually did. They will therefore argue that we must try to sustain monetary income for the poorest segments of the population, and the policy implication may be different from the ones derived from the first objective. Some others can choose to put a stronger emphasis on the preservation of such or such element of the environment, such as biodiversity or the quality of landscapes, and so on…

Making choices in this respect goes once again far beyond the normal job or normal responsibility of statisticians: they can help clarify the options or help implementing correctly the index once the choices have been made, but they can in no way fully assume the definition of objectives.

Message 3: Measuring sustainability raises an additional difficulty in an international context. The question is not exclusively to assess relative sustainabilities of each country taken separately. The problem is global, at least in its environmental dimension. In that case, what is at stake is rather the contribution of each country to global sustainability or unsustainability.

We have seen that the treatment of this dimension is actually one strong source of divergence between differents approaches to sustainability, and explains contradictory results of these different approaches. More developed countries are, in some sense, the most sustainable ones, because they can devote significant resources to capital accumulation, both in its physical and human forms. It is not unsurprising to find that many less developed ones are on much more fragile trajectories on this economic side. But, from on another point of view, it is developed countries that are often the largest contributors to worlwide environmental unsustainability, at least in its climatic dimension.

All these messages must be looked at jointly. They all give arguments in favor of a non-monodimensional view of sustainability. Too much information undoubtedly goes against readability and public impact. But trying to put too much information in a too restrictive set of numbers, or even in only one number can symetrically leads to lose important points of the phenomenon that we try to capture.

On the whole, problems with sustainability measurement are large, but we need to offer solutions, how imperfect they can be. We shall formulate five recommandations in this direction.

Recommendation 1: Sustainability assessment requires a well-identified sub-dashboard of the global dashboard to be recommended by the Commission.

First of all, the complexity of the task strongly advocates in favor of treating it separately from the other issue that has been considered by the commission, i.e. the monetary or non monetary dimensions of *current* well-being. This recommendation to separate the two issues might look trivial. Yet it deserves emphasis, because some existing approaches fail to adopt this principle, leading to potentially confusing messages. The confusion reaches a peak when one tries to combine these two dimensions into a single indicator. This criticism applies not only to composite indices, but also to the notion of green GDP. To take an analogy, when driving a car, a meter that weighed up in one single value the current speed of the vehicle and the remaining level of gasoline would not be of any help to the driver. Both pieces of information are critical and need to be displayed in distinct, clearly visible areas of the dashboard.

Recommendation 2: The distinctive feature of all components of this sub-dashboard should be interpretable as variations of some "stocks".

Second, on a complicate subject where many misunderstandings can take place, it is good practice to first start by elaborating a common language or a common general framework. The one that we have tried to emphasize is the so-called "stock-based" or "capital-based" or "wealth-based" approaches to sustainability. The argument is that, ultimately, the sustainability issue is about how much stocks of resources we leave to future periods or future generations, and the question is whether we leave enough of these to maintain opportunity sets at least as large as the one we have inherited of. Saying this does not entail any a priori limitation in the list of these assets that matter for future well-being. Quite the contrary: their list can be extended as far as possible. Assessments of sustainability must be made on complete inventories of these stocks, and a good assessment of how they are currently changing, and of what are their expected paths of evolution. On the purely economic side, such an inventory is provided by capital accounts. Some solutions also exist for the measurement of human capital, both in terms of stocks or in terms of net flows. On the environnemental side, such an large scale inventory has been made for instance in the contect of initiatives such as the Millenium Ecosystems Assessment coordinated by the United Nations between 2001 and 2005. "Social capital" is also another dimension that is important to consider even if, on this point, quantification is a much more difficult issue.

Just to illustrate the relevance of this stock approach and the way it articulates to the measurement of current well-being, let's recall again how it very neatly solves one of the most well-known criticisms of standard GDP, the fact that, viewd as a well-being indicator, it can send the aberrant message that a natural catastrophe is a blessing for the economy, because of the additional economic activity generated by repairs. If we are able to apply it in a correct way, the stock approach will clearly recorded as a form of exceptional loss of natural or physical capital. The resulting increase in economic activity that can be induced by the catastrophe has a positive value only insofar as it helps to restore the initial level of the capital stock. It does not contribute to current well-being: without the catastrophe, people would have enjoyed more pleasant activities. It only helps to avoid a reduction in future well-being caused by the greater scarcity of these capital goods.

To conclude on this point, we must also recall that phrasing the issue of sustainability in terms of preservation of some "capital" goods does not mean that one consider that these

goods must be managed or traded as ordinary capital goods. Economists indifferently use the terms of wealth or capital for all forms of goods that can be transferred from one period to another, without any prior upon whether these goods should be privately or collectively owned, and upon whether their management should be fully left to market forces or not. To avoid such a misunderstanding, we have tried here to retain the more neutral term of wealth as much as possible. Now, whatever the words that are chosen, it should be possible to agree that the question of sustainability can be formulated as the question of knowing whether we pass enough of all these components of wealth to future periods or future generations. This is the reason why we have chosen to adopt here this formulation of the problem.

Recommendation 3: A monetary index of sustainability has its place in a sustainability dashboard, but under the current state of the art, it should remain essentially focused on economic aspects of sustainability.

Now, this stock-based approach to sustainability can in turn be broken down into two versions. One version would just look at variations in each stock separately with a view to doing whatever is necessary to keep it from declining or at least to keep it above some critical threshold beyond which further reductions would be highly detrimental to future well-being. Or one could attempt to summarize all stock variations in synthetic figures.

This second track is the one followed by so-called "extended wealth", "inclusive wealth" or "adjusted savings" approaches, which share the idea of converting all these assets into a monetary equivalent. We have discussed at length the potential of such an approach, but also several of its limitations. In certain conditions, it offers the ability to anticipate many forms of non-sustainability, but the requirements for such a capacity are extremely high. This is because the aggregation required by this approach cannot be based on market values: market prices are nonexistent for quite a large number of the assets that matter for future well-being. Even when they are available, there is no guarantee that they adequately reflect how these different assets will matter for future well-being. In the absence of such price messages, we have to turn to imputations, which raise both the normative and informational difficulties that have been outlined in our liminary messages.

All this suggests staying with a more modest approach, i.e. focusing the monetary aggregation on items for which reasonable valuation techniques exist, such as physical capital, human capital and fossil resources. This more or less corresponds to the hard part of "adjusted net savings" as computed by the World Bank and further developed by several authors. "Greening" this index more intensively is of course a relevant objective, and we can keep it on the agenda. We know what kind of analytical apparatus would be required for accomplishing this: large-scale projection models of interactions between the environment and the economy, including a proper treatment of the uncertainty that exists concerning this interaction by resorting to scenarios based on different relative prices as well as, possibly, on "stress tests". But, meanwhile, we must focus this indicator essentially on what it does relatively well, i.e. the assessment of the "economic" component of sustainability, that is, the assessment of whether or not countries overconsume their economic wealth.

Box 3: Physical and other non-monetary indicators: which ones to choose?

The commission's general position has been to avoid formulating definitive turnkey proposals on any the different issues it has raised. All proposals rather intend to stimulate further debate. This is all the more true in the domain of *physical* sustainability indicators where the expertise of specialists from other disciplines is crucial and was only indirectly represented in the commission's composition.

Some suggestions can however be made, in connection with conclusions of some recent related reports.

In 2008, a OECD/UNECE/Eurostat working group has produced a report on measuring sustainable development whose messages have several points in common with ours. It strongly advocates the stock-based approach to sustainability as the relevant way of structuring a micro-dashboard of sustainability indicators gathering both stock and flow variables. It also suggests a line of demarcation between determinants of 'economic' well-being (those that are the most directly amenable to monetary evaluation) and the determinants of 'foundational' well-being, among which four couples of stock/flow environmental indicators devoted respectively to global warming, other forms of atmospheric pollution, quality of water and biodiversity. The details and positions of these indicators in the dashboard can be visualized (in bold) on the table below.

Small set of sustainable development indicators proposed by the UNECE/OECD/Eurostat working group on sustainability measurement

Indicator domain	Stock indicator	Flow indicator
Foundational well-being	Health-adjusted life expectancy	Index of changes in agespecific mortality and morbidity (place holder)
	Percentage of population with post-secondary education	Enrolment in post-secondary education
	Temperature deviations from Normals	**Greenhouse gas emissions**
	Ground-level ozone and fine particulate concentrations	**Smog-forming pollutant emissions**
	Quality-adjusted water availability	**Nutrient loadings to water bodies**
	Fragmentation of natural habitats	**Conversion of natural habitats to other uses**
Economic well-being	Real *per capita* net foreign financial asset holdings	Real *per capita* investment in foreign financial assets
	Real *per capita* produced capital	Real *per capita* net investment in produced capital
	Real *per capita* human capital	Real *per capita* net investment in human capital
	Real *per capita* natural capita	Real *per capita* net depletion of natural capital
	Reserves of energy resources	Depletion of energy resources
	Reserves of mineral resources	Depletion of mineral resources
	Timber resource stocks	Depletion of timber resources
	Marine resource stocks	Depletion of marine resources

Source: UNECE/OECD/Eurostat (2008)

More recently, the French Economic, Social and environmental Council (CESE) has produced a report (Le Clézio, 2009) whose initial aim was the assessment of the Ecological Footprint but that has more widely explored the different tracks available for quantifying sustainability. It has the same messages as the current report concerning the limits of this EF index, and the fact that most of the relevant information that it conveys is more directly and more neatly reflected in one of its subcomponents, the carbon footprint. As a consequence, it strongly advocates in favor of this index. Compared to Global GHG emissions suggested in the OECD/UNECE/Eurostat Dashboard presented above, the Carbon Footprint has the advantage of being expressed in this 'footprint' unit that is intuitively so appealing and that has made the success of the EF. In addition to this, this CESE report has suggested emphasizing the other physical indicators already present in the large international dashboards such as the one elaborated for the European Strategy for Sustainable Development whose complete list is given in appendix 1. Some of them are those already quoted in the OECD/UNECE/Eurostat dashboard.

As far as climate change is concerned, some other indicators can be considered. Direct observation of mean temperature is one possibility but not the best suited, because it has a tendency to run behind the main components of climate change and because there can always be disagreements about the causes of temperature rises, hence about their permanent or transient character. Consequently, climatologists prefer

to make use of a thermodynamic concept, the CO_2 radiative forcing, that measures the earth energy imbalance created by the action of CO_2 as a greenhouse gas.

Alternatively, it is possible to directly use a notion of CO_2 remaining budget: according to climatologists, there is an upper limit of 0.75 trillion tonnes of carbon that might be discharged in the atsmosphere if the risk of temperatures exceeding 2° Celsius above pre-industrial levels is limited to one-in-four, this upperbound at 2°C being largely accepted among climate experts as a "tipping point" opening the door to unstoppable feedback effects (methane from melting permafrost , CO_2 and methane from decaying tropical forests ,all sorts of greenhouse gases released by saturated warming oceans, etc...). Of this 0.75 total budget, emissions to 2008 have already consumed circa 0.5. Hence the importance of monitoring this remaining CO_2 budget. The attractiveness of this indicator is to be strongly consistent with the stock-based approach to sustainability. It can be also rephrased in the very expressive terms of a countdown index, i.e. the time that remains until exhaustion of this stock, under the assumption of emissions remaining on their current trend. This kind of representation is often used for other forms of exhaustible resources.

Still other indirect indicators of global warming are the regression of permanent ice or the oceanic pH. The regression of permanent ice has the advantage of being an advanced one and to be directly related to manifest effects. The oceanic pH increases with the amount of CO_2 that is naturally pumped into the oceans. A consequence of this increase is a decrease in the quantity of phytoplancton, which is itself a carbon sink no less important than the forests. One may therefore say that the physical sink (sea water dissolving atmospheric CO_2) destroys the biological one. This is the reason why the oceanic pH appears to be another good tentative indicator of climate change, pointing to one of the most vicious feedback effects.

Among criteria for choosing between all the indicators, two are of particular importance. One is their appropriability by the public, the other is the capacity of declining them at national or even subnational levels: in this respect, the carbon footprint has quite a lot of advantages.

As far as biodiversity is concerned, the issue is currently under review by the TEEB group ("the economics of the environment and biodiversity") working at the initiative of the European Union (Sukhdev, 2008) and it has been also recently addressed by a report by the French *Conseil d'Analyse Stratégique*, in this case with the idea of pushing as far as possible the monetization of this dimension (Chevassus-au-Louis, 2009). The reason for this search of monetary equivalent is essentially that it may foster incorporation of this dimension in investment choices: many public decisions such as building a new motorway imply some potential biodiversity loss through fragmentation of natural habitats. But the report provides also a very detailed and technical review of available physical measures of biodiversity, to which the reader is referred for further information.

At last, moving away from environmental preoccupations, but still on the "non-monetary" side, one important issue is the issue of social capital and "institutional assets" that we transmit to future generations. One will have noticed that the UNECE/OECD/Eurostat dashboard presented above did not propose any indicator of this kind, not because the question is not relevant, but mainly because of lack of consensus about the way to measure it. Subgroup 3 was not in a position to explore this question further, but efforts along this direction remain undoubtedly necessary.

Recommendation 4: The environmental aspects of sustainability deserve a separate follow-up based on a well-chosen set of physical indicators.

As far as environmental sustainability is concerned, the limitations of monetary approaches do not mean that efforts to monetize damages to the environment are no longer needed: it is well known that fully opposing any kind of monetization often leads to policies that act as if environmental goods had no value at all. The point is that we are far from being able to construct monetary values for environmental goods that at the macro level can be reasonably compared to market prices of other capital assets. Given our state of ignorance, the precautionary principle legitimates a separate follow-up of these environmental goods.

In fact, the fundamental reasons for a separate treatment of these environmental issues directly stem from our liminary messages: these are the items for which projections are the most difficult to produce, with strong technological uncertainties, a high difficulty to

appreciate how they are likely to be valued by future generations and a separate follow-up of the corresponding physical of indicators can be viewed, at least at this stage, as a simple way to this need of a specific treatment. In addition, these goods often have this characteristic of being global public goods, as it is the case for climate. All these elements go in the same direction of favoring a separate follow-up.

Now, what are the stock/flow indicators that are the best suited for this separate follow-up ?

The ecological footprint was one of the options for this separate follow-up. In particular, in contrast to net adjusted savings, it essentially focuses on contributions to global non-sustainability, with the message that the main responsibility lies with the developed countries. Yet the group has taken notice of its limitations, and in particular that it is far from being a pure physical indicator of pressure on the environment: it retains some aggregation rules that may be problematic. In fact, it appears that much of the information that it conveys about national contributions to non-sustainability is imbedded in a simpler indicator, the carbon footprint, which is therefore one good candidate for monitoring humanity's pressure on the climate.

Now, on this general issue of physical indicators, a group of economists cannot claim for any comparative advantage. The discussion of which stock measures should be ideally included in the micro-dashboard we are calling for necessarily requires other forms of expertise, before being submitted to public discussion. Box 3 just gives some examples of how the same problem has been treated recently by equivalent groups or commissions, or suggest some additional tracks for the specific case of global warming.

In short, our pragmatic compromise is to suggest a small dashboard, firmly rooted in the logic of the "stock" approach to sustainability, which would combine:

– An indicator more or less derived from the extended wealth approach, "greened" as far as possible on the basis of currently available knowledge, but whose main function, however, would be to send warning messages concerning "economic" non-sustainability. This economic non-sustainability could be due to low savings or low investment in education, or to insufficient reinvestment of income generated by the extraction of fossil resources (for countries that strongly rely on this source of income).
– A set of well-chosen *physical* indicators, which would focus on dimensions of environmental sustainability that are either already important or could become so in the future, and that remain difficult to capture in monetary terms.

The points of convergence between this scenario and conclusions reached by some of the other reports mentioned in Box 3 are reassuring: they suggest that, from the relatively confused situation that was described in section 2, we are steadily moving towards a more consensual framework for the understanding of sustainability issues[11].

A subsidiary question concerns a user's guide to such a dashboard. A warning should be given that no limited set of figures can pretend to forecast the sustainable or unsustainable character of a highly complex system with certainty. The purpose is, rather, to have a set of indicators that give an "alert" to situations that pose a high risk of non-sustainability. Whatever we do, however, dashboards and indices are only one part of the story. Most of the efforts involved in assessing sustainability have to focus on increasing our knowledge about how the economy and the environment interact now and are likely to interact in the future.

11. Some other points of convergence can be found in reactions by the European Environmental Agency to the first draft of this commission's summary report (European Environmental Agency, 2009).

Appendix 1: Reviewed list of European sustainable development indicators

(Source: Eurostat, 2007)

Level 1	Level 2	Level 3
Theme 1: Socio Economic Development		
1. Growth rate of GDP per inhabitant	***Sub-theme: ECONOMIC DEVELOPMENT***	
	2. Gross investment, by institutional sector	3. Dispersion of regional GDP per inhabitant
		4. Net national income
		5. Gross household saving
	Sub-theme: INNOVATION, COMPETITIVENESS AND ECO-EFFICIENCY	
	6. Labor productivity per hour worked	7. Total R&D expenditure
		8. Real effective exchange rate
		9. *Turnover from innovation, by economic sector**
		10. *Effects of innovation on material and energy efficiency**
		11. Energy intensity
	Sub-theme: EMPLOYMENT	
	12. Total employment rate	13. Employment rate, by gender and by highest level of education attained
		14. Dispersion of regional employment rates
		15. Unemployment rate, by gender and age group
Theme 2: Sustainable consumption and production		
1. Resource productivity	***Sub-theme: RESOURCE USE AND WASTE***	
	2. *Generation of total waste*, by economic activity and GDP* (proxy: Municipal waste generated per inhabitant)	3. Components of Domestic Material Consumption
		4. *Environmental impact of material consumption** (proxy: Domestic Material Consumption, by material)
		5. Municipal waste treatment, by type of treatment method
		6. *Generation of hazardous waste, by economic activity**
		7. Emissions of acidifying substances, ozone precursors, and particulate matter by source sector, and GDP
	Sub-theme: CONSUMPTION PATTERNS	
	8. Electricity consumption by households	9. Final energy consumption, by sector
		10. Consumption of certain foodstuffs per inhabitant
		11. Motorisation rate
	Sub-theme: PRODUCTION PATTERNS	
	12. Enterprises with a registered environmental management system	13. Eco-label awards
		14. Area under agri-environmental commitment
		15. Area under organic farming
		16. Livestock density index
Theme 3: Social inclusion		
1. At-risk-of-poverty rate after social transfers	***Sub-theme: MONETARY POVERTY AND LIVING CONDITIONS***	
	2. At-persistent-risk-of poverty rate	3. At-risk-of-poverty rate, by gender, by age group, and by household type
		4. Relative at risk of poverty gap
		5. Inequality of income distribution
	Sub-theme: ACCESS TO LABOUR MARKET	
	6. People living in jobless households, by age group	7. In-work poverty
		8. Total long-term unemployment rate
		9. Gender pay gap in unadjusted form
	Sub-theme: EDUCATION	
	10. Early school leavers	11. At-risk-of-poverty rate, by highest level of education attained
		12. Persons with low educational attainment, by age group
		13. Life-long learning
		14. *Low reading literacy performance of pupils**
		15. *ICT skills**

* Indicators under development.

Level 1	Level 2	Level 3
Theme 4: Demographic changes		
1. Employment rate of older workers	***Sub-theme: DEMOGRAPHY***	
	2. Life expectancy at age 65, by gender	3. Total fertility rate
		4. Net migration, by age group
	Sub-theme: OLD-AGE INCOME ADEQUACY	
	5. Aggregate replacement ratio	6. At-risk-of-poverty rate for persons aged 65 years and over
	Sub-theme: PUBLIC FINANCE SUSTAINABILITY	
	7. General government consolidated gross debt	8. Average exit age from the labor market
Theme 5: Public Health		
1. Healthy life-years and life expectancy at birth, by gender	***Sub-theme: HEALTH AND HEALTH INEQUALITIES***	
	2. Death rate due to chronic diseases, by age group	3. Healthy life-years and life expectancy at age 65, by gender
		4. Suicide death rate, by gender and by age group
	Sub-theme: DETERMINANTS OF HEALTH	
	5. Salmonellosis incidence rate in human beings 6. Index of production of chemicals, by toxicity class	7. *Overweight people, by age group**
		8. *Present smokers, by gender and by age group**
		9. Population exposure to air pollution by particulate matter
		10. Population exposure to air pollution by ozone
		11. *Population living in households considering that they suffer from noise**
		12. Serious accidents at work
Theme 6: sustainable development		
1. Total greenhouse gas emissions 2. Consumption of renewables	***Sub-theme: CLIMATE CHANGE***	
	3. Greenhouse gas emissions by sector	4. Greenhouse gas intensity of energy consumption
		5. Projections of greenhouse gas emissions
		6. Global surface average temperature
	Sub-theme: ENERGY	
	7. Energy dependency	8. Gross inland energy consumption, by fuel
		9. Electricity generation from renewables
		10. Consumption of biofuels by transport
		11. Combined heat and power generation
		12. Implicit tax rate on energy
Theme 7: sustainable transport		
1. Energy consumption of transport	***Sub-theme: TRANSPORT GROWTH***	
	1. Modal split of freight transport 2. Modal split of passenger transport	4. Volume of freight transport and GDP
		5. Volume of passenger transport and GDP
		6. Energy consumption, by transport mode
	Sub-theme: TRANSPORT PRICES	
	7. Road fuel prices	
	Sub-theme: SOCIAL AND ENVIRONMENTAL IMPACT OF TRANSPORT	
	8. Greenhouse gas emissions by transport, by mode 9. People killed in road accidents, by age group	10. Average CO2 emissions per km from new passenger cars
		11. Emissions of ozone precursors from transport
		12. Emissions of particulate matter from transport
Theme 8: Natural resources		
1. Common Bird Index 2. Fish catches taken from stocks outside safe biological limits	***Sub-theme: BIODIVERSITY***	
	3. Sufficiency of sites designated under the EU Habitats *and Birds* directives	4. *Red List Index for European species* *
	Sub-theme: FRESH WATER RESOURCES	
	5. Surface and groundwater abstraction	6. Population connected to wastewater secondary treatment systems
		7. Biochemical oxygen demand in rivers
	Sub-theme: MARINE ECOSYSTEMS	
	8. *Concentration of mercury in fish and shellfish**	9. Size of fishing fleet
	Sub-theme: LAND USE	
	10. *Land use change*, by category* (proxy: Built-up areas) 11. Forest increment and fellings	12. Forest trees damaged by defoliation
		13. *Critical load exceedance for nitrogen* *

* Indicators under development

Level 1	Level 2	Level 3
Theme 9: Global Partnership		
1. Official Development Assistance (ODA)	***Sub-theme: GLOBALISATION OF TRADE***	
	2. EU imports from developing countries, by income group	3. EU imports from developing countries, by group of products
		4. EU imports from least developed countries, by group of products
		5. Aggregated measurement of support
	Sub-theme: FINANCING FOR SUSTAINABLE DEVELOPMENT	
	6. Total EU financing for development, by type	7. Foreign direct investment in developing countries, by income group
		8. ODA, by income group
		9. Untied ODA
		10. Bilateral ODA dedicated to social infrastructure and services
		11. Bilateral ODA dedicated to debt
	Sub-theme: GLOBAL RESOURCE MANAGEMENT	
	12. CO2 emissions per inhabitant in the EU and in developing countries	13. Bilateral ODA dedicated to water supply and sanitation
Theme 10: Good governance		
	Sub-theme: POLICY COHERENCE AND EFFECTIVENESS	
	1. New infringement cases, by policy area	2. Transposition of Community law, by policy area
	Sub-theme: OPENNESS AND PARTICIPATION	
	3. Voter turnout in national and EU parliamentary elections	4. E-government on-line availability
		5. E-government usage by individuals
	Sub-theme: ECONOMIC INSTRUMENTS	
	6. Shares of environmental and labor taxes in total tax revenues	

* Indicators under development.

Appendix 2: Measuring human capital

What it is and why it matters?

Assessing the sustainability of a given development path ideally requires monetary estimates for all the types of capital stocks that contribute to people's well-being. This raises challenges for measurement that go well beyond those pertaining to natural and environmental resources and that extend to all types of assets that are not traded in markets. This is the case of human capital.

Human capital refers to the knowledge, skills and attributes that are embodied in each person and that facilitate the creation of different forms of well-being (Keeley, 2008). Adam Smith talked about human capital more than three centuries ago arguing that economic activity is fuelled not by workers as a collective mass but by the "acquired and useful abilities of all inhabitants or members of society" and that these abilities, once attained, stand as "a capital fixed and realised, as it were, in his person". Since the second half of the XXth century, human capital has played an increasingly important role in discussions about the determinants of economic growth, poverty and inequality.

The boundaries of what is conventionally labelled as human capital are not always well-defined. The formal education system obviously plays a critical role in sustaining the accumulation of human capital, but this accumulation can also reflect learning processes developed outside schools, such as early childhood education and on-the job training. Health can also be regarded as part of human capital, although the challenges posed by its measurement are distinct from those relating to skills and competences.

The benefits of human capital are also broad: they are both economic and social, and they may accrue both the person making this investment and to the community of which they are part. On the economic side, the benefits of human capital investment take the form of higher income and earnings capacity for the individual making this investment: these private returns to education have increased strongly over the past few decades in many OECD countries, contributing to wider income inequalities between people with different education. At the aggregate level, investment in human capital is also a critical factor for economic growth, for the simple reason that this growth is, more than ever, based on technical advances that demand workers always more skilled and qualified. Investment in human capital also delivers social returns, such as higher life-expectancy for more educated people, lower undesired fertility in less-developed countries, and greater participation in civic and social life.

Because of this range of payoffs, and of its links to a variety of other fields (such as health, paid work and caring), the concept of human capital enters contemporary debates in a variety of forms: as a driver of economic growth and innovation; as an investment to secure greater access to jobs, higher income and lower poverty; and as one of the assets that should be preserved and developed – on par with natural capital and other types of resources – to secure sustainable development.

Measuring human capital: from indicators to accounts

Despite its importance, the measurement of human capital remains elusive. Of course, physical indicators (such as the average years of schooling for the population of working age) exist, but they have important limits: for example, they ignore differences in competencies among people with the same educational attainment; they disregards what people learn outside school, both on the job and through exposure to more accessible and extensive information flows; and they neglect the amount of resources (in terms of both time and money) that feeds these learning processes. Direct measures of people's competencies, administered through standardised tests, also provide important indicators of the "quality" of the competencies gained, but those currently available are mainly limited to students of a given age (15 years old, in the case of the OECD's *Programme on International Student Assessment*). While these physical indicators allow benchmarking countries' performance in the educational field from a variety of perspectives, they lack a framework that allows relating to each other the various elements that shape the accumulation of human capital and assessing their contribution through a common yardstick.

While the task of developing such a framework may have appeared far-fetched only a few years ago, it is feasible today (Fraumeni, 2008). Comprehensive and reliable accounts on human capital can be constructed using the same accounting framework that is used for standard economic statistics. Practical guidance to the construction of these accounts for a variety of non-market activities is provided by a report prepared by a panel of the US National Academy of Sciences (Abraham and Mackie, 2005). This report stressed the importance of developing independent measures of both the inputs and the outputs of non-market activities, and of quantifying both the values and the volumes of each of the two sides of these accounts.[12] In the human capital field, these accounts could initially be limited to formal education, and then be extended in steps to other dimensions, such as the investments made to prepare children for formal schooling and that made to maintain and develop the competencies of people of working-age through on-the-job training and adult education. "Formal education", in the perspective of non-market accounts, encompasses the joint use of inputs by households and educational establishments to produce educational services, as opposed to the more narrow definition of the education industry in the national accounts, which largely ignores the role of private households in the production of these services.

A comprehensive satellite account for formal education would integrate values of both the inputs into these learning processes, and of the outputs produced. This requires indentifying the most salient items entering the two sides of the accounts and solving specific measurement problems.

- On the inputs side, the most important categories are the monetary expenditures of both households and educational institutions for paid work (teachers and support personnel), intermediate inputs (materials such as books) and capital inputs (school building, equipment and software). On top of these expenditures, the unpaid time devoted to education by the students themselves, by their parents providing home-

12. While, for market transactions, the two sides of the accounts are expected to balance out (with the statistical discrepancy providing an indicator of the overall quality of these estimates), large imbalances are possible in the case of non-market accounts; this may reflect failures in the measuring of all relevant inputs or in the pricing of various outputs.

support, and by other people and associations who may support the learning of students with specific difficulties or in special areas (e.g. mentors, tutors) constitutes inputs to the production of education services. The valuation of these unpaid time-inputs requires, on the volume side, good estimates of how people allocate their time (from either administrative records, for the time spent by students in classes, or from time-use surveys); and, on the value side, estimates of appropriate accounting prices for these inputs. According to Abraham and Mackie (2005) non-market inputs that could have been provided by a third party should ideally be valued based on their replacement costs (e.g. the time spent by parents in helping their kids to perform their home-work could be valued at the wage of a private teacher providing the same support, possibly adjusted for differences in the skills and efforts between the paid-service provider and the nonmarket worker); while inputs for which own time is required, such as students' time in classes, should ideally be valued based on opportunity costs (possibly adjusted for the value of enjoyment derived from non-market work).[13]

- On the output side, the value of education services produced is measured as the addition to human capital through education. Human capital will increase as a consequence of education if the latter leads to higher earnings and productivity on the workplace of better educated people. Other items that could also be included on the output side are the effect of education on non-market productivity (such as the greater rewards from some leisure activities enjoyed by more educated people) and, ideally, the returns of education for society at large, such as those that flow from creating citizens that are better informed, more tolerant, and more willing and capable of interacting with others.

Valuing human capital stocks: the discounted life-time income approach

In practice, comprehensive accounts of human capital, even when limited to formal education, would need to be developed in steps. From a sustainable development perspective, the immediate focus of research is likely to be measuring the total stock of human capital per capita, and to monitor how this stock changes over time. Different approaches have been used for this goal. One approach derives measures of the stock of human capital as a "residual" from an accounting identity. For example, World Bank's estimates of "intangible capital" are computed as the difference between a total measure of the capital stock (the discounted sum of average consumption expenditure over the past three years, summed over an horizon of 25 years) and estimates of the values of natural resources, physical capital and financial holdings (World Bank, 2005). Similarly, other authors have measured human capital by estimating, first, a "return of human capital" (the difference between Net National Income in a given year and net returns on fixed, financial, and natural resources, i.e. resource rents); and, second, by measuring the stock of human capital as the present value of these human capital returns (Greaker, 2007). However, these "indirect" estimates of the stock of human capital are affected by errors in measuring all the terms entering these identities; also, by construction,

13. The extent to which society fosters and values the learning of youths, and the peer effects of other students (which are forms of social capital) may also be considered as inputs that enhances students' learning experiences and opportunities; monetary valuation of these effects, however, do not exist.

they reflect the assumption that some monetary measure of income or consumption provides as a suitable proxy of human well-being. [14]

While these approaches may be useful for research purposes, most practitioners in the field of human capital measurement favour a "direct" measure based on the discounted life-time income of each individual. This approach, pioneered by Jorgenson and Fraumeni for the United States (1989, 1992)[15], relies on data on the observed earnings of people with various levels of educational, as well as on information on the number of people by age and gender, their highest level of education attained and their participation in the labour force. In theory, the wage premium for people with different education may signal their innate competencies, and the sorting of people that occurs through the schooling system, rather than the effect of education in improving people's productivity. In practice, the view that these wage premia mainly reflect the effects of education in enhancing people's productivity is shared by most researchers in this field.

Empirical estimates of human capital based on variants of the discounted life-time income approach already exist for several OECD countries such as Australia, Canada, New Zealand, Norway, Sweden and the United States. While these estimate differ in a number of respects (e.g. the population covered, the inclusion of non-market income), and assumptions (e.g. the future pace of real earnings growth by education level, the rate used to discount these future income flows), they agree in showing that the value of human capital is substantial, and much larger than that of conventional types of capital: in other terms, currently available measures of a country's total wealth, as available from national balance-sheets produced by national statistical offices, omit the largest component of this wealth.[16] These approaches based on discounted life-time income also allow describing the composition of human capital across educational attainment levels, gender and age (Table 1); and to decompose the pace of human capital accumulation in terms of various drivers such as population ageing, migration and changes educational attainment of various cohorts. Aggregate estimates of human capital can also be tested for their sensitivity to various assumptions about future earnings growth and discount rates.

14. Arrow et al (2008) provide proxy estimates of human capital for China and the United States measured as the product of its total quantity (a function of the average education attainment of each country and of their adult population) and its rental price (the discounted sum of each country's wage bill per unit of human capital employed, cumulated over the average number of working-years remaining). Based on these estimates, for both China and the United States, the increase in human and reproducible (fixed and financial) capital greatly exceeds the loss from the depletion of natural capital.

15. This was the conclusion reached at a workshop on the measurement of human capital organised by the OECD and the Fondazione Giovanni Agnelli on 3 - 4 November 2008 in Turin, Italy.

16. For example, estimates presented at the Turin workshop, suggest that the value of human capital (excluding non-market income) is about 1.7 times that of net fixed produced assets in the United States, 3.4 times in Australia, 4 times in Canada, and six times in Norway. Because of differences in assumptions and population coverage, these estimates are not comparable across countries.

Table A.3.1.: Estimates of the value of human capital in Australia

	1981	1986	1991	1996	2001
			Male		
Higher degree	42 917	52 562	92 185	127 009	161 362
Bachelor degree	244 123	315 558	448 212	607 439	733 190
Skilled labour	840 709	943 680	1,039949	1 143 195	1 259 752
Unskilled labour	1,540987	1 685 260	1 889 659	1 950 974	1 957 450
Total	2 668 736	2 997 060	3 470 005	3 828 618	4 111 754
			Female		
Higher degree	9 485	14 002	30 389	55 730	90 579
Bachelor degree	106 458	160 347	305 251	489 443	663 789
Skilled labour	349 437	420 986	429 201	488 993	553 664
Unskilled labour	1 251 790	1 353 062	1 569 421	1 623 914	1 616 411
Total	1 717 170	1 948 398	2 334 262	2 658 080	2 924 442
			Entire population		
Total	4 385 906	4 945 457	5 804 266	6 486 698	7 036 196
Memorandum Item					
Produced capital	..	..	..	..	2 057 912
Ratio of human capital to fixed assets	..	..	..	..	3.4

Source: Estimates presented at the workshop on the measurement of human capital organised by the OECD and the Fondazione Giovanni Agnelli, 3- 4 November 2008, Turin.

Developing human capital accounts would provide a number of payoffs. First, they provide a single monetary estimate of human capital, which can be compared across countries and over time. Second, the possibility to decompose changes in the value of human capital into various drivers, would also allow projecting its evolution in the future, for example when older cohorts of less educated people are replaced by more educated ones. Last, by looking at the full range of inputs into education, they would favour policy dialogue between the various ministries and agencies that share responsibilities in this field, in particular when these accounts are extended beyond formal education. Human capital accounts based on individual data could also allow measuring inequalities in its distribution, and to address concerns linked to equity in access and outcomes for people with various characteristics. The potential of offered by better measures in this filed is huge; to realise it will require involving the statistical community in individual countries, undertaking investment to produce these estimates at regular intervals, and favouring cross-country collaboration through a phased time-plan (Fraumeni, 2008).

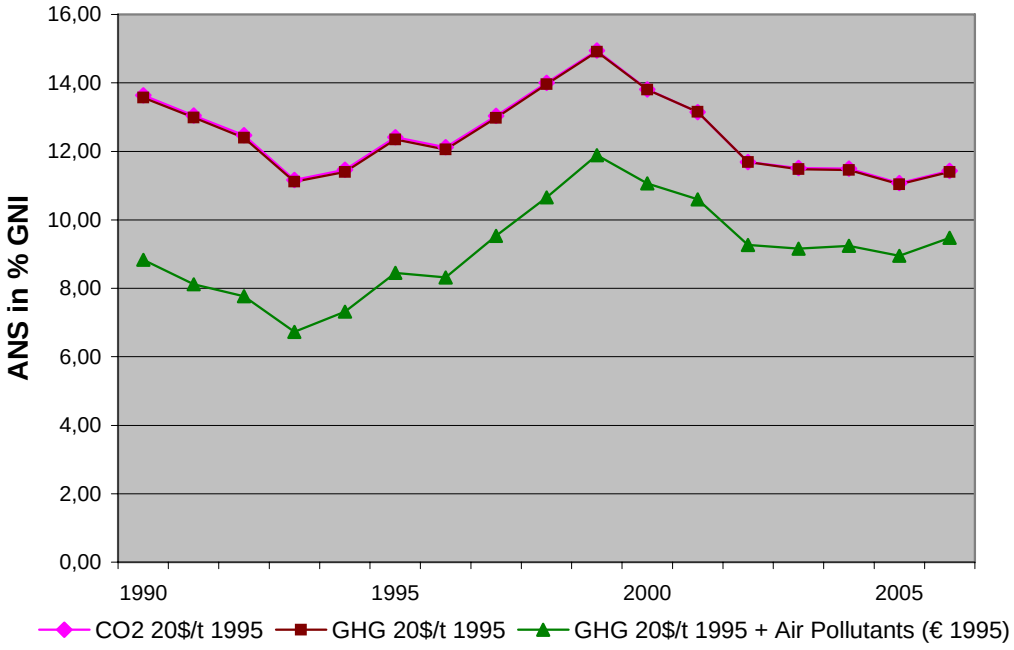
Appendix 3: Adjusted Net Savings and environmental assets: some robustness checks

As emphasized in the text, the relevance of the ANS approach crucially depends on **what** is counted (the different forms of capital passed through to future generations), namely what is included in the "extended wealth", and on **the price** used to count and aggregate (what is counted) in a context of imperfect valuation by markets. This appendix examines how the index is sensitive to variants of both kinds, focussing on natural resources.

Shortcomings and improvements: the "what"

One 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" (Where is the wealth of Nations, 2006, pp. 154), and *a fortiori* biodiversity loss. " This results from the lack of internationally comparable data, rather than intended omissions".

Figure A.3.1: Additional air pollution damages in France: substantial level effects



To perform sensitivity tests according to the scope of pollution damages covered, we focus on France and replicate what is done in Nourry (2008). We extend CO2 damages to all greenhouse gazes (GHG) expressed in tons of equivalent CO2 and replicate the incorporation of additional damages linked to Sulphur Dioxide (SO2), Nitrogen Oxide (Nox), Carbon Monoxide (CO) and Volatile Organic Components (VOC) in line with Noury (2008), and the damages linked to particulate matters (PM10) emissions (World Bank). The marginal damage costs proxies are based on previous studies by Rabl and Spadaro (2001) and calculations by Noury (2008): 5245 €/t of So2, 8093 €/t of Nox, 970 €/t of CO, 5762 €/t of VOC and 7265 €/ t of PM10.

Those calculations lead to substantial changes in the level of the final ANS, as illustrated for France on figure A.3.1. However, France still faces a comfortable positive value of its ANS indicator. On average, the "augmented" ANS is 3 percentage points below the standard ANS calculations (excluding PM10), namely slightly more than 9% instead of an initial 12.5% of GNI.

Shortcomings and Improvements: relative prices

Another key ingredient of the World Bank's estimates of ANS is their reliance on current prices for natural exhaustible resources. In theory, the use of market prices to evaluate flows and stocks is only warranted in a context of complete, competitive 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[17].

Moreover, whether competitive or not, the World Bank approach takes current prices as relevant metrics for all future periods. But those prices cannot be expected to remain stable. According to the Hotelling rule, they should increase as the resources are becoming scarcer. Consequently, those changes would have to be incurred as capital gains by exporting countries and corresponding losses for importing ones (while being neutral at the worldwide level), as highlighted by Arrow et ali. (2008).

As for pricing environmental degradations, things turn out to be even trickier. Indeed, in the absence of any market valuation, accounting prices must be evaluated by modeling long term consequences of given changes in environmental capital and how they impact prospective well-being. This point is elaborated more systematically in the text. We shall only examine here by how much the ANS can be sensitive to alternative numerical values of these prices, focussing on the pricing of CO_2 emissions, namely the estimation of the marginal social cost of one additional ton of carbon. Debates have been lively on that topic but as a consequence, estimations of the good price for carbon have been flourishing. As sustainability always involves a view of the future, pricing CO_2 emissions is only partially a problem of what CO_2 emissions costs now, but rather how much those should be priced tomorrow.

17. Alternatively, the El-Serafy's method could have been used. Dietz and Neumayer (2004, pages 284-285) noticed that the WB calculus leads to extremely large negative values of ANS for some oil exporting countries at some periods (up to -30% for Saudi Arabia in the 1970s), meaning a quasi-exhaustion of this country's wealth within only a few decades. The El-Serafy's method would have produced much more realistic values.

In the following graphs, we displayed revised values of the ANS according to different marginal social costs that have been used or proposed. The original release of the ANS retained a value of 20$ (1995) per ton of carbon (deflated for subsequent years by GDP deflator). A recent update by Arrow et ali. (2008) proposed a value of 50$ (2005) per ton of carbon. In France, the government has been involved in that topic for several years as public choice for public investments does take into account the environmental costs induced in the cost/benefits tradeoff (Boiteux and Baumstrark 1994 and 2001). A study led by Boiteux proposed a value of 100€ per ton of carbon in 2000 (i.e 27€ per ton of CO2). A recent update of this report (Quinet, 2008) targeted a value of roughly 370€ per ton of carbon in 2030, with a 4% discount rate. Proceeding backward, it provides us with a value of 45€ per ton of CO2 in 2010.

Those substantial differences in prices do not alter that much the overall ANS, as shown in the graphs. Indeed, the main differences appear when doubling the price from the 50$ per ton in Arrow *et al. (2008)* to the 100€ per ton in the French official assessment. Pricing assumptions affect only marginally the overall level of the ANS.

The interesting part of the 2008 Quinet report is the projected path of the cost of CO2 emissions. Assuming a soft recovering from the value of 27€ per ton of CO2 in 2000 to the target of 100€ per ton of CO2 in 2030, we are able to compute several forecasts of the ANS, along different scenarios.

Two sets of assumptions are addressed here, as an example. First, the ANS (excluding the CO2 damage part) could continue from 2006 following the trend observed since 1990 or return smoothly to a (supposed) long-run value (equal to the average ANS on 1980-2006). Secondly, the intensity ratio (CO2 emissions per units of GNI) may be still decreasing over that horizon at the pace observed since 1990, or remain constant at this 2006 level (less probably). Those assumptions therefore leave us with 4 scenarios, under the same path for the CO2 price.

Figure A.3.2.: Valuation effects for CO2 damages: France

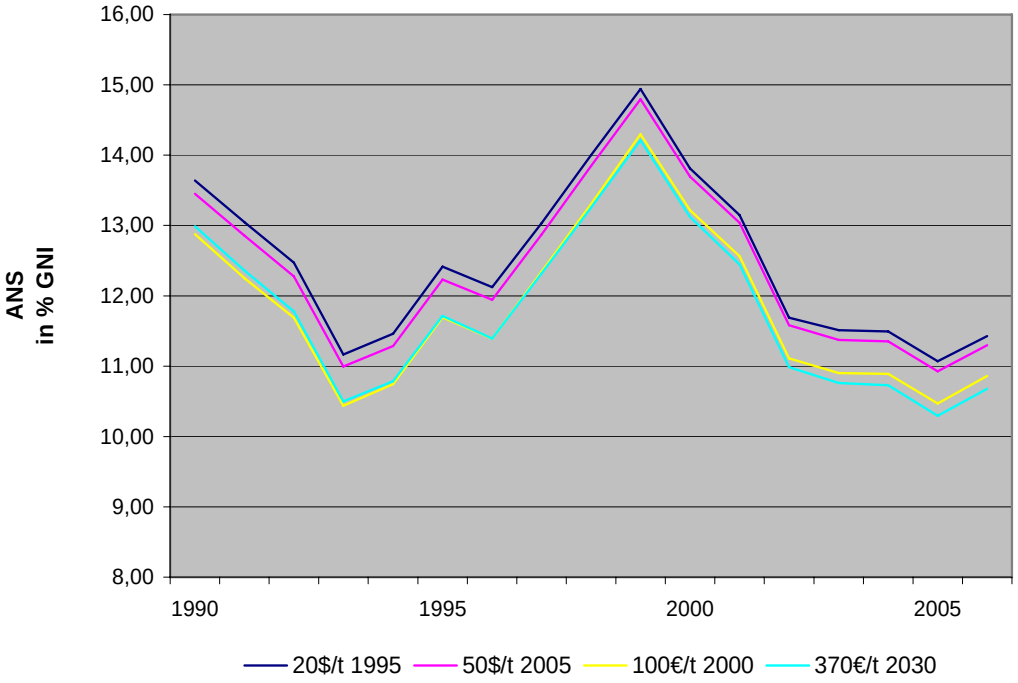


Figure A.3.3: Valuation effects for CO2 damages: USA

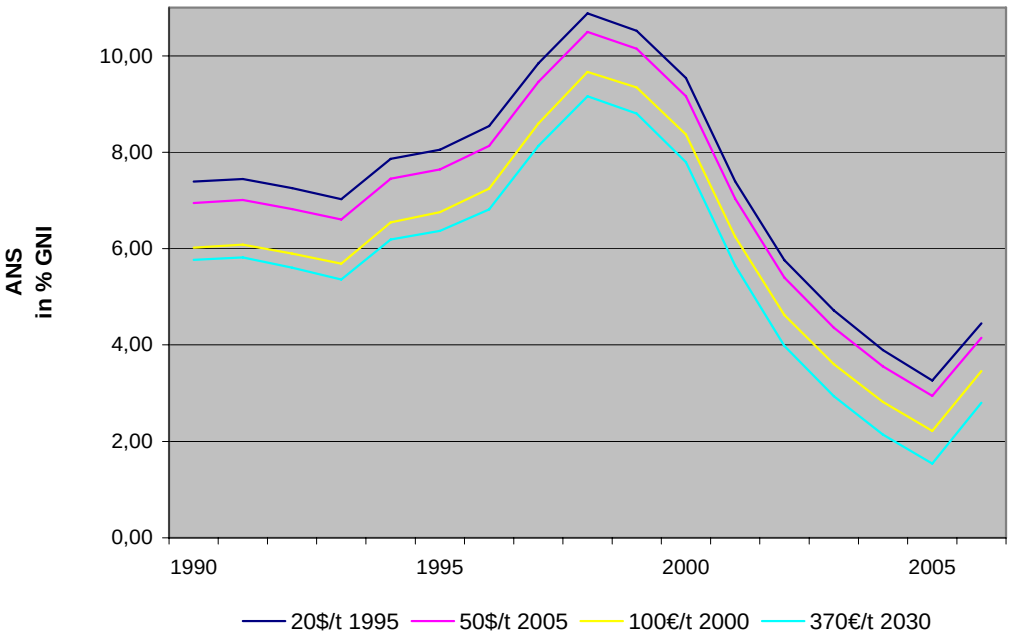


Figure A.3.4: ANS forecasts along different scenarios: France, 100 €/t CO2 2030

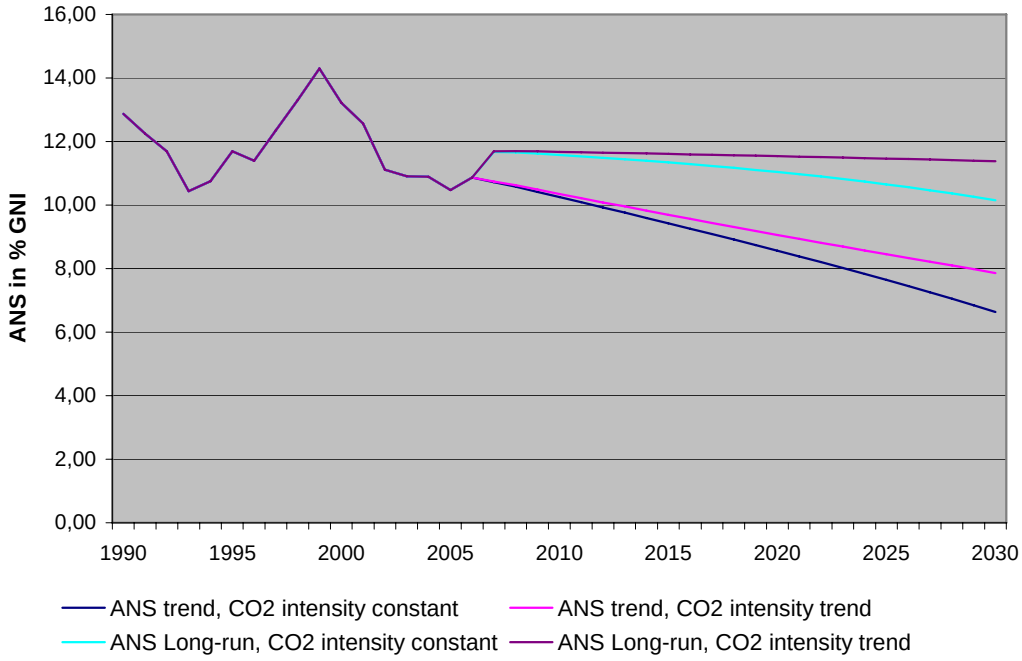
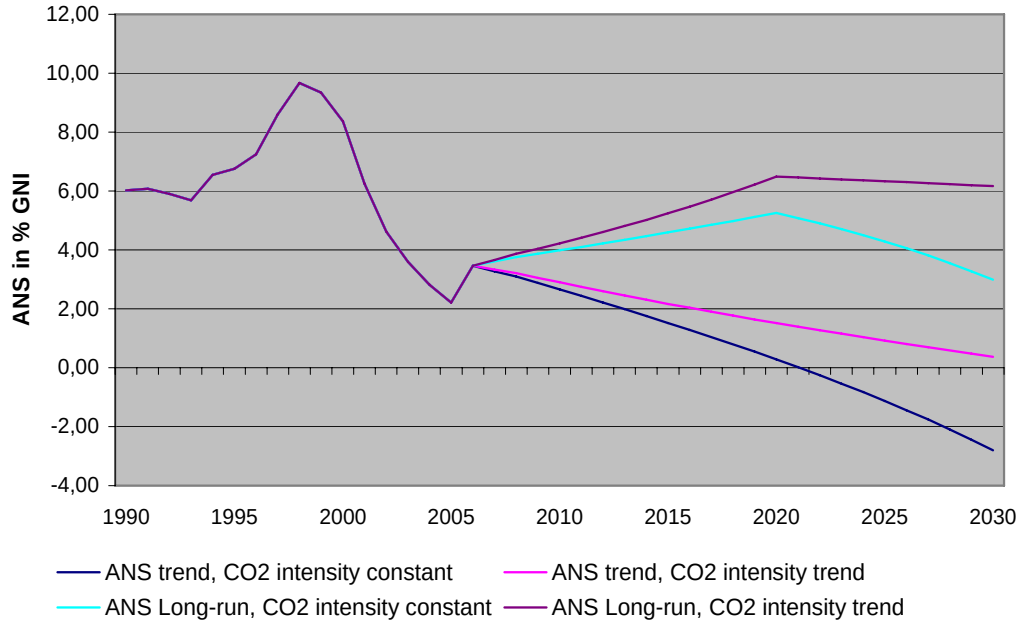


Figure A.3.5: ANS forecasts along different scenarios: USA, 100 €/t CO2 2030



Here again, the most striking result is that we need the twin strongly pessimistic (unrealistic?) assumptions of steady CO2 intensity ratio *and* decreasing trend in gross saving to see the USA ending up as unsustainable around 2020 according to the ANS indicator.

Upper-level criticisms

At least two aspects have been left aside so far and would need further elaboration.

First, we offered no treatment of **uncertainty**, a problem which is not fully dealt with by the latest improvements in valuation methods (see, in particular, Henry and Henry, 2002; Weitzman, 2007; Heal, 2008). Section 3 elaborates more on that. Secondly, 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 (Proops et al., 1999) for imputing consumption of exhaustible resources to their final consumers, i.e. the importing countries. In fact, if scarcities were fully reflected in the prices at which these exhaustible resources are sold on international markets, there would be no reason for doing such a correction. "Dissavings" of worldwide natural resources by final consumers is already taken into account in their commercial balances and it is indeed the responsibility of the exporting country to reinvest a sufficient share of the corresponding income in other assets to warrant its own sustainability. Nevertheless, when prices are non competitive, this result is no more valid. If 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. A discrepancy appears between this country's own sustainability and its contribution to worldwide sustainability. It is precisely this discrepancy that explains the *global* inefficiency of the consumption path.

In fact, the valuation of global pollution damages raises a very similar problem. Do we aim at measuring the sustainability of well-being for one specific country, or the contribution of a given country to the *global* unsustainability? The difference between these two stances is exemplified in the treatment of CO_2 emissions by the World Bank and by Arrow et al. (2008). The World Bank attributed to each country all the cost of its CO_2 emissions, i.e. the reference situation is one where each country would have to pay for the worldwide consequences of its own emission. In the second case, Arrow et al. (2008) consider global CO_2 emissions and compute an ANS indicator by evaluating how much each country is going to be affected by these common damages. As major polluters (such the USA) are not the ones who will suffer the more from global warming, there are substantial (anti-)redistributive effects in Arrow's estimates.

Appendix 4: More on the ANS and climate change

The mission for the group was to discuss sustainability. This question is linked to the problems of intergenerational equity: if we are in a non-sustainable situation, then there is a strong case in favor of some action in favor of future generations. However, the link is not completely univocal and this also deserves some clarifications. The issue of global warming is one good case study for such a clarification,

The point of departure is what can be called a "CO_2 paradox". On the one side, the Stern review (2006) has strongly argued in favor of rapid action to avoid the long term negative consequences of these CO_2 emissions. On the other side, appendix 3 has shown that, for developed countries, it is hard to introduce CO_2 emissions in aggregate net savings in a way that clearly sends a message of unsustainability. This would require a very high valuation of these CO_2 emissions, much higher that the highest figures generally proposed in the literature and used in the Stern review. The result is all the more puzzling that both approaches use the same conceptual apparatus, i.e. long term discounting of costs and benefits. What are the true links between the two approaches?

Let's first start by recalling how consequences of climatic change are generally evaluated. As many other assesments of consequences of climate change, the Stern Review is based on an Integrated Assessment Model (IAM) of eco-environmental interactions. The model used is the PAGE model (Hope, 2006). Another model frequently used in this literature is the DICE model (Nordhaus, 2007). These models allow economic evaluations of future damages to the environment. None of them question the fact that these consequences will be negative, but both do so with a large degree of uncertainty, and this is in line with the message of the present report concerning the difficulty to assess these costs with single headline numbers. The PAGE model as used by Stern estimates damages in 2200 to lie in a bracket of -1 to -35% of Gross World Product (GWP) projected for that period. This bracket itself has been a matter of controversy, some commentators finding it too large, and some others too optimistic, putting a too low probability on the risk of major eco-environmental disruptions.

Now, the problem faced by the Stern review has been to translate these figures in terms of one single suggestive figure of how large is the problem. The point is that, even if these costs are large, they occur quite far away in the future, generally much later than 2050. In order to translate them in a number that can also make sense from the point of view of the present, they are converted in terms of an equivalent permanent loss of well-being. This actuarial equivalent falls in a bracket of -5% to -20% of GWP, large figures indeed, based on a discount rate of 1.4%.

What about policies that could help limit this loss? Knowing whether they are worth being implemented is done through actuarial calculus. An abatment policy will generally have an immediate cost, but provide returns in terms of a higher level of well-being tomorrow. We can evaluate whether the gains outweigh the costs by computing the difference between their discounted sums. This is still done over a time horizon that goes until 2200, and the choice of the discount rate becomes here crucial. The low value chosen by Stern gives strong justifications to rapid action while other authors argue in favor of much more progressive abatment policies, based on much higher values of the discount rate. For instance, Nordhaus retains a value of 4.5%.

It is on this point that has concentrated most of the debate upon the Stern report, with many authors arguing that Stern's assumption on discounting was unduly low, much lower than typical values of market interest rates, therefore giving excessive weight to concern for the future.

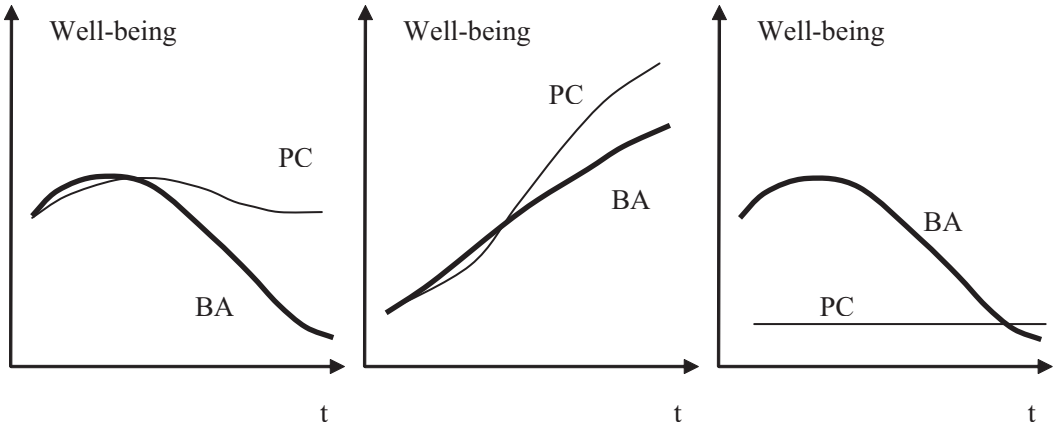
Other authors have however refined the analysis to point at additional reasons why the message of the Stern report had to be taken seriously. A review of all the arguments is provided in Heal (2008). Among these arguments, we find the fact that market rates of interest are not a good normative guide for intergenerational comparisons, especially when we think that markets do not work efficiently. There is also the idea that the problem is not that much the consequences assessed in central scenarios, but the risk of very extreme ones -the application of the precautionary principle-. There is also an idea suggested by Weitzman that the true value of the discount rate may be itself uncertain and that, in the long run, precedence must be given to the lowest of its plausible values. Another major point is that the analysis must take into account the imperfect substitutability between produced goods and natural capital, a feature that is ignored both by Stern and Nordhaus. As soon as this sustainability is imperfect, the divergent paths of production and environmental amenities lead to changes in relative prices that have to be taken into account in the cost-benefit analysis of environmental policies. Reference to a unique discount rate is no more valid, following an argument already provided long ago by Malinvaud (1953). This point has been emphasized by Guesnerie (2004) and Sterner and Persson (2007). The latter have shown that a modified version of Nordhaus's DICE model incorporating such an heterogeneity can lead to conclusions that are still more in favor of strong immediate action than the Stern review did. Further elaboration of this line of argumentation will be found in Guéant, Guesnerie et Lasry (2009, under progress).

All this may suggest that the apparent opposition between messages sent by the ANS and messages of cost-benefit analysis of immediate action only comes from different assumptions on discount rates. But this is not that clear however. As was illustrated in the simulations of section 3.3, the choice of the discount factor in the ANS or in an extended wealth framework has no incidence on the message concerning sustainability. The reason is that non-sustainability is a characteristic of the time profile of current well-being, i.e. the fact that it will decline someday below its current level: this characteristic is independent of the value of the discount rate chosen for building the ANS. This discount rate will only have an influence upon how much in advance the decline in well-being will be anticipated by the indicator (Fleurbaey, 2009).

This has a clear consequence: if the policy recommendation of Stern-like assessments depends on the discount factor, and if sustainability is a characteristic of the projected path of well-being profile that does not depend on this discount factor, this means that there cannot be any one-to-one connection between messages derived from the two approaches. How is that possible? The explanation is that a difference must indeed be made between the *observation* that a situation is not sustainable and the *prescription* of a policy change that can improve intergenerational welfare.

Let's present a few examples to show how these two notions differ. On all the following graphs, we present two paths for prospective well-being. The one represented by the thick line is the spontaneous "business-as-usual" scenario in which no policy change occurs. The thin line corresponds to a "policy change" scenario, that systematically reduces well-being in the short run, but helps improving it in the long run. We use labels BAU and PC to characterize these two scenarios.

Figure 3.7: Three different scenarios concerning sustainability and the opportunity of a policy change



The first figure on the left displays a configuration where messages concerning unsustainability and the opportunity of a policy changes will be consistent. The BAU path is a non-monotonic one. Non-sustainability is a property of this BAU scenario that can be detected sufficiently in advance with the ANS index if we fix the discount rate at a sufficiently low level. The PC scenario imposes small short run costs that allow restoring sustainability. The opportunity of moving from BAU to PC for a current planner will depend on its choice of the discount rate but, for the profiles that are proposed, we can expect that this shift will be considered desirable even for relatively high values of this discount factor. In that case, the message from the ANS index will appear consistent with the policy recommandation.

The second figure more or less reflects the reference scenario of the Stern review. Even if it projected large negative impacts of climate change on global well-being, up to minus 35% of GWP in 2200, it applied them to a path of continuing economic growth at a rate making all future generations much more affluent than we are, despite climatic change. Such an assumption is of course debatable, and this brings us back to the debate upon long run uncertainty. For instance, it may well be that climatic disorders could lead to major conflicts or social disorders that would have a strong negative effect on our physical or human capital. One can also consider that productivity extrapolations are based on biased estimates that ignore that recent productivity gains have been precisely at the expense of environmental quality. But let's assume here that this productivity assumption is correct. In this case, the BAU scenario passes the ANS sustainability test. But this does not mean that there is no justification for shifting from this BAU to the PC scenario. As in the previous scenario, the short run cost is modest, and the long run advantages of a safer climate are considerable. In that case, the PC scenario will appear desirable, at least to a planner who does not discount the future at an excessively high rate. In that case, the ANS index does not orient the policy maker in the right direction.

The last scenario is the exactly opposite -and unpleasant- one where sustainability is not warranted, but without this being a sufficient motive for action. The BAU curve has the hump shape that it already had in the first scenario. Sustainability could be restored with a PC scenario, but at a very high cost for current generations, putting everybody close to starvation.

In that case, only a very high level of collective concern for these future generation would lead the planner to engage in this PC policy.

The central scenario shows that there is not necessarily a contradiction between computing a positive rate of net adjusted savings -as the World Bank does for many countries- and being in favor of a strong immediate action in the domain of climatic change. We have seen indeed that it is hard to arrive at negative values of this ANS, even with economic valuations of CO_2 emissions close to the order of magnitudes used in the Stern report. But this raises a strong problem *vis-à-vis* policy makers. Can we let them with the only messages of an indicator that tells us that BAU is sustainable when we think that there are good reasons for action?

One answer to this dilemna could be to strengthen the sensitivity of the ANS index to environmental problems by focussing on dark scenarios, i.e. building a "precautionary" ANS with imputed prices for natural resources computed to reflect these worst cases perspectives.

The other possibility can be to treat climate change or other environmental changes as a fully separate matter. This can still be done in the extended wealth framework simply by replacing the objective function with a function that exclusively depends upon the environmental good that we want to focus on, as was done in section 3.4. But this solution may seem excessively formal. It can be more self-speaking to build *ad hoc* physical indicators comparing stocks or emissions to predefined targets. Whatever we do, this does not mean that the standard ANS becomes unuseful: it stills has its role of anticipating "global" unsustainability, but we are warned that it is not a sufficient guide for taking right decisions on specific matters, this being true for environmental as well as for non-environmental issues.

REFERENCES

Abraham K. and C. Mackie, eds. (2005), *beyond the Market Designing Nonmarket Accounts for the United States,* National Academy of Sciences, Washington D.C.

Afsa, C., Blanchet, D., Marcus, V., Mira d'Ercole, M., Pionnier, P.A., Ranuzzi, G., Rioux, L. and Shreyer, P. (2008) "Survey of existing approaches to measuring socio-economic progress", background paper for the first meeting of the CMEPSP.

Alfsen, K.H, Hass, J.L. Tao, H. and You, W. (2006) "International experiences with 'green GDP'", Report 2006/32, Statistics Norway.

Arrow, K.J., Dasgupta, P. And Mäler, K.G. (2003a) "Evaluating projects and assessing sustainable development in imperfect economies", *Environmental and Resources Economics*, 26: 647-685.

Arrow, K.J., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P., Heal, G., Levin, S., Mäler,K.G., Schneider, S., Starrett, D . and Walker, B.(2004) "Are We Consuming Too Much?", The Journal of Economic Perspectives, 18 (3): 147-172

Arrow, K.J., Dasgupta, P., Goulder, L.H., Mumford, K. and Oleson, K. (2008) "China, the U.S., and Sustainability: Perspectives Based on Comprehensive Wealth", Working Paper No. 313, Stanford Center for International Development, Stanford University.

D'Autume, A. and Schubert, K. (2008) "Le concept d'épargne véritable est-il adapté pour mesurer la durabilité du développement économique?", 12eme Colloque de l'association de Comptabilité Nationale.

Bovar, O., Demotes-Mainard, M., Dormoy, C., Gasnier, L., Marcus V. and Tregouët, B. (2008) "Les indicateurs de développement durable", *L'économie française: comptes et dossiers*, Edition 2008, INSEE.

Centre d'Analyse Stratégique (2009) « Approche économique de la biodiversité et des services liés aux écosystèmes: Contribution à la décision publique », Rapport du groupe de travail présidé par Bernard Chevassus-au-Louis, La Documentation Française.

Cobb, J. and Daly, H. (1989) *For the common good, redirecting the economy toward community, the environment and a sustainable future*, Boston, Beacon Press.

Cobb, C.W. and Cobb, J. (1994) *The green national product*, Lanham, University Press of America.

Commissariat Général au Développement Durable (2009), Une expertise de l'empreinte écologique-version provisoire, *Etudes et Documents*, n°4, mai.

Dasgupta, P. (2001) "Valuing objects and evaluating policies in imperfect economies", The Economic Journal, 111: C1-C29.

David, M. (2008) "Potentialités et limites de l'approche en termes d'indicateurs de développement durable", 12eme Colloque de l'association de Comptabilité Nationale.

Desai, M. (1994) "Greening the HDI?" in The New Economic Foundation (Eds) *Acccounting for change*, MacGillivray.

Diamond, J. (2005) *Collapse: How Societies Choose to Fail or Succeed*. New York: Viking Books (French Translation: *Effondrement. Comment les sociétés décident de leur disparition*, Gallimard, 2006).

Dietz, S. and Neumayer, E. (2004) "Genuine savings: a critical analysis of its policy-guiding value", *International Journal of Environment and Sustainable Development*, 3 (3/4): 276 - 292

Dreze, J. and Stern, N. (1990) "Policy Reform, Shadow Prices and Market Prices", *Journal of Public Economics*, Vol. 42 pp1-45.

Estes, R., Levy, M., Srebotnjak.T. and de Shrebinin, A. (2005) *2005 environmental sustainability index: benchmarking national environmental stewardship*, New Haven: Yale Center for Environmental Law and Policy.

European Environmental Agency (2009) "Comments on the draft summary report of the Commission on the Measurement of Economic Performance and Social Progress and proposals for a swift implementation of basic adjustments to the SNA", draft.

Ewing, B., Reed, A., Rizk, S., Galli, A., Wackernagel M. and Kitzes, J. (2008) Calculation Methodology for the National Footprint Accounts, 2008 Edition. Oakland: Global Footprint Network.

Fleurbaey, M. (2008) "Individual well-being and social welfare: some notes on the theory", background paper for the first meeting of the CMEPSP.

Fraumeni B. M. (2008), "Human Capital: From Indicators and Indexes to Accounts", paper presented at the OECD / Fondazione Giovanni Agnelli Workshop on the Measurement of Human Capital, 3 – 4 November Turin, Italy.

Gadrey, J. et Jany-Catrice, F. (2007) *Les nouveaux indicateurs de richesse*, 2nd edition, Repères-La Découverte.

Guéant, O., Guesnerie, R. and Lasry, J.M. (2009) "Ecological intuition versus economic 'reason'", draft.

Guesnerie, R. (2004) "Calcul économique et development durable", *La Revue Economique,* 55(3): 363-382.

Hamilton, K. (1996), "Pollution and Pollution Abatement in the National Accounts", Review of Income and Wealth, vol. 42(1), 13-33.

Hanley, N., Moffatt, I., Faichney, R. and Wilson, M. (1999) "Measuring sustainability: A time series of alternative indicators for Scotland", *Ecological Economics*, 28:55-73.

Hartwick, J.M. (1977) "Intergenerational Equity and the Investing of Rents from Exhaustible Resources", *The American Economic Review*, 67 (5): 972-974

Heal, G. (1998) "Valuing the Future: Economic Theory and Sustainability", Columbia University Press.

Heal, G. and Bengt Kriström, B. (2002) "Uncertainty and Climate Change", *Environmental and Resource Economics*, 22, 3-39.

Heal, G. (2005) "Intertemporal Welfare Economics and the Environment", *Handbook of Environmental Economics*, Volume 3. Edited by K-G Mäler and J.R. Vincent, Elsevier, Chapter 21, 1105-1145.

Heal, G. and Kriström, B. (2005) "National income and the environment", *Handbook of Environmental Economics*, Volume 3. Edited by K-G Mäler and J.R. Vincent, Elsevier, Chapter 21, 1148-1217.

Heal, G. (2008a) "What do we know about the economics of climate change?", Draft

Heal, G. (2008b) "Climate economics: a meta-review and some suggestions", NBER Working Paper 3927

Henry, C. and Henry, M. (2002) "Formalization and applications of the Precautionary Principle", Working Paper, Laboratoire d'Économetrie de l'École Polytechnique.

Henry, C. and Henry, M. (2002) "Formalization and applications of the Precautionary Principle", Working Paper, Laboratoire d'Économetrie de l'École Polytechnique.

Henry, C. (2007) "Somewhat different ways, but the same destination. On some controversial points in the Stern report". Working Paper, Laboratoire d'Économetrie de l'École Polytechnique.

Hotelling, H.J. (1931) "The Economics of Exhaustible Resources", *The Journal of Political Economy*, 39 (2): 137-175

Jollands, N., Lermit, J. and Patterson, M. (2003) "The usefulness of aggregate indicators in policy making and evaluation: a discussion with application to eco-efficiency indicators in New Zealand "

Jorgenson D. W. and B. Fraumeni (1989), "The Accumulation on Human and Nonhuman Capital, 1948-1984, in R. E. Lipsey and H. S. Tice (eds.), *The Measurement of Saving, Investment , and Wealth,* University of Chicago Press, Chicago.

Jorgenson D. W. and B. Fraumeni (1992), "The Output of the Education Sector", in Griliches Z. (ed.), *Output Measurement in the Services Sector,* University of Chicago Press, Chicago.

Keeley B. (2008), *Human Capital How what you know shapes your life,* OECD, Paris.

Lange, G.M and Wright, M. (2004) "Sustainable development in mineral economies: the example of Bostwana", *Environment and Development Economics*, 9: 485-505.

Lasso de la vega, M.C., Urrutia, A.M. (2001) "HDPI: a framework for pollution sensitive human development indicators", *Environment, Development and Sustainability* (3), 199-215.

Le Clézio, P. (2009) « L'empreinte écologique et les indicateurs du développement durable », Avis du Conseil Economique, Social et Environnemental.

McDonald, G.W. and Patterson, M.G. (2004) "Ecological Footprints and interdependencies of New Zealand regions", *Ecological Economics,* 50 (1-2): 49-67

Mäler, K.G., Aniyar, S. and Jansson, A. (2008) "Accounting for ecosystem services as a way to understand the requirements for sustainable development", *Proceedings of the National Academy of Sciences*, 105 (28): 9501–9506.

Malinvaud, E. (1953) "Capital accumulation and the efficient allocation of Resources", *Econometrica* Vol 21 (2): 233-268

Méda, D. (1999) *Qu'est-ce que la richesse?*, Aubier.

Moran, D., Wackernagel, M., Kitzes, J., Godlfinger, S. and A. Boutaud (2008), Measuring sustainable development - Nation by Nation, *Ecological Economics*, vol 64(3), 470-474.

Nordhaus, W. and Tobin, J. (1973) "Is Growth Obsolete?" in *The Measurement of Economic and Social Performance, National Bureau of Economic Research*, 1973.

Nordhaus,W.J. (2007a) "The Stern Review of the Economics of Climate Change", *Journal of Economic Literature*.

Nordhaus,W.J. (2007b) "Critical Assumptions in the Stern Review on Climate Change », *www.sciencemag.org*, 317(13)

Nourry, M. (2007) "La croissance économique est-elle un moyen de lutte contre la pollution?: Les enseignements de la courbe de Kuznets environnementale", *Revue française d'économie*, 21 (3): 137-176.

Nourry, M. (2008) "Measuring sustainable development: Some empirical evidence for France from eight alternative indicators", *Ecological Economics*, 67 (3): 441-456.

Osberg, L. and Sharpe, A. (2002) "An index of economic well-being for selected countries", *Review of Income and Wealth*, September.

Pearce, D.W. and Atkinson, G. (1993) "Capital theory and the measurement of sustainable development: an indicator of weak sustainability" , *Ecological Economics*, 8(2): 103-108.

Pezzey, J.C.V., Hanley, N., Turner, K. and Tinch, D. (2006) "Comparing augmented sustainability measures for Scotland: is there a mismatch?", *Ecological Economics*, 57: 60-74.

Proops, J.L.R., Atkinson, G., v. Schlotheim, B.F. and Simon, S. (1999) "International trade and the sustainability footprint: a practical criterion for its assessment", *Ecological Economics*, 28: 75-97.

Quinet, A. (sous la présidence de) (2008) *La valeur tutélaire du carbone*, Rapport du Centre d'Analyse Stratégique.

Randall, A (2008) "Is Australia on a sustainability path? Interpreting the clues », *The Australian Journal of Agricultural and Resource Economics,* 52, pp. 77–95

Repetto, R., Malgrath W., Wells, M., Beer C. and F. Rossini (1989), Wasting Assets: Natural Resources in the National Income Accounts, *World Resources Institute*.

Samuelson, P.A. (1961), "The Evaluation of Social Income: Capital formation and Wealth", The Theory of Capital, Proceedings of an IEA Conference, Lutz and Hague, eds. (New York: St. Martin's Press).

Stern, N. (2006) "The Economics of Climate Change: The Stern Review", H.M. Treasury, U.K.

Stern , N. (2008) "The economics of climate change", *American Economic review*, Papers and Proceedings, 2008

Sterner, T. and M. Persson, M. (2007) "An Even Sterner Review: Introducing Relative Prices into the Discounting Debate", Discussion Paper, Resources for the Future.

Sukhdev P. (ed.) (2008), The Economic of Ecosystems and Biodiversity, Interim Report, European Community, 68 p.

Talberth, J. Cobb, C. and Slattery, N. (2006) *The Genuine Progress Indicator 2006: a tool for sustainable development*, Redefining Progress, Oakland CA, (available at www.rprogress.org)

UNECE/OECD/Eurostat (2008) "Report on measuring sustainable development: statistics for sustainable development, commonalities between current practice an theory", Working Paper ECE/ CES/2008/29.

United Nations (2003), Handbook of National Accounting: Integrated Environmental and Economic Accounting 2003.

United Nations (2005) Living Beyond Our Means: Natural Assets and Human Well-being

Vanoli, A. (1995) "Reflections on environmental accounting issues", *Review of income and wealth*, Series 41, Number 2.

Vanoli A. (2005) *A History of National Accounting*, IOS Press.

Viveret, P. (2002) *Reconsidérer la richesse*, Editions de L'Aube

Wackernagel, M. and Rees, W. (1995) *Our ecological footprint: reducing human impact on the earth*, New society publishers, The New Catalyst Bioregional Series, Gabriola Island,BC.

Weber, J.L. (2008) "Comptabilité des ecosystèmes et de leurs services", 12eme Colloque de l'Association de Comptabilité Nationale.

Weitzman, M.L. (1976), "On the Welfare Significance of National Product in a Dynamic Economy", *Quarterly Journal of Economics* 90, 156-162.

Weitzman, M. (2007). "The Stern Review of the Economics of Climate Change", *Journal of Economic Literature*, .

World Bank (2006) *Where is the Wealth of Nations? Measuring Capital in the 21st Century*. The World Bank, Washington D.C.