The Human Development Index: a search for a measure of human values

Makiko Ito Harrison

The London School of Economics and Political Science

Ph.D. in Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract of the thesis

The thesis investigates methods of evaluating indexes that measure concepts of human values. My understanding of indexes, especially on how they relate to the real world and concepts (that are the objectives of the measurement), is influenced by my study of literature on models used in economic and in physics.

We learn from this study of models the following:

(1) regularities described in theories do not represent real world phenomena, which consist of many different forces acting simultaneously;

(2) but such regularities are true in models, because they describe specific conditions under which regularities in nature are displayed;

(3) there are more than one model that can represent the same phenomenon depending on which particular aspect of the phenomenon to focus on; and

(4) the success of a model has to be evaluated partly by criteria that are independent from theoretical ones.

Since the role indexes play in relation to real world and concepts are similar to the role models play in relation to theories, I have applied the above knowledge to propose the following three criteria to evaluate successful indexes:

(1) Purpose-dependent criteria: criteria that are based on particular motivations of the measurement project;

(2) Theory-dependent criteria: criteria that are reflected in the theories that expressly or implicitly guide the development of the project of measurement; and

(3) Conditions-dependent criteria: criteria that are based on the conditions under which the index measures what it is designed to measure.

I apply these three criteria of successful indexes to examine two projects of measuring human values, one called the Human Development Index developed by the United Nations Development Programme and the other called the Life Satisfaction Indicator developed by an officer at the Economic Planning Agency in Japan. Among the findings from the examination of those two indexes are that they can be the products of a mixture of concerns that include convenience, conventions, practicality, politics and consistency with relevant theories, and some of these concerns may conflict with each other. Another important finding is that because there are many assumptions made and simplifications applied in order to choose a quantitative representation of a human value, the application of the measure is limited. I conclude that both in using and in evaluating indexes of human values, it is important that we are aware of such limitations, so that we can more effectively know both how to avoid misusing the indexes and how to improve them over time.

Contents

Abstract of Thesis 2

Table of Contents 4

Acknowledgements 10

Chapter one: Models, indexes and the real world 12

1-1: Accounts of successful models 13

1-2: Models versus indexes 18

1-3: My account of successful indexes 25

1-3-1: Purpose-dependent criteria of successful indexes 26

1-3-2: Theory-dependent criteria of successful indexes 28

1-3-3: Condition-dependent criteria of successful indexes 29

Chapter two: Purpose-dependent criteria for the HDI and the LSI 31

2-1: Background of the construction of the Human Development Index and the Life Satisfaction Indicator 31

2-2: The method of measurement 34

2-2-1: The HDI 34

2-2-2: The LSI 36

2-3: Purpose-dependent criteria of the index 45

2-3-1: The HDI 45

2-3-1-1: Universality 46

2-3-1-2: Differentiation from a GDP measure and more discrete orderings of countries 52

2-3-1-3: Policy usefulness 68

2-3-1-4: Comparability across space 69

2-3-1-5: Comparability across time 77

2-3-2: The LSI 80

2-3-2-1: Universality 80

2-3-2-2: Discrimination from a GDP measure 81

2-3-2-3: Policy usefulness 82

2-3-2-4: Comparability across space 83

2-3-2-5: Comparability across time 84

Chapter three: Theory-dependent criteria for the HDI and the LSI 85

3-1: Relationship between Amartya Sen's capability theory and the HDI 86

3-1-1: Sen's capability approach to evaluating well-being 86

3-1-2: A historical relationship 91

3-1-3: Relative aspects of well-being and absolute criteria for the measurement 93

3-1-4: The commodity-based approach and its limit 96

3-1-5: The capability approach and its ability to reveal relative necessities of individuals 98

3-1-5-1: Problems in implementing the capability approach to the measurement of well-being 99

3-1-5-2: Problems in implementing a commodity-based approach to the measurement of well-being 107

3-1-5-3: The HDI and the practical problems 110

3-1-6: Capability approach, relative aspects of well-being and the HDI 112

3-1-6-1: Life expectancy at birth 112

3-1-6-1-1: Comparison to other alternatives 114

3-1-6-1-2: Identification of relative necessities 122

3-1-6-2: Adult literacy rate and the school enrolment ratio 124

3-1-6-2-1: Comparison to other alternatives 125

3-1-6-2-2: Identification of relative necessities 128

3-1-6-3: The real GDP per capita in purchasing power parity dollars 131

3-1-6-4: Summary of the subsection 139

3-2: The LSI and theory 142

3-2-1: The general relationship of Bentham's utility theory to the LSI 143

3-2-2: The LSI and Bentham's method of measuring pleasures and pains 144

Chapter four: Condition-dependent criteria for the HDI and the LSI 149

4-1: Quality of data and analytic problems with the approaches used in the HDI and the LSI 151

4-1-1: Quality of data 151

4-1-2: Analytic problems with using the composite approach to the HDI 153

4-1-2-1: The weighting of the dimensions 154

4-1-2-2: The HDI and the policy trade-offs 160

4-1-2-3: Limitation in the elements included in the HDI 164

4-1-3: Analytic problems with the LSI index 166

4-2: Conditions under which the HDI and the LSI can reasonably be used for policy purposes 169

4-2-1: The HDI 170

4-2-2: The LSI 172

Conclusion 177

Postscript 181

Appendix 186

References 187

Illustrations

Figure 1.1 How successful models and successful indexes are made.....27

2.1 The pictorial representation of the structure and normalised factors reflected in the HDI.....36

2.2 The pictorial representation of the structure and factors reflected in the LSI 43

Table 2.1 The Life Satisfaction Indicator: a time-series result between 1975-93.....44

2.2 Country rankings according to 'mean years of schooling' (among countries with a 99% adult literacy rate).....56

2.3 Country rankings according to the old 'educational attainment index,' 1994 Report (among countries with a 99% adult literacy rate).....58

2.4 Country rankings according to the new 'educational attainment index,' 1995 Report (among countries with a 99% adult literacy rate).....59

2.5 Educational attainment index rankings and GDP index rankings based on the 1991 Human Development Report (Selected comparisons).....63

2.6 Educational attainment index rankings and the GDP index rankings based on the 1995 Human Development Report (Selected comparisons).....64

2.7 Country rankings according to different educational indexes (Selected comparisons) 66

2.8 HDI rankings of the lowest 20 countries from the 1998 Human Development Report 73

Acknowledgements

This dissertation has enormously benefited from many individuals' good will and support, every one of which was important for me and for the completion of this work.

First I would like to thank my parents. Without their understanding, patience and love I would never even have dreamt of writing a dissertation, and I am grateful to them for the opportunities they have given me.

My sincerest gratitude goes to my two supervisors at the London School of Economics and Political Science, Professor Nancy Cartwright and Professor Mary S. Morgan, both of whom have given me wonderful opportunities to challenge and expand my ability, and have provided guidance and encouragement both conscientiously and with warmth.

I also would like to thank my previous supervisors: Professor Sahoko Kaji at Keio University, who has provided both intellectual and moral support since my undergraduate days; Dr. Peter Holmes at the University of Sussex, who greatly encouraged me to pursue my interest in economic methodology and; Professor Daniel Hausman (currently at the University of Wisconsin-Madison) and Dr. Thomas Uebel (currently at the University of Manchester), who patiently guided me to philosophy during my two years at the London School of Economics and Political Science.

I also owe debt for their intellectual inputs and support to a number of people, including Mr. Max Steuer, Dr. Carl Hoefer, Dr. Richard Bradley and Dr. Joseph Berkovitz at the London School of Economics and Political Science; Professor Kotaro Suzumura at Hitotsubashi University; Professor Eishi Hosoda at Keio University; Professor Takuo Dome at Osaka University; Dr. Marcel Boumans and Dr. Harro Maas at the University of Amsterdam; Mr. Kosei Fukuda at the Economic Planning

Agency; Ms. Frances Cairncross at the Economist; and all the people I interacted with in the Centre for Philosophy of Natural and Social Science at the London School of Economics and Political Science, especially Ms. Cynthia Ma, Ms. Antigone Nounou, Mr. Julian Reiss, Mr. Sang Wo Yi, Mr. George Zouros, Dr. Marco Del Sata, Dr. Sebastiano Bavetta (now at University of Palermo), Dr. Francesco Guala (now at the University of Exeter), and Dr. Johannes Hunger (now at the University of Heidelberg).

My special thanks for their additional assistance go to Ms. Frances Townshend and Ms. Sadaf Haviz from the philosophy department, Ms. Dorota Rejman, Ms. Klementina Balint, and Ms. Francisca Ahanonu from the Centre for Philosophy of Natural and Social Science, and Ms. Cathy Van Brocklin in Washington, D.C.

Finally, I would like to give my most special thanks to my husband, Donald, who not only gave me enormous energy and hope to complete this dissertation, but also helped me edit and improve my previous drafts with patience, conscience and love which went beyond my imagination.

I am very grateful for where I stand now, and I thank God for His love and for having made it always so obvious to me.

Chapter one: Models, indexes and the real world

This dissertation is concerned with attempts to quantify ideas about human values or well-being primarily through the use of indexes developed for this purpose. Many may object to such attempts on moral, practical or metaphysical grounds. They may say: it is not 'right' to quantify human values or well-being; it is not possible with our knowledge to quantify them; or even in principle such values or well-being cannot be quantified.

The purpose of this dissertation is not to discuss the status of such objections, but to investigate ways to evaluate efforts to quantify concepts that have to do with human well-being. My own metaphysical position, which will become clear in the dissertation, is that it is not possible to find *the* index to quantify human values or well-being, but we can find various measurement instruments each of which can be useful and informative for different purposes for which we require a numerical representation of human values. It is the criteria of successful indexes that are relative to these different purposes that I propose and defend.

This first chapter consists of three parts. The first part briefly discusses different views of successful scientific models in order to find out the ways in which these models relate to real world phenomena. I introduce these different views on models because despite the differences in their principal purposes, models and indexes are similar in the way they relate to the real world. Three works on models I use are: Nancy Cartwright's views on theories, models and phenomena and her account of causation as fundamentally involving 'capacities' in the world rather than 'nomological laws'; Mary Morgan's account of the range of purposes for satisfactory models developed by early econometricians; and Marcel Boumans' account of the integration of different 'ingredients' to realise satisfactory models in economics. The

second part of the chapter explains why understandings of the relationship between models and the real world are not only relevant but also important in examining successful indexes. The last part introduces my own account of successful indexes. Here, I provide differences and similarities between my account of successful indexes and accounts of successful models.

1-1: Accounts of successful models

In her latest book, Nancy Cartwright expresses her views that real world is much messier and unsystematic than many scientists assume it to be, and that regularities can only be observed under an artificially shielded environment:

> This book supposes that, as appearances suggest, we live in a dappled world, a world rich in different things, with different natures, behaving in different ways. ---- For all we know, most of what occurs in nature occurs by hap, subject to no law at all. What happens is more like an outcome of negotiation between domains than the logical consequence of a system of order. The dappled world is what, for most part, comes naturally: regimented behaviour results from good engineering.[1]

Consistent with the view quoted above, Cartwright rejects the 'theoretical' in favour of the 'phenomenological' approach to scientific investigations, explaining as follows:

> In modern physics, and I think in other exact sciences as well, phenomenological laws are meant to describe, and they often succeed reasonably well. But fundamental equations are meant to explain, and paradoxically enough the cost of explanatory power is descriptive adequacy. Really powerful explanatory laws of the sort found in theoretical physics do not state the truth.[2]

Since such explanatory laws do not state truths about the world, models that try to link fundamental laws directly to reality are unsuccessful models. Cartwright argues instead that

---

[1] Cartwright (1999), p. 1.

> [t]he route from theory to reality is from theory to model, and then from model to phenomenological law. The phenomenological laws are indeed true of the objects in reality – or might be; but the fundamental laws are true only of objects in the model.[3]

According to Cartwright's account of models, models imitate or represent forces, causal relationships, or what Cartwright calls 'capacities' that we believe exist in the world. Cartwright explains 'capacities' as follows:

> The generic causal claims of science are not reports of regularities but rather ascriptions of capacities, capacities to make things happen, case by case. 'Aspirins relieve headaches.' This does not say that aspirins always relieve headaches, or always do so if the rest of the world is arranged in a particularly felicitous way, or that they relieve headaches most of the time, or more often than not. Rather, it says that aspirins have the capacity to relieve headaches, a relatively enduring and stable capacity that they carry with them from situation to situation; a capacity which may if circumstances are right reveal itself by producing a regularity, but which is just as surely seen in one good single case.[4]

Cartwright views successful models as representations of such 'capacities':

> My claim then is that models serve as blueprints for nomological machines. There are three separate theses involved in this claim. The first is that the general scientific knowledge that we use to construct models is not knowledge of laws.[5]

> [Sets of probability and causal laws] need a socio-economic machine [which is a type of nomological machine] to generate them. The view of course only makes sense if the kind of knowledge that we need to understand the operation of a socio-economic machine is not itself more knowledge of 'deeper' probabilistic and causal laws. And this is just the claim I argue: the knowledge we need here is knowledge not of laws but of capacities.[6]

Let me summarise Cartwright's arguments that are relevant for the purpose of my thesis:

---

2 Cartwright (1983), p. 3.

3 Ibid., p. 4.

4 Cartwright (1989), p. 3.

5 Cartwright (1997), p. 1.

6 Cartwright (forthcoming), p. 7.

1. Generic causal claims of science are reports not of regularities but of capacities that we believe exist in the world.
2. Models link theories (that consist of generic causal claims relating to capacities) and real world.
3. Specifically, models serve as blueprints for nomological machines that generate regularities described in theories.

If we accept Cartwright's account, there is no 'right' theory for a phenomenon, because real world is made up of a combination of forces acting simultaneously, some revealing their capacities, some not, while a theory can only focus on a particular set of such forces assuming that all of them fully display their capacities. Neither is there a 'right' model to describe a phenomenon in Cartwright's account. Any model can be successful as long as it provides blueprints for a nomological machine that, if followed properly, would generate expected regularities described in a theory. For example, in an attempt to represent in a model the reasons for changes in the level of unemployment, many different sets of economic and social variables and relationships among them can be chosen. The choice depends on one's different hypotheses about the labour market, the behaviours of firms and workers, and the economy in general, as well as the level (micro or macro, for example) at which one wants to find a relationship between unemployment and other phenomena.

So how do we distinguish successful models from non-successful ones? One criterion that derives from Cartwright's account of models is related to the way we use them: that is, models (and theoretical claims they support) are used (to explain a phenomenon or to intervene in the world) only when the circumstances resemble the blueprints. This is because, according to Cartwright, a successful model represents capacities in the world, but because they are capacities, what the model tells us is true

only when a set of conditions under which a relevant capacity can be actualised are met in the real world.

But this criterion does not help us to choose among competing models that explain the same phenomenon. On the latter type of criterion, Cartwright talks about the role of 'bridge principles' in mathematising phenomena in the world so that the phenomena fit into theory, and she also explains that various *purposes* for building models determine the final structure of models. She writes:

> In physics it is usual to give alternative theoretical treatments of the same phenomenon. We construct different models for different purposes, with different equations to describe them.[7]

Cartwright, therefore, does not focus on the identification and classification of factors other than theoretical ones to determine a successful model.[8] The accounts of successful models by Morgan and Boumans provide broader analyses for this purpose.[9]

Mary Morgan, like Cartwright, views a model as a bridge between an abstract theory and observable data. She examines a particular model, therefore, with respect to its relationship with both sides of this bridge: the model's consistency with a relevant theory as well as its power to explain empirical data. Morgan argues that early econometricians' focuses were on finding models that could satisfactorily explain empirical data, rather than on falsifying or verifying fundamental economic theories.[10] She finds that models were examined relative to one or more of the following purposes for which models are made:

---

[7] Cartwright (1983), p. 11.

[8] Ibid.

[9] Morgan (1988) and Boumans (1999a).

[10] Morgan (1988).

1. To measure theoretical laws, that is, to satisfy certain theoretical requirements.
2. To explain (or describe) the observed data, that is, to fit the observed data (statistical or historical).
3. To be useful for policy, that is, to allow the exploration of policy options or permit predictions about future values.
4. To explore or develop theory, that is, to expose unsuspected relationships or develop the detail of relationships.
5. To verify or reject theory, that is, to be satisfactory or not over a range of economic, statistical and other criteria.

Morgan's study from 1988 points out that early econometricians did not aim at finding models to provide *the* true mathematical representation of a theory, but built models for a much wider range of purposes consistent with the range of purposes noted above.

Marcel Boumans' view of models is slightly different from those of Cartwright and Morgan.[11] He views models, not as bridges between data and theories, but as outcomes of a mixture of ingredients that include data and theory. Examples of such 'ingredients' described in his paper are:

(1) Theoretical notions
(2) Mathematical concepts
(3) Mathematical techniques
(4) Stylised facts

[11] Boumans (1999a).

(5) Empirical data

(6) Policy views

(7) Analogies

(8) Metaphors

Boumans claims that models that are satisfactory are the ones that integrate these ingredients in ways that meet certain criteria.[12] He provides the following examples of such criteria:

1. Satisfaction of theoretical requirements
2. Satisfaction of mathematical requirements
3. Satisfaction of statistical requirements
4. Usefulness for policy

Boumans' account of satisfactory models thus identifies various concerns other than theoretical ones that enter into the process of constructing a model. In his account, particular theories or theoretical notions for which we build models are not necessarily the dominant concern compared to various other concerns, as long as the integration of the range of concerns satisfies one or more of the *a priori* criteria.

1-2: Models versus indexes

As noted above, according to Morgan, models are made for various purposes, one of which is to measure phenomena that are defined in terms of theories.[13] Indexes, on the other hand, are developed, at least primarily, to measure social, economic and other phenomena of interest. Because indexes are made for measuring certain states of the world, they are not primarily used to test the causal relationships affecting such

[12] Ibid.

[13] Morgan (1988).

states, whereas models often are constructed out of hypothesis we have (or would like to test) on such relationships.

Despite the differences between the roles of models and the roles of indexes, the relationship between models and the real world is somewhat similar to that between indexes and the real world.

Firstly, similar to the case of models versus real world phenomena, there is no one-to-one relationship between a concept and any instrument to quantify the concept, except in the very limited circumstances when the concept represents a single natural quantity.[14] In other words, similar to the case of models versus real world phenomena, more than one index can be accepted as measuring a concept. Let us consider here the ideas of Hasok Chang, which illustrate this principle.[15] According to Chang, even concepts such as 'temperature,' 'length' or 'weight,' do not represent single natural quantities, even if we believe that there is 'real' temperature, 'real' length, 'real' weight etc. that exist independently from the instruments that measure those concepts. Chang shows that historically, there have been many different instruments used to measure the 'real temperature' or the 'real degree of heat.' He explains, however, a fundamental problem in confirming the reliability of such an instrument. In order to show, for example, that the mercury thermometer is reliable, we need to know that mercury expands uniformly (or linearly) with temperature between certain 'fixed points' (and beyond them), one of which is marked when the water freezes and the other when the water boils. To test the uniform expansion of mercury, we need to find out the relationship between the height of the column of mercury (h) and the

---

[14] An example of this latter case is a concept that refers to mathematical fractions (for example, one-half), where there is a one-to-one relationship between the concept and its measurement.

[15] See Hasok Chang (1996) on his paper titled 'Spirit, Air and Quicksilver: the search for the "Real" Scale of Temperature.'

temperature (T). That is, we need to know the function h(T) by making observations about the correlation between the two variables. But in order to do so, we have to have a reliable measure of temperature (T) in advance. There is accordingly circularity in the testing procedure.

Chang introduces various attempts made by scientists in the 19th century to overcome this circularity problem and their scepticism about the accuracy of the mercury thermometer. One of the 19th century scientists, Henri Victor Regnault, succeeded in providing a solution to the circularity problem, based on the idea of 'comparability.' Chang defines this 'comparability' idea as follows: if a thermometer were to give us the true temperature, it should at least always give us the same reading under the same circumstances; and if a type of thermometer is believed to be accurate, all thermometers of that type should agree in their readings. Under this criterion of comparability, the spirit thermometer is not accurate because spirit expands differently depending on its strength, so that spirit thermometers (made out of spirits with different degrees of strength) are not comparable. Regnault also rejected the mercury thermometer according to this criterion because the readings of mercury thermometers made with different types of glass, or even the same type of glass which had undergone different thermal treatments, would not agree with each other. Gas thermometers also did not satisfy the comparability requirement because Regnault found that the results of thermometers made with different gases were not always comparable. Regnault concluded that thermometers made with atmospheric air gave the best results in terms of their comparability because the results of air thermometers with different densities were quite comparable with each other, certainly better than those of the mercury thermometers.

'Comparability' is an alternative to the need to prove that expansions of air, mercury, spirit or gas are uniform in order to accept thermometers made with those materials, because application of the criterion of comparability eliminates the need to prove the linear expansion of those materials. All types of thermometers except for the air thermometer were unsatisfactory because they failed to pass the comparability requirement, and the 'comparability' evidence provided sufficient reason for Regnault to support the air thermometer. We could say that Regnault regarded comparability as a necessary but not sufficient condition for choosing a reliable thermometer, assuming that such a thing existed.

For our purposes, it is important to recognise that in order to select one means of measuring even a relatively straight forward concept like 'temperature,' it was necessary to introduce a new criterion, in this case, 'comparability.' This new criterion may be applied to certain basic concepts such as temperature, length, weight and time.

On the other hand, there are other concepts that we do not have any conventional method of measuring, nor do we know conditions under which each competing measure is useful and reliable. Concepts such as 'human development' and 'general level of satisfaction,' which I will be discussing in the later chapters, fall into this latter category. Chang's study on the measurement of temperature shows that most concepts, even ones that are conventionally associated with quantities (such as temperature and length), do not represent a single natural quantity, and therefore do not have one-to-one relationship to instruments or indexes that measure those concepts. Chang concludes that in order to choose among competing instruments, we need criteria independent from concepts themselves, such as the criterion of comparability. This conclusion is even clearer in the case of concepts such as 'human development' and 'general level of satisfaction.'

A second similarity between models and indices in their relationship to real world involves the fact that some indexes are made out of constituents that are considered to be causally related to the phenomena that the indexes aim to measure, just as models typically include factors intended to express causal relationships in the world. As we will see in the later chapters, the Life Satisfaction Indicator (LSI) consists of variables that are believed to cause the object of measurement, that is, the general level of satisfaction in Japan. The general level of satisfaction is measured by adding up factors that are considered to cause the phenomenon, and each factor is supposed to have a different degree of contribution to the occurrence of the phenomenon. Such degrees are represented as correlation co-efficients. The equation used to measure the general level of satisfaction is exactly the same as the equation that explains how the satisfaction is caused.

Finally, analogous to the case of models versus real world phenomena in Cartwright's analysis, an index for a particular abstract concept measures phenomena in the world successfully only when a certain set of conditions is met. Cartwright's analysis says that models represent causal relationship as capacities in nature, and they describe specific contexts in which those capacities display themselves in a lawful manner. This view of models leads us to believe that when the circumstances are not 'right,' a given capacity can be there but will not give rise to regular behaviour. Models will accordingly be able to explain the causal structure of phenomena in the world only when particular conditions are met.

Similarly, an index for a particular concept provides a numerical representation of the concept, but because there is no one-to-one relationship between a number and a concept, certain data used in the index does not necessarily reveal information we are seeking. As with models, in order for indexes to successfully represent abstract concepts, certain conditions have to be met. For example, literacy rates reveal information about the population's actual capability to acquire knowledge only when there are books to read and guidance available to systematise the inflow of new knowledge. The index may have the ability to measure an abstract concept such as literacy, but whether or not the index actually displays the ability depends on the conditions under which the measurement instrument is used.

Understanding these similarities between models and indexes, we can see three main benefits of recognising similarities between the scientific models and indexes for measuring phenomena in the world. Firstly, understanding of Cartwright's view of models leads us to accept that a model is true of causal relationships in the world only when the conditions in which the model can exhibit its capacities are met. Accordingly when we use indexes we also need to examine the conditions under which we use the indexes. In practice, we need to be aware of the fact that the data used in the index reveal information we are looking for only when a certain set of conditions are met. Our knowledge of the limitation of models thus helps us to understand the relationship between indexes and the world.

Secondly, the awareness that there are many different ways of modelling even the same theoretical concept, which is brought out by Cartwright's studies, leads us to accept that there is no one-to-one relationship between a model and a concept. As a consequence, a satisfactory model is chosen based on various concerns, which are independent from the theory or theoretical concept itself, such as the purposes for

which the models will be used or whether the model meets mathematical requirements. The latter point was reinforced by Morgan and Boumans. I claim that a successful index for an abstract concept is also selected for reasons that are independent from the definition of the concept because there is no one-to-one relationship between the concept and its numerical representation.

Chang's study the on the measurement of temperature discussed above illustrates this point, and his approach to selecting the most reliable measurement of temperature can be applied to other concepts that are not defined in terms of single natural quantities. I apply this approach to the Human Development Index (HDI), which has been developed to measure a concept of 'human development.' As detailed below, the HDI is expressed as a function of three variables, that is, HDI = f (A, B, C). There is circularity in testing the HDI, which is similar to the case of testing thermometers, because in order to test how effectively the index measures human development using regression models, we need to first have a reliable measure of human development. One way to overcome this circularity and to select the most successful HDI is to introduce Chang's criterion of 'comparability': to choose the index that is most stable under the same circumstances, and to choose a type of index that produces consistent results across different versions of the same type.

A third main benefit of recognising similarities between scientific models and indexes involves the importance of theory in testing the success of an index. Specifically, I believe that although the link between theory and index is inherently not as strong as in the case of theory and model, theory and theoretical properties may still play an effective role in testing whether or not an index is successful, which is the

case in the models' relationship to theories shown in Morgan and in Boumans.[16] In particular, I will show in Chapter three the important role that theories play in the HDI and the LSI.

1-3: My account of successful indexes

The account of successful indexes introduced in this section is my own and is based on the role and the nature of indexes that we learn from our understandings of models. My account of successful indexes consists of criteria that are relative to the motivations and purposes of a particular quantification project by a particular index maker. I hold a relativistic account of indexes because of my view, consistent with the above discussion, that there is no one-to-one relationship between a concept and any instrument for its quantification when the concept does not represent a single natural quantity. In other words, I believe that there is more than one non-trivially different index that can be accepted as measuring, for example, 'human development' or 'satisfaction' with our current knowledge. I also conclude that each index that has met some appropriate standard as a measure of 'human development' has a 'capacity' to measure the concept.

If one holds this particular view about the quantification of abstract concepts, it is not particularly interesting to ask the questions whether or not the HDI measures human development, or whether or not the Life Satisfaction Indicator LSI describes or explains the general level of satisfaction in Japan. The HDI may be just one of many indexes that measure human development. The same is likely to be true in the case of the LSI as a measure of the general satisfaction level in Japan. Also, answers to such questions do not give us instructions on how to use these indexes in order for them to

[16] Morgan (1988) and Boumans (1999a).

exhibit their capacity to measure 'human development' or the 'general satisfaction level.' The questions to be asked, instead, are 'Do the HDI and the LSI satisfy the main motivations behind each quantification project?' and 'When or under what conditions do these indexes exhibit their capacity to measure or to explain the phenomena in question?'

As detailed below, in responding to these questions I apply three types of criteria, specifically purpose-dependent criteria, theory-dependent criteria, and conditions-dependent criteria.

1-3-1: Purpose-dependent criteria of successful indexes

My account of the purpose-dependent criteria of successful indexes is similar to Morgan's account of satisfactory models introduced earlier in this chapter. As noted, Morgan emphasises that the early econometricians' interests were not always to prove fundamental economic theory true. Rather, they were interested in models that serve less fundamental and more practical goals. My account similarly consists of criteria that are relative to motivations and purposes. I call a set of such criteria 'purpose-dependent criteria of successful indexes.'

However, one difference between the two accounts is that my account takes motivations to be much more project specific and to include considerations such as who does the quantification and what are the backgrounds (histories) of a project. As developed in the later chapters, for example, the principal motivation behind the HDI was the United Nations Development Programme's interests in developing concept of human development as an alternative to the use of the Gross Domestic Product (GDP), which had proved in many cases to be unsuccessful in improving human good. In the case of the LSI, the motivation involved the index-maker Fukuda

Kousei's attempt to resolve the contradictory relationship between changes in the levels of GDP and the levels of satisfaction with life in general revealed in a public opinion poll. Fukuda also aimed to identify policies that would improve the general level of satisfaction in Japan.

My view of successful indexes is also similar to Boumans' account of satisfactory models, because both are based on awareness that there are many different ways of building successful models (or indexes) for a single concept or a particular theory. As a consequence, both Boumans' account and mine believe that evaluations depend on the purposes for which scientists and policy-makers build models (or indexes). The difference between Boumans' account and my view relates to the way in which various concerns relate to purposes. In Boumans' account, concerns have to be integrated in ways that serve the purpose for which a particular model is made. (Picture A). In my view, the concerns that permit us to evaluate indexes are derived from project specific purposes. (Picture B).

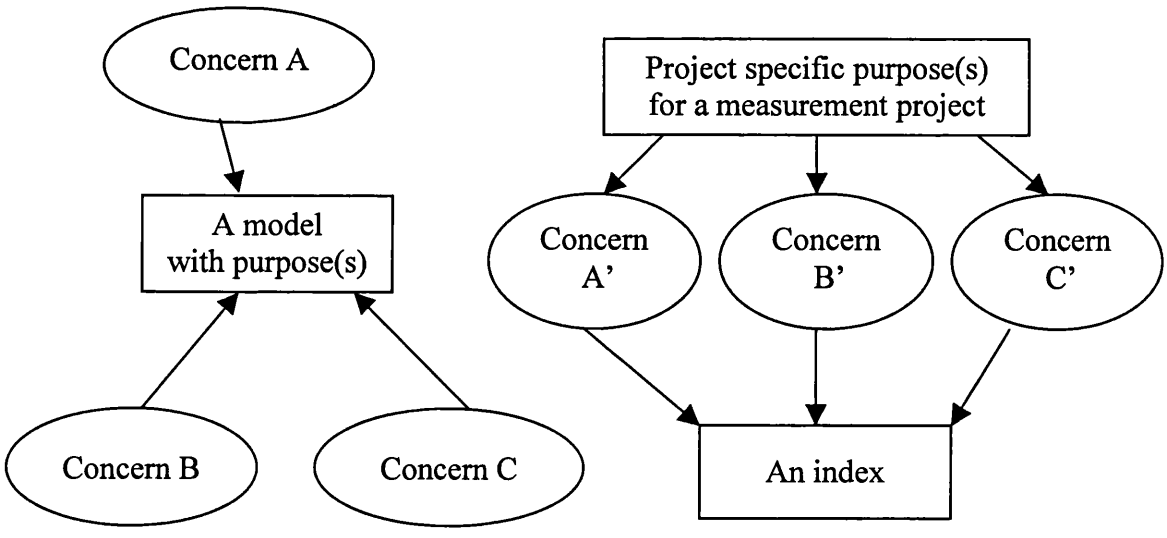

Figure 1.1 How successful models and successful indexes are made

1-3-2: Theory-dependent criteria of successful indexes

The stronger emphasis I put on particular motivations of quantification of indexes compared to Morgan's and Boumans' account of models may arise from a fundamental difference between models and indexes, that is, that the latter are explicitly and primarily made for the purpose of measurement while the former may have different purposes often having to do with theoretical investigations. When we talk about models, they normally embody theoretical considerations – models are made having some theories or of theoretical concepts in mind. In Boumans' account, theories may make a minimum contribution to the creation of a model. But theories are still necessary to make models, because they are models of a reality that is captured by theories that define concepts, or by theories that imply certain regularities in the world.

By contrast, because of the primary purpose indexes have, and because objects of the measurement (concepts or ideas) can be chosen without a particular theory behind them, the relationship between theories and indexes is not clear. Rather, how important theory is in constructing an index is largely dependent on the particular motivation of the project. This can be seen from the different ways in which theories influence the two indexes – the HDI and the LSI – upon which I focus in this dissertation. As explained in the following chapters, the HDI has a strong link to *the capability approach to well-being* developed by Amartya Sen; Sen's ideas were utilised from the beginning of the UNDP's effort to develop an alternative policy objective to higher income and to produce an index to measure the alternative concept. Therefore, one important aspect of the capability theory by Sen – that is, if used for the measurement of well-being, 'capabilities' can identify necessities of

individuals that are relative to their personal characters (such as age, health conditions) – can be effectively used for evaluating the HDI.

The LSI, on the other hand, was less obviously influenced by any particular theory. To some extent this index was influenced by Bentham's idea that the improvement of the general level of pleasure in society should be the ultimate policy objective of a government. However, the link between the LSI and Bentham's theory remains at a very general level; that is, the index measures the aggregation of some positive mental state (satisfaction) and the index shows the type of policies that will improve the level of the general level of satisfaction. The theory does not play a role, for example, in choosing between different kinds of data on the degree of happiness, satisfaction or pleasure. But the limited application of Bentham's theory in the LSI does not form a basis to criticise the LSI, because nowhere is there an indication that the particular notion of 'satisfaction' that the index-maker chose to measure is the mental state defined by Bentham[17]. (This circumstance is unlike the case of the HDI, where the notion of 'capability' used in the Human Development Report, which publishes the index, expressly comes from Sen's theory.)

1-3-3: Condition-dependent criteria of successful indexes

Under my view, whether an index is successful in revealing information of interest to us depends on whether a certain set of conditions are met where the measurement is conducted. Therefore, in my account of successful indexes it is also important to check whether an index is applied in circumstances where it can exercise its ability. I call this checking procedure 'condition-dependent criteria of successful

indexes,' which, as explained in section 1-2, is influenced by Cartwright's view on models as describing the exercise of capacities in the world.

In practice, an index having the conditional ability to convey useful information would imply the following: (1) empirical data used in the index (for example *adult literacy rates* in the HDI) are capable of revealing such useful information (for example, about 'people's ability to acquire knowledge'), but only in countries that satisfy a certain set of conditions; and (2) the ways in which different variables are combined and the weights assigned to them are governed by how the index is used and whether a set of conditions are met where the index is applied.

The 'condition-dependent criteria of successful indexes' does not appear in Boumans' analysis.[18] This may be because he regards such a problem as a problem for model-users rather than for model-builders. But in the case of indexes this distinction is not possible, primarily because the purpose of indexes is to measure, and they cannot measure without fitting data into indexes. The selection of data sets for each sub-dimension of the HDI, for example, is an important part of the measurement of human development.

With this general background, I now proceed in Chapter two to the first of the three criteria to be reviewed, the purpose-dependent criteria for successful indexes as applied to the HDI and the LSI.

---

[17] This does not mean that there is no link between the LSI and Bentham's theory at a more concrete level. In fact, Fukuda identifies an aspect of the general satisfaction data he uses that is consistent with Bentham's definition of pleasure. I will discuss such connections in Chapter three.

[18] Ibid.

Chapter two: Purpose-dependent criteria for the HDI and the LSI

In this chapter, I will test two indexes – the Human Development Index (HDI) and the Life Satisfaction Indicator (LSI) – according to one of the three kinds of criteria of successful indexes discussed in the previous chapter, specifically purpose-dependent criteria. I have chosen these two indexes because, despite the fact that they share the broad aim of measuring a concept of well-being that is richer than mere growth in income, the two indexes have very different structures. I want to find out why such differences arise and also how successful the two indexes are in achieving the aim of each measurement project. I conclude that differences in the specific purposes for which the indexes are designed play an important role both in choosing the particular concept of well-being and in the method of measurement used in each index. I also conclude that such differences affect the criteria by which we evaluate the success of each index.

The chapter is organised as follows: First, I will describe the background of the two indexes and in particular the motivations and purposes behind each of them. Second, I will give a brief description of the content of the two indexes, including discussions about why particular measurement methods were chosen for the two indexes. The last section will set purpose-dependent criteria for each index and examine the indexes accordingly.

2-1: Background of the construction of the Human Development Index and the Life Satisfaction Indicator

The Human Development Index (HDI) aims to measure and to make international as well as inter-temporal comparisons of 'human development,' a concept defined by the Human Development Report of the United Nations

Development Programme (UNDP) as 'a process of enlarging people's choices.' The UNDP's main motivation was to set an alternative policy objective to growth in GDP, because policies which have GDP growth as the main objective had been associated with undesirable consequences in many countries (both developed and developing): drug-related crimes, pollution-related disease, the breakdown of families, and terrorism and violence. People in the UNDP's Human Development Report Office believed that by shifting the policy objective to one in which humans are at the centre of concern would reduce such undesirable side-effect consequences. The definition of human development in the Human Development Report was developed to meet this requirement, on the implicit understandings that improved human development is by definition choice enlargement and that having wider opportunities to choose from results in an improvement, not essentially of economic conditions, but of human well-being.[1]

On the other hand, the 'Life Satisfaction Indicator' (LSI) designed by an officer in the Japan Economic Planning Agency,[2] Kosei Fukuda, was developed to measure people's well-being in terms of their mental state: how happy or how satisfied they are about their lives in general.[3]

Fukuda's original motivation in developing this index was to solve the paradox he found in the relationship between individuals' satisfaction levels revealed in public opinion polls and the growth rate of the per capita GDP in Japan between 1987 and 1993. He discovered that although the relationship was positive between 1975 and 1986, from 1987 to 1993 the relationship turned negative, that is, people's

---

1 Problems associated with these implicit understandings are discussed in the postscript of the thesis.

2 One of the ministries in Japan that is responsible for economic forecasting and analysis.

3 See Fukuda (1995).

satisfaction levels were declining as per capita GDP grew. This negative relationship between satisfaction and GDP growth was a paradox because we would normally expect, and economists or policy-makers generally assume, that the growth of GDP would be associated with an increase in people's satisfaction with their lives, in addition to their material well-being. Fukuda tried to find out the reason for this paradoxical relationship between the satisfaction level and GDP by searching for factors other than GDP that are likely to be causing the change in people's satisfaction levels. One of Fukuda's implicit assumptions was that in fact GDP had not became a variable that negatively correlates with the level of satisfaction. Rather, Fukuda conducted his investigation under the assumption that changes in other social or economic conditions must have distorted the relationship between the growth of GDP and people's satisfaction level.

The LSI was also designed to be used for policy intervention by the government. For the government to improve the degree to which people find their lives satisfactory, government officials need to know which factors other than GDP are related to the level of satisfaction and to find out the conditions, *i.e.*, the combinations and the weightings of those factors, under which people's satisfaction levels improve. Accordingly, as detailed below, the LSI assesses the satisfaction level of individuals through adding up the satisfaction levels achieved in different areas of their lives, and associating these areas with objective indexes found statistically to cause changes in the related satisfaction levels. For example, if the relationship between a particular objective index such as consumer prices and the satisfaction level of the relevant area of life such as the area of 'consumption' is negatively correlated, and if it is found that the objective index (consumer price) is rising, then *ceteris paribus* we can infer that there is a pressure for the satisfaction level of this particular

area of people's lives to fall, which in turn is likely to cause the general satisfaction level to fall, *ceteris paribus*. Accordingly, as a policy maker the government would act to improve overall satisfaction by policies designed to reduce consumer prices, again *ceteris paribus*.

As reflected in the above discussion, both the UNDP and Fukuda are interested in improving societies in ways that cannot necessarily be achieved by income growth. But the objectives of the improvement as well as the purpose for which the UNDP and Fukuda create indicators are different. The UNDP sets human development as an alternative policy objective and designed the HDI to facilitate international and intertemporal comparisons of human development. On the other hand, Fukuda aims at maximising the general satisfaction level of the people in Japan, and the LSI was constructed to find policy variables that can be used to achieve the goal. As we see in the following sections, the methods of measurement used in the two indexes are also very different.

2-2: The method of measurement

2-2-1: The HDI[4]

In the HDI, three dimensions are chosen to measure the basic concept of human development. They are (1) for people to lead a long and healthy life, (2) for people to acquire knowledge and (3) for people to have access to resources needed for a decent standard of living. The index is a composite one that combines proxies for each of the chosen dimensions. These proxies or sub-indicators are, respectively, for

[4] The definition of the HDI used in this paper is based on the 1998 Human Development Report unless otherwise stated.

each of the three dimensions, (1) life expectancy at birth; (2) educational attainment, which is a composite indicator currently consisting of (a) the adult literacy rate at two-thirds weight and (b) the combined school enrolment ratio at one-third weight; and (3) the per capita real GDP in purchasing power parity dollars, adjusted for the assumption of diminishing marginal utility. In order to combine the three sub-indicators into a single index, they are normalised by taking the difference between the minimum and the maximum values of each sub-indicator as a denominator. Each component of the HDI of a country can thus be calculated as:

$$\text{Human Development Index component for } x_i = \frac{(\text{Actual } x_i \text{ value} - \text{minimum } x_i \text{ value})}{(\text{Maximum } x_i \text{ value} - \text{minimum } x_i \text{ value})}$$

The HDI of a country is thus the unweighted average of the life expectancy index component, the educational attainment index component and the adjusted real GDP per capita (PPP$) index component. In other words, the HDI value is derived by dividing the sum of the three normalised index components by 3. A pictorial representation of the HDI is included in Figure 2.1.

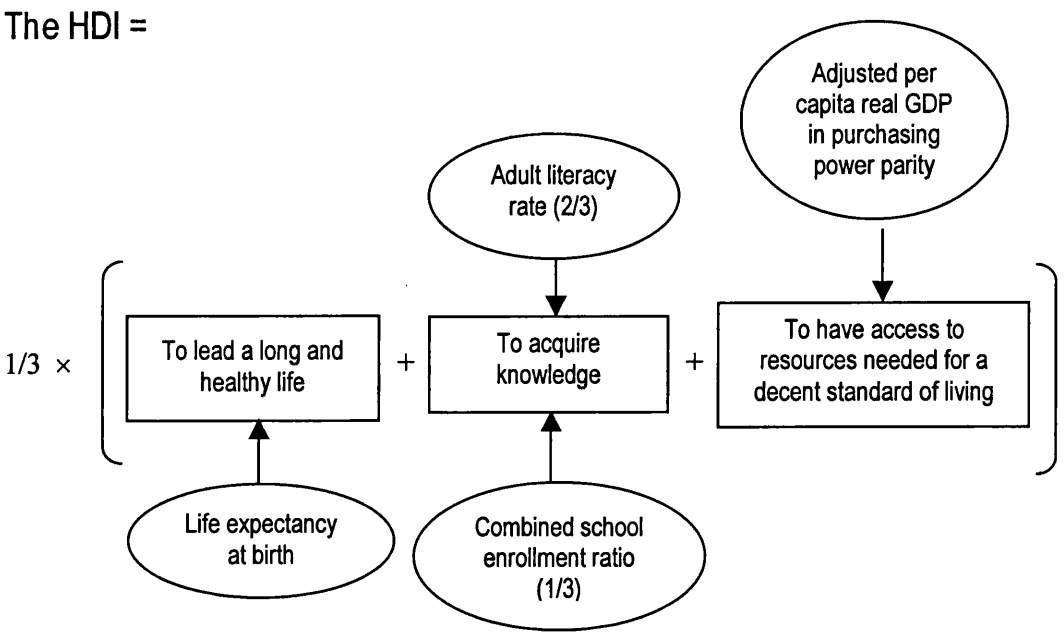

Figure 2.1 Pictorial representation of the structure and normalised factors reflected in the HDI

2-2-2: The LSI

The LSI uses data from a public opinion poll conducted every May of 10,000 Japanese residents over the age of 20. The opinion poll consists of about 30 questions, which include questions from which data is obtained concerning 'the degree of satisfaction in life in general,' 'the degree of fulfillingness of life in general,' 'which aspects of your life you want to improve in the future,' and 'desires for government policy.' The LSI uses data on 'the degree of satisfaction in life in general' (DSL), which is based on answers to the following question: 'Over all, how satisfied are you with your current life?' People are asked to choose one answer from six alternatives, specifically 'satisfied,' 'almost satisfied,' 'rather unsatisfied,' 'unsatisfied,' 'indifferent,' and 'cannot answer.' The DSL is defined as the proportion of people who chose 'satisfied' or 'almost satisfied' among all the people who replied.

One methodological issue involves Fukuda's choice of the DSL data for the LSI. Fukuda selected the DSL among other alternatives because he wants the LSI to be a subjective index representing how rich one feels that one's life is, and because he believes that the DSL is most suitable in these respects.

There were other responses to the public opinion poll that may have been used as alternatives to the DSL to evaluate the richness in people's life, including answers to questions on 'the degree of fulfilment in life,' 'the degree of insecurity (or uncertainty) in life,' or the 'standard of living.' 'The degree of fulfilment in life' is defined as the proportion of all the respondents who chose 'sufficiently fulfilled' or 'rather fulfilled' as an answer to the question, 'To what extent do you feel that you are fulfilled in your life?' (The six alternative answers were 'sufficiently fulfilled,' 'rather fulfilled,' 'not very fulfilled,' 'not at all fulfilled,' 'indifferent,' and 'unable to answer.') 'The degree of insecurity (or uncertainty) in life' is defined as the proportion of all the respondents who chose 'I am worried and feel insecure' as an answer to the question, 'Are you worried or feel insecure about your life?' (The three alternative answers were 'I am worried and feel insecure,' 'I am not worried and do not feel insecure,' and 'unable to answer.') The 'standard of living' data are derived from answers to the question, 'Relative to the general public, to which category do you think you belong with respect to your standard of living?' when people are asked to choose from 'upper,' 'upper middle,' 'middle,' 'lower middle,' 'lower,' and 'unable to answer.' The 'standard of living' data are calculated by giving 100 points to those who answered 'upper,' 80 points to those who answered 'upper middle,' 60 points to those who answered 'middle,' 40 points to those who answered 'lower middle,' and 20 points to those who answered 'lower,' adding all the points, and then dividing the result by the number of the total respondents.

Fukuda explains why he did not use any of these alternatives to the DSL. The reason Fukuda gives for not using 'the degree of fulfilment in life' for the LSI is his judgement that these responses focus more on the degree to which individuals have achieved their 'personal' (meaning inner or moral) goals, and less on the degree to which their everyday life has resulted in satisfaction from external conditions. Assuming that Fukuda's interpretation of the term 'fulfilment' is correct, his defence is consistent with his aim of creating an index for use as a policy instrument through which government can improve richness in Japanese life otherwise than by income growth. Government objectives generally involve ones that can be influenced by policy variables affecting external conditions. But personal goals may include internal objectives that cannot be directly affected by government policies. For example, personal goals such as 'to be honest,' 'to be generous to others' or 'to be industrious' cannot be (at least directly) achieved by government interventions, such as those through taxes, interest rates, money supply or the provision of public goods.

Similarly, Fukuda does not use 'the degree of insecurity (or uncertainty) in life,' in part because, as with the case of 'the degree of fulfilment in life,' 'the degree of insecurity (or uncertainty) in life' tends to relate to personal issues over which governments do not have a direct influence. That is, uncertainties and a feeling of insecurity arise not only from aspects of well-being that governments can affect but also from an individual's personal problems. However, this reason cannot fully explain Fukuda's rejection of this alternative, since the same is true for the DSL. How 'satisfied' one is in life depends not only on economic and social factors that government can improve, but also on factors such as the person's social life, his

relationship to his family, how he gets on with his colleagues at work, etc.[5] Another reason why Fukuda does not use 'the degree of insecurity (or uncertainty) in life' for the LSI involves a concern that the data are affected not only by the person's evaluation of the current state of affairs but also by his evaluation of his future prospects. But why does this factor matter for the LSI? My explanation is that governments need to have a fixed point of reference for their effort to improve well-being using various policy instruments. In the case of the LSI, such reference point is the current state of the public satisfaction: given this evaluation by the Japanese public, the government tries to make improvements in the future. If, as Fukuda claims, 'the degree of insecurity (or uncertainty) in life' reflects anticipation over conditions at some unspecified time in the future, this data set is arguably not suitable for the purpose of the LSI.

Finally, Fukuda claims that 'the standard of living' alternative is not appropriate for the LSI because this evaluation expressly focuses on how a respondent's life is relative to others. However, such a relative evaluation is not inconsistent with the subjective evaluation in which the LSI is interested. A subjective assessment of one's current life can be made in comparison to the state of life of others. Instead, the real problem Fukuda intends to make about 'the standard of living' alternative must be that its virtually total emphasis on the person's relative evaluation may limit the scope of his subjective evaluation of his well-being, that is, how a person sees his own life irrespective how others in the society are doing. For one of the requirements of a subjective evaluation is that each subject is free to use his own criteria for the assessment.

---

[5] For convenience, the reference to 'his' or 'he' will be used to refer to 'his or her' or 'he or she.'

Once the DSL is chosen as the indicator for subjective well-being, the next stage in the development of the LSI is to find factors that explain changes in the DSL. Here, Fukuda assumes that the DSL should be an aggregate of the degrees of satisfaction in different areas of life. The assumption is not free from problems, which I discuss in the next chapter, but here I will focus on the procedure of selecting such areas. Fukuda identified candidates for the different areas of life that affect the general satisfaction level using the results of the opinion poll in two areas, specifically the questions asking people about the 'requirements for government policies' and the questions on the 'areas of life on which you want to work towards improvements.' Fukuda notes that the best data set for the purpose may be the answers to the section of the opinion poll on 'the degree of satisfaction in different areas of life,' but these data are available only from 1986, which is not sufficient to run regressions. For this reason he chose two alternative data sets for which the data are available for a relatively long period of time, specifically (1) the 'requirements for government policies' (available since 1967) and (2) 'areas of life on which you want to work towards improvements' (available since 1974). Fukuda takes areas in which people see the need for government policy intervention and areas in which people want to work towards improvements as areas in which people are not satisfied, and regards these areas as candidates to explain the change in the DSL.[6]

Regression models are used to identify which of these areas of life have satisfaction levels (the strength of the requirements people have for government in different areas of life) that are most strongly related to the changes in the level of the DSL. The method here is to select the areas of life that have a significant effect on the

[6] Some assumptions are required for this connection to be sound. I will discuss these assumptions in Chapter four.

DSL and to find the correlation coefficient of the regression function explaining the level of the DSL using the following methodology: first, multiply (-1) by the variables that are expected to have negative coefficient; then run a regression using all the n variables available that are expected to be relevant to the DSL; eliminate the variable that has the smallest t-value, and run the regression again, this time with n-1 variables; again, eliminate the variable with the least t-value and run the regression with n-2 variables; continue the process until all the remaining variables pass the t-test (for example with 5% significance). Five areas were selected under this statistical selection process, specifically leisure, prices, education, employment, and housing.[7]

The next step in the process of developing the LSI assumed that the feeling of satisfaction or dissatisfaction people have about each of those five areas of life is further influenced by phenomena in society that are measured by objective factors, such as total working hours or the inflation rate. Accordingly, this next step was to find such phenomena for each of the five areas. The method of choosing such objectively measurable phenomena was first, to identify from the 'social life statistical index' compiled by the Secretariat (one of the ministries of the Japanese government) individual indexes that are relevant for each area.[8] The second part of this process was to choose one of the individual indexes as the 'basic index' for each of the areas and then to select further non-basic indexes (three to six) that are most strongly correlated to the basic index.[9] The final part was to vary the combinations of the non-basic

---

[7] Among the items that were not selected by this process are, for example, social security, traffic safety, and energy.

[8] Fukuda categorises the individual indexes regarded as relevant for the degree of satisfaction in each area of life based on the New National Life Indicators, which divides aspects of life into 8 areas (such as education, housing, working) and 4 criteria (such as security and equality). (Fukuda 1995, pp. 146-7, 167).

[9] The decision on which objective index should be chosen as the basic index involves a subjective judgement of the index-maker. (Fukuda 1995, p. 153)

indexes (combinations of two to three), and use the regression function that has the DSL as an explanandum and the objective indexes for each area of life as explanans to select combinations of indexes for each area that best explain the DSL.[10] These operations were used to obtain the correlation coefficient (weight) assigned for every objective index (both basic and non-basic) that explains the satisfaction level of the five different areas of life, as well as to obtain the correlation coefficient (weight) assigned for each of five areas of life used to calculate the satisfaction levels in the DSL. This process results in a LSI that measures the level of satisfaction in life in general and that is composed of different satisfaction levels for five areas of life, each of which is measured by several objective indexes.[11] A pictorial representation of this process is included in Figure 2.2. Table 2.1 lists an example of the LSI based on a time-series result between 1975 and 1993.

---

[10] The problem of how to aggregate objective indexes with different units is solved by the normalisation method, which is to divide the difference between the actual value and the mean value by the standard deviation. (Fukuda 1995, p. 152) This method of normalisation was used by W. M. Persons as early as 1919 when he attempted to make the cyclical fluctuations of economic data comparable within and between data series. (See Morgan 1990, p. 60)

[11] Satisfaction levels of different areas of life are all measured by percentage (the same unit) and the weighted summation is done by regression. The trend-elimination operation is conducted for the satisfaction level of life in general.

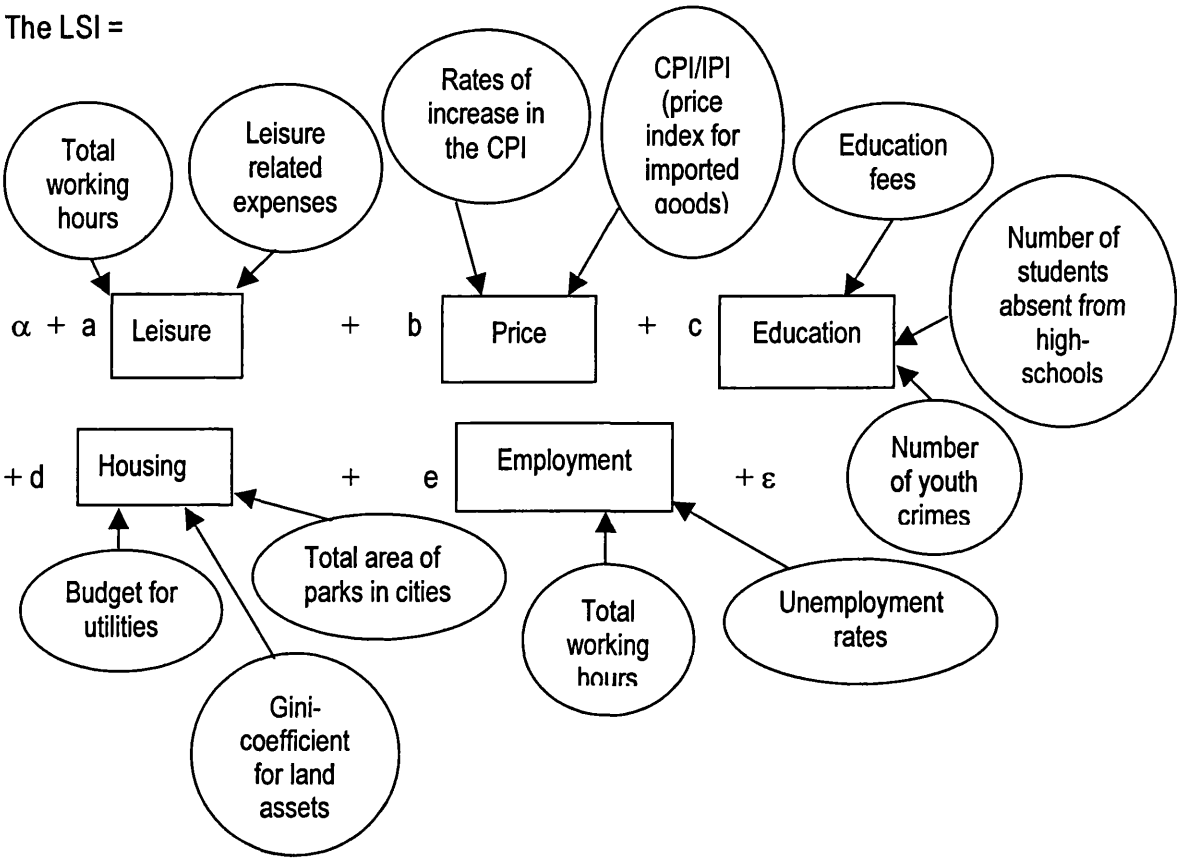

Figure 2.2 Pictorial representation of the structure and factors reflected in the LSI

Table 2.1 The Life Satisfaction Indicator: a time-series result between 197593[12]

Constant	Leisure	Prices	Education	Housing	Employment	Adj – R^2	St Err	DW
-0.102	0.350	2.694	1.215	0.592	1.443	0.8870	0.3661	1.2363
(-1.29)	(2.05)*	(6.80)**	(6.97)**	(2.64)**	(4.90)**			

Leisure factor:

Constant	Total working hours	Leisure Related Expenditures	Adj – R^2	St Err	DW
-0.011	-0.885	-0.764	0.6988	0.5633	1.1004
(-0.08)	(-5.99)**	(-5.16)**			

Price factor:

Constant	Rates of increase in the CPI	CPI/IPI (Price index for imported goods)	Adj – R^2	St Err	DW
0.067	-0.653	-0.414	0.6997	0.5378	0.6246
(0.54)	(-5.05)**	(-3.20)**			

Education factor:

Constant	Education fees	Students absent from high schools	Numbers of youth crime	Adj – R^2	St Err	DW
-0.034	1.121	-0.422	-2.155	0.8212	0.4289	2.2188
(-0.35)	(4.63)**	(-2.17)*	(-6.49)**			

Housing factor:

Constant	Budget for utilities	Gini-coefficient for land assets	Total area of parks in cities	Adj – R^2	St Err	DW
0.1	0.196	-0.294	0.656	0.8949	0.298	0.8518
(1.46)	(2.13)*	(-4.16)**	(7.13)**			

Employment factor:

Constant	Total working hours	Unemployment rates	Adj – R^2	St Err	DW
-0.064	-0.609	-0.457	0.895	0.3191	0.9956
(-0.87)	(-6.83)**	(-5.12)**			

Notes: ** indicates 5% significance and * indicates 1% significance in null-hypothesis tests.

[12] Table translated from Fukuda 1995, Appendix page 28.

2-3: Purpose-dependent criteria of the index

As discussed above, although both indexes are primarily concerned with improving the well-being of a society, their more specific objectives as well as the methods used to measure this well-being are very different. As detailed below, this is partly because of differences in philosophical ideas about what is good for a society, but also because of differences in (1) who makes the index, (2) what the expected ranges of applicability are, and (3) how the index is to be used in practice. The discussion reveals that wide varieties of concerns other than those related to concepts enter into the process of constructing indexes.

This chapter and the two chapters that follow do not ask the more fundamental question of which index is better to use in improving well-being of a society. The question is discussed in the postscript of the thesis.

2-3-1: The HDI

As noted above, the HDI was developed to measure and thus to encourage improvements in human development, which the UNDP proposed as an alternative policy objective to that of higher income level. The HDI is intended to be used to compare degrees of human development across countries and across time. I use five criteria to examine the extent to which the HDI achieves this purpose. The five criteria, discussed in the succeeding sub-sections of this chapter, are:

1. Universality
2. Differentiation from a GDP measure and more discrete orderings of countries
3. Policy usefulness

4. Comparability across space
5. Comparability across time

2-3-1-1: Universality

According to Basic Facts about the United Nations, one of the principles of the United Nations is to act based on the sovereign equality of all its Members.[13] The principle of equality fundamentally affected the structure of the HDI by making one of the purpose-dependent criteria of the HDI its universality. Actually, the concept of universality applied in two different ways, specifically requiring that the HDI be universally *applicable* and that it be universally *acceptable*.

The former purpose – that the HDI be universally applicable – results from the fact that the UNDP, which is a branch of the United Nations and which created and publishes the results of the HDI, had to equally consider the member countries' different ideas about what human development is. In other words, the HDI has to conform to diverse values that people in the member countries have: the HDI has to be a universal index in a sense that it has to apply universally, notwithstanding these diverse values.

The concept of human development defined as 'enlarging choice' by the Report seems to pass the test that it be universally applicable. For example, a review of discussions about the UNDP's human development project in newspapers and magazines did not show objections that the HDI is inadequate as a definition of universal value. Difficulties in meeting the criterion of universal applicability arise

[13] See United Nations (1998), p. 5.

once an attempt is made to represent or redefine the concept in a concrete way, as I show in the following discussion.[14]

The consideration of universal applicability fundamentally affects the identification of the different dimensions of choice in the HDI. Specifically, in order for the index to be universally applicable, the dimensions of the HDI must measure choices in life that are important notwithstanding diverse aims that people have across different cultures, age groups, or genders. On this point, the Report says,

> - - - a broad definition [of human development] makes it possible to capture better the complexity of human life – the many concerns people have and the many cultural, economic, social and political differences in people's lives throughout the world.[15]

As noted above, the three dimensions used in the HDI are: 'to lead to a long and healthy life,' 'to acquire knowledge,' and 'to have access to resources needed for a decent standard of living.' These each qualify as broadly defined expressions of choices. All appear to meet the criteria that they be universally applicable. Less broad and less general expressions of the third dimension, for example, may be 'having a TV, a phone, a car, a washing machine, a refrigerator, and a house with at least two bedrooms plus a bathroom.' But such expressions are not appropriate for the HDI, because they are too detailed and culturally specific to serve as universally applicable criteria for evaluating the choice people have. By contrast, the broadly defined dimensions of the HDI are well suited for the UNDP's measurement project.

Another less obvious example of a dimension that does not meet the criterion of being universally applicable is 'being employed,' which some may say is an

---

[14] There is a related issue of whether the HDI is *comprehensive*, that is, whether it covers the range of factors that are considered important to human development. I have not discussed this issue separately, but it appears to be the basis for some suggested addition to the dimensions of the HDI, for example, the suggestion discussed below of including the dimension of political and civil rights.

[15] UNDP (1990), p. 11.

important dimension of choice. If the UN consisted only of cultures and societies that value employment as an important option in life, then this dimension should be included in the HDI (as it is in the LSI). However, this is not actually the case. There are some societies, for example, that do not consider women working as an important element of choice enlargement. Here again, the fact that there is not a universal agreement on the applicability of this measure distinguishes the three dimensions used in the HDI, which do appear to be universally applicable.

As noted above, a second aspect of the universality requirement is the test that the dimensions chosen for the HDI measurement be universally *acceptable*. The impact of this criterion can be seen in the consideration of including political and civil rights, which the UNDP's Human Development Report Office acknowledges as important aspects of human development. The Human Development Report Office's view is expressed in the 2000 Report, which features specifically issues related to human rights, notes:

> At all levels of development the three essential capabilities are for people to lead a long and healthy life, to be knowledgeable and to have access to the resources needed for a decent standard of living. *But the realm of human development extents further: other areas of choice highly valued by people include participation, security, sustainability, guaranteed human rights* – all needed for being creative and productive and for enjoying self-respect, empowerment and a sense of belonging to a community.[16]
>
> *An adequate conception of human development cannot ignore the importance of political liberties and democratic freedoms.* Indeed, democratic freedom and civil rights can be extremely important for enhancing the capabilities of people who are poor.[17]

However, on the issue of political and civil rights, again there is a conflict among member countries. But the reason for the conflict is different from the example

---

[16] UNDP (2000), p. 17, emphasis added.

[17] Ibid., p. 20, emphasis added.

of whether to include the dimension of employment. As explained below the reason why this dimension is not included in the current HDI is not fundamentally because some countries believe that the dimension is not appropriate for evaluating human development, nor that the member countries are opposed to the proposed method of measuring political and civil rights. The problem exists in the UNDP publishing the result.

As a matter of substance, Partha Dasgupta, Meghnad Desai and many others believe that the HDI is incomplete as a measurement of choice enlargement without the dimension of political and civil rights.[18] Thus, Desai claims as follows:

> - - - human development was to ensure that everyone had certain basic capabilities – to lead a long and healthy life, to be able to engage in, and profit by, productive work, and to communicate freely and have access to information, and to participate in the political and social life of the community. *Political freedom has to be seen as providing a capability, especially the last two: to communicate and to participate in the community.*[19]

Desai, who contributed in creating the HDI, has designed 'a political freedom index' that consists of five clusters (integrity of self, rule of law, political participation, freedom of expression, and equality before law), each of which is in turn related to a set of indicators for which qualitative and quantitative data must be gathered. Although Desai is aware of the immense task that would be required to implement his methodology for measuring political freedom on a consistent basis for a number of countries, he also notes that 'there are enough sources for data – from non-governmental organisations such as Amnesty International and Human Rights Watch to official bodies such as US State Department - - -.'[20]

---

[18] See Dasgupta (1990) or Desai (1995).

[19] Desai (1995), p. 201, emphasis added.

[20] Ibid., pp. 218 – 9.

Given the plausibility of the project of measuring political freedom, the fact that this factor is not included in the HDI must be caused either because there is no universally accepted method of measuring political freedom, or because the inclusion of the factor itself is a problem for some countries. We can see that the latter is the case from Desai, who explains the following:

> At this stage in any social science research done in an academic environment, one would present the results. *But research on political freedom is not so straightforward.* Since the research was undertaken with a view to incorporation in the United Nations Development Programme's (UNDP's) Human Development Report for 1992, it had a diplomatic rather than academic environment to cope with. The methodology proposed here was explained to the relevant committee of the UN Economic and Social Council and was approved. But when the actual score for 101 countries was computed, *a number of countries objected officially in the UN General Assembly and elsewhere. It was not contended that the results were wrong or contained errors. The argument was made that the UNDP had no mandate to work on human rights and so could not publish such an index!*[21]

Consistent with Desai's explanation on this matter, the 1993 Report notes:

> Further work [on measures of political freedom] is needed, preferably by academics who can look at this question in an environment free from international political pressures.[22]

This discussion is illustrative of the fact, noted above, that because of the nature of the HDI's index-maker, the dimensions in the index must meet not only the criterion of universal *applicability* – that is, that they conform to different ideas people have about what are the important elements in enlarging choice – but also the criterion of universal *acceptability*. To satisfy this latter criterion, the dimensions of an index for human development must be accepted by all the countries that are the members of the UN. It follows from this discussion that an index may be universally applicable but not universally acceptable for a variety of reasons. For example,

---

[21] Ibid., p. 219, emphases added.

[22] UNDP (1993), p. 105.

universally applicable dimensions of the HDI may not be universally accepted by mistake, or because of bribery, or because of the type of political pressures that Desai points out.

A further issue concerns how to identify dimensions of an index that will be universally acceptable. In general, if an evaluation of well-being affects international resource allocation or other international policy decisions, that dimension is not likely to satisfy this criterion, since different countries are likely to have different opinions about how the evaluation should be made based upon their circumstances. For example, UN member country *A*, which seeks grants from international organisations, may want the index to focus on areas of development in which the country is relatively poor and in need, thus advantaging it compared to other countries competing for the same grants. Or another UN member country *B* may not want an index that emphasises particular aspects of the society where its development is slow because the government of country *B* regards the publication of such information as disadvantageous to its reputation. The UNDP's HDI must be accepted by UN member countries like *A* and *B* that have different interests in how the HDI will be used.

In short, because the HDI has to be both universally applicable and universally acceptable, aspects of choice that are universally regarded as important – such as political and civil rights – are not included in the measurement. The HDI must thus be seen in light of the fact that it was produced under strong political pressure. The broad concept of human development as choice enlargement did not prompt such political pressures, but efforts to quantify and define the concept more specifically did prompt political pressures.

2-3-1-2: Differentiation from a GDP measure and more discrete orderings of countries

A second purpose-dependent criterion of the HDI involves the interest in differentiation. Here the differentiation involves two aspects: (1) the interest in an index that will give a ranking that need not coincide with the one based upon GDP, and that is better for what we want to measure; and (2) the interest in a measure that will permit more discrete orderings of countries that are genuinely different in their levels of human development.

The former interest developed naturally as a purpose-dependent criterion because the initial goal of the UNDP's human development project was to facilitate an alternative policy objective to that of simply encouraging higher income levels. The project was also motivated by the fact that even among countries with similar levels of income, there are diverse ranges of development as measured by the choices people have in their lives. Therefore, the country orderings according to the HDI will not usually resemble that of GDP rankings.

On the issue of the index's ability to discriminate in general, the Report says that this condition is generally required for the index to be universal. The Report says, 'As a measure of *universal* index, the HDI needs variables that discriminate among countries.'[23] However, the reason the Report gives for treating this need as part of the criterion of a universal index is not a convincing one. If the ability to discriminate is a criterion of a universal index, we may in principle reject an index that is more accurate than others in measuring human development only because it is less discriminatory compared to rival indexes. Unless we regard the index's ability to

[23] Ibid., p. 105, Italics by the Report.

discriminate as more important than accuracy, however, such a judgement is not appropriate. Universality and the ability to discriminate may complement each other in producing an accurate index, but as I show in the following discussion, universality does not *require* that an index discriminate well.

This is not to say that an index's ability to discriminate is unimportant. But its importance is not because this ability is a general requirement for universal indexes, but rather because we want an index to be useful in providing discrete orderings of countries (probably for policy reasons).

On examining rankings in the HDI, we find that the HDI is indeed successful in giving orderings of countries that are distinct from the country rankings produced by a GDP measure. For example, using the 1998 Human Development Report, I find the following facts:

1. Among the 20 countries with the highest HDI ranking, only two (Belgium 12th and Austria 13th) are ranked exactly the same as in the real GDP per capita (PPP$) rankings. The largest differences in the ranking is 17 places (for Finland, which ranks 6th in the HDI and 23rd in the real GDP ranking). The average absolute difference between the two rankings among these 20 countries is 8.75 places.
2. Among the 20 countries with the lowest HDI ranking, only one country (Guinea-Bissau 164th) is ranked exactly the same as in the real GDP per capita (PPP$) result. The largest difference in the ranking is 28 places (for Angola, which ranks 156th in the HDI and 128th in the real GDP ranking, and for Senegal, which ranks 158th in the HDI and 130th in the real GDP ranking). The average absolute difference between the two rankings among these 20 countries is 5.8 places.

In order for the HDI to be a successful measure of human development, however, we need to know not only that the index gives a ranking different from a GDP measure, but also that the index is better than the GDP for what we want to measure. For this purpose, we would need to do a more detailed comparison between rankings of the different types of the measurement of human development. The discussion that follows concerning the interest in a more discrete ordering of countries provides examples of such micro level analyses of the different versions of the HDI.

As noted above, the HDI was also motivated by an interest in facilitating a more discrete ordering of countries that are genuinely different in their levels of human development. This motivation is apparent from a comparison of the three alternative types of educational attainment indexes used in the HDI, one in use prior to 1991, one in use from 1992 to 1994 and the current one in use from 1995. I use this comparison also to explore a number of difficulties involved in implementing this purpose of obtaining a more discrete ordering of countries.

Originally, to measure educational attainment, only the adult literacy rate was used. To explain this choice, the Report stated that 'literacy is a person's first step in learning and knowledge-building,' and therefore 'for basic human development, literacy deserves the clearest emphasis.'[24] However, there developed a concern that adult literacy rates did not discriminate among developed countries, since 24 countries had a 99% adult literacy rate (1985 data). Accordingly, from the 1991 Report through the 1994 Report, the indicator for educational attainment was modified to include both adult literacy and the mean years of schooling, with two-third weights on the

[24] UNDP (1990), p. 12.

former and a third weight on the latter.[25] This change was consistent with the Report's aims of constructing an index that gives different rankings of development from those based on GDP levels and that provides a more discrete ordering of countries. (The purposes of using such a more discriminating index may also include more pragmatic reasons, such as convenience in using the discriminatory indicator for making policy decisions.) In any case, the alternative educational attainment index used between the 1991 Report and the 1994 Report gave results that did provide a more discrete ordering of countries, as detailed below.

I have examined the effect of the changes in the educational attainment index on countries' educational attainment rankings. Included in Table 2.2 are the rankings based on the 1991 Report figures for mean years of schooling (1980 data) including the 24 countries that had a 99% adult literacy rate (1985 data).

[25] See UNDP (1993), pp. 105-6.

Table 2.2 Country rankings according to 'mean years of schooling' (among countries with a 99% adult literacy rate)

Ranking	Country	Mean years of schooling
1	USA	12.2
2	Canada	11.4
3	United Kingdom	10.8
4	Japan	10.4
5	Denmark	9.7
6	Norway	9.6
	Austria	9.6
8	Sweden	9.4
	France	9.4
10	Australia	9.3
11	Finland	9.2
12	New Zealand	8.9
13	Germany	8.8
14	Hungary	8.6
15	Switzerland	8.3
16	Netherlands	7.9
	Belgium	7.9
18	Czechoslovakia	7.8
19	Luxembourg	7.7
	Ireland	7.7
21	USSR	7.6
22	Iceland	7.5
23	Poland*	7.3
24	Bulgaria*	7
25	Cyprus*	6.9
26	Korea, Rep. Of*	6.6
	Romania*	6.6
	Philippines*	6.6
29	Greece*	6.5
30	Italy*	6.4
31	Barbados	6.3
32	Hong Kong*	6.2
	Bahamas	6.2
	Chile*	6.2

*Countries that have lower than a 99% adult literacy rate.

The above table reveals that when in 1991 the educational index was changed to add mean years of schooling to the adult literacy rate, the 24 countries ranked

equally by literacy rate (countries with 99% adult literacy rates) were discriminated into 20 different levels. The countries with a large drop in their educational attainment rankings because of the new educational index were: Bahamas (1^{st} to 32^{nd}), Barbados (1^{st} to 31^{st}), Iceland (1^{st} to 22^{nd}), USSR (1^{st} to 21^{st}), Luxembourg (1^{st} to 21^{st}), Ireland (1^{st} to 19^{th}), Czechoslovakia (1^{st} to 18^{th}), Netherlands (1^{st} to 16^{th}), and Belgium (1^{st} to 16^{th}).

But the fact that an indicator provides a more discrete ordering of countries does not say whether or not the indicator is correct. In fact, the latest educational attainment index that has been used since the 1995 Report, which is the aggregate of adult literacy rates and the combined school enrolment ratio (with the two-third weights for the former and a third weight for the latter), results in a very different ordering of countries than the index used from 1991 through 1995.[26] Again, I have examined the effect of the change on the countries' rankings of educational attainment. To do this I have compared the results in the 1994 and the 1995 Reports. The following table lists the rankings of educational attainment in the 1994 Report HDI – rankings based on the aggregation of adult literacy rate (1992) and the mean years of schooling (1992), with two-thirds and one-third weights on each – among countries with a 99% literacy rate:

---

[26] I used data from the 1991 Report (with mean years of schooling [1980] and literacy rates [1985]) for the comparison between the *1990 Report type index* and the *1991 – 1994 Report type index*, and used data from the 1994 Report (with mean years of schooling [1992] and literacy rates [1992]) and data from the 1995 Report (with combined school enrolment ratio [1992] and literacy rates [1992]) for the comparison between the *1991 – 1994 Report type index* and the *1995 Report type index*. The different changes in the country-ranking found in two sets of comparisons show that different countries are sensitive to different types of educational data, *assuming* no change during this 1980-92 period of the general structure of the country's educational sector relative to other countries (e.g. USSR is relatively high in its literacy rates but the period of schooling tends to be relatively low, whereas the educational structure in the USA produces high performances in both aspects), but only changes in the absolute levels of each aspect within the structure.

Table 2.3 Country rankings according to the old 'educational attainment index,' 1994 Report (among countries with a 99% adult literacy rate)

Rankings	Countries	Educational Attainment Index	Rankings	Countries	Educational Attainment Index
1	USA	2.81		Latvia	2.58
2	Canada	2.8	26	Lithuania*	2.57
3	Norway	2.78		Russian Federation*	2.57
	France	2.78	28	Korea, Rep. Of*	2.55
	Australia	2.78	29	Argentina*	2.53
6	United Kingdom	2.76	30	Poland	2.52
7	Switzerland	2.75	31	Uruguay*	2.47
	Germany	2.75	32	Italy"	2.45
9	Sweden	2.74	33	Spain*	2.42
	Austria	2.74		Belarus*	2.42
11	Belgium	2.73	35	Bahamas	2.39
12	Netherlands	2.72		Chile*	2.39
13	Denmark	2.71	37	Saint Kitts and Nevis	2.38
14	Japan	2.7		Mongolia*	2.38
	Finland	2.7	39	Cyprus*	2.35
16	New Zealand	2.69		Samoa*	2.35
17	Luxembourg	2.68	41	Bulgaria*	2.32
18	Hungary	2.63	42	Jamaica*	2.32
19	Barbados	2.61		Moldova, Rep. Of*	2.32
20	Czechoslovakia	2.6	44	Armenia*	2.31
21	Iceland	2.59		Georgia	2.31
22	Israel*	2.58		Philippines*	2.31
	Ireland	2.58		Philippines*	2.31
	Estonia	2.58			

* Countries that have lower than a 99% adult literacy rate

The following table, on the other hand, lists the rankings of educational attainment in the 1995 Report HDI – rankings based on the aggregation of adult literacy rate (1992) and combined first-, second-, and third- level gross enrolment ratio (1992), with two-third and one-third weight on each, respectively – among countries with a 99% literacy rate:

Table 2.4 Country rankings according to the new 'educational attainment index,' 1995 Report (among countries with a 99% adult literacy rate)

Rankings	Countries	Educational Attainment Index	Rankings	Countries	Educational Attainment Index
1	Canada	0.99		Kyrgyzstan*	0.9
2	USA	0.98		Uzbekistan*	0.9
	Finland	0.98	37	Israel*	0.89
4	Netherlands	0.95		Barbados*	0.89
	Norway	0.95		Belize*	0.89
	France	0.95		Czech Rep.	0.89
7	Spain*	0.94		Estonia	0.89
	Belgium	0.94		Latvia	0.89
	Austria	0.94		Russian Federation*	0.89
	Denmark	0.94		Antigua and Barbuda*	0.89
	New Zealand	0.94		Moldova, Rep. Of*	0.89
	Ireland	0.94	46	Italy*	0.88
13	Iceland	0.93		Greece*	0.88
	Germany	0.93		Cyprus*	0.88
15	Japan	0.92		Trinidad and Tobago*	0.88
	Sweden	0.92		Hungary	0.88
	Australia	0.92		Lithuania*	0.88
	United Kingdom	0.92		Korea, Dem. People's Rep. Of*	0.88
	Saint Kitts and Nevis	0.92		Philippines*	0.88
	Armenia*	0.92		Guyana*	0.88
	Georgia	0.92	55	Chile*	0.87
22	Switzerland	0.91		Ukraine*	0.87
	Korea, Rep. Of*	0.91		Kazakhstan*	0.87
	Poland	0.91		Saint Lucia*	0.87
	Grenada*	0.91		Azerbaijan*	0.87
	Saint Vincent*	0.91		Tajikistan*	0.87
	Turkmenistan*	0.91	61	Fiji*	0.86
	Samoa (Western)*	0.91	62	Luxembourg	0.85
29	Bahamas*	0.9		Costa Rica*	0.85
	Argentina*	0.9		Cuba*	0.85
	Uruguay*	0.9		Suriname*	0.85
	Slovakia	0.9		Romania*	0.85
	Belarus*	0.9		Lebanon*	0.85
	Dominica*	0.9			

*Countries that have lower than a 99% adult literacy rate.

A comparison of these two tables shows the dilemma of trying to decide among the different measures of educational attainment. By the standard of their ability to distinguish among countries the measure of educational attainment has become less effective, because the latest educational attainment index provides a less discrete ordering of countries than the previous index. As noted above, in the 1991 Report, the 24 countries with a 99% adult literacy rate (and thus an identical ranking under the pre 1991 Report) were discriminated into 20 different levels. In the 1995 Report, however, the 27 countries with a 99% literacy rate were separated into only 11 categories.

However, as noted above, the fact that an index is more discriminatory than the other is not a sufficient reason to prefer the former. The discrete ordering has to reflect differences in countries that are genuinely different in their levels of human development. Therefore, there is another aspect of the dilemma, which relates to the very different country rakings that resulted when in 1995 the educational attainment index was changed from adult literacy rates plus mean years of schooling to adult literacy rates plus combined school enrolment ratio. Specifically, among countries whose adult literacy rates remained at 99% (the highest score for the indicator) for both the 1994 and 1995 Human Development Reports, some countries faced large drops or large increases in their rankings from 1994 to 1995. Countries with significant drops in their rankings are: Luxembourg (17^{th} to 62^{nd}), Hungary (18^{th} to 46^{th}), Switzerland (7^{th} to 22^{nd}), Australia (3^{rd} to 15^{th}), the United Kingdom (6^{th} to 15^{th}), Germany (7^{th} to 13^{th}) and Sweden (9^{th} to 15^{th}). Countries with large improvements in their educational attainment rankings are: Georgia (44^{th} to 15^{th}), Saint Kitts and Nevis (37^{th} to 15^{th}), Ireland (22^{nd} to 7^{th}), Finland (14^{th} to 2^{nd}), New Zealand (16^{th} to 7^{th}), Netherlands (12^{th} to 4^{th}), Iceland (21^{st} to 13^{th}) and Denmark (13^{th}

to 7th). Because the adult literacy rates in these countries remained the same throughout the period, and because we would expect mean years of schooling and combined school enrolment ratio to be relatively stable from 1994 to 1995, we can attribute most of the rakings changes to the change in the type of data used.

In reality, it appears that the new educational attainment index for the 1995 Report was chosen without examining the implication of the significant changes in many of the country rankings. Instead, it appears that this latest change in the educational attainment index was made only for reasons related to the quality and the source of the data. The Report explains that the change was made

> - - mainly because the formula for calculating mean years of schooling is complex and has enormous data requirements. Data on mean years of schooling are not provided by any UN agency or international organizations. As a result, estimates must sometimes be used, which are not always acceptable. The combined enrolment ratio overcomes both these problems. It shows the stock of literacy quite easily for those under age 24. And it is based on the work of UNESCO.[27]

The problem with this data-based rationale, of course, is that it does not tell us whether the new index is actually a more effective or more accurate measurement of educational attainment. There is a particular problem in choosing the newest index (1995 index for short) over the 1991 index, without knowing whether or not both are measuring a general capability to acquire knowledge. This is because the better quality of data should justify a change in the index only if the two indexes are measuring the phenomenon we are interested in, but for one of them we have data sets with better quality than the other. This chapter is concerned with purpose-dependent, *a priori* criteria of indexes, not empirical criteria such as data quality. I talk about the issue related to quality of data to emphasise that the ability to give discrete orderings and the quality of data are two separate criteria, so that indexes that pass the former

[27] UNDP (1995), p. 134.

criterion do not necessarily use better data, nor are indexes that use better data necessarily produce more reliably discrete orderings of countries.

It is possible, for example, that the difference in results of the two indexes arises not because they are measuring the same phenomenon and one has better quality data than the other, but because the two indexes are in fact measuring different phenomena. In this case it might well be advisable not to change the index but to focus on better collection of data to improve the index. In this case, two considerations in choosing between the two indexes, the better quality of data available or better measurement-device, would be in conflict.

Is there any way to know the difference (or the similarity) in the potential capacity of the indexes to meet a particular purpose independent from the results they produce? Especially, are we able to find out which index has the better capacity to measure the capability 'to acquire knowledge'? As indicated above, a problem common in creating a measurement-device is that we do not know apart from the measurement what the world looks like with respect to the concept in question.

One potential way to deal with this problem of not having prior knowledge about the rankings of countries with respect to the concept 'to acquire knowledge' itself is to look to information on the relationship between this concept and other concepts for which we have good data.

One possible such relationship is that between the educational attainment index and adjusted real per capita GDP in purchasing power dollars (PPP$), for which we have good data, plus the knowledge we have that countries that have achieved the highest levels of educational attainments are not necessarily the ones that are at the top of the income scale. Using the 1991 Human Development Report, I found the following:

1. Among the top 12 countries in educational attainment index, only two countries (Norway and Finland) have the same rankings in the GDP and in the educational attainment index;
2. Only two countries (Japan and Sweden) get higher rankings in the GDP compared to the educational attainment index, but both only by one ranking; and
3. Eight countries have significantly higher rankings in the educational attainment index compared to the GDP index, as reflected in the following table:

Table 2.5 Educational attainment index rankings and GDP index rankings based on the 1991 Human Development Report (Selected comparisons)

Country	Educational attainment index ranking	GDP index ranking
Canada	2	10
USA	1	6
Australia	8	21
France	8	12
UK	3	20
Denmark	5	9
New Zealand	13	23
Austria	5	14

4. Overall, the result from the analysis shows that countries that obtain the highest educational attainment scores are not the ones that have the highest income levels.

I also examined the relationship between the educational attainment index in the 1995 Report and the adjusted real per capita GDP (PPP$). Using this data I found the following:

1) Among the top 12 educational attainment countries, only the USA obtains a higher ranking in the GDP compared to the educational attainment, but only by one ranking;

2) A total of 11 countries get higher results in the educational attainment rankings compared to the GDP rankings. The differences in their rankings are in general large, which can be seen from the following table:

Table 2.6 Educational attainment index rankings and GDP index rankings based on the 1995 Human Development Report (Selected comparisons)

Country	Educational attainment index ranking	GDP index ranking
Canada	1	8
Netherlands	4	20
Finland	2	14
Norway	4	15
France	4	11
Spain	7	29
Belgium	7	14*
Austria	7	13*
Denmark	7	12*
New Zealand	7	26
Ireland	7	30

* Rankings between the two indexes for these countries are not as large as the numbers suggest, because there are six countries which are at tie in 7th rankings.

3) Again, the overall observation is that countries which obtain the top rankings in their educational attainments are not the ones that enjoy the highest income levels.

As the above results show, unfortunately, neither analysis is useful for our purpose. Specifically, both the 1991 index and the 1995 index are consistent with the general observation that countries that have achieved the highest levels of educational attainments are not necessarily the ones at the top of the income scale. So this particular observation about the relationship between educational attainment and income does not help in choosing one index over the other.[28]

Given the limits in testing measurement-devices with respect to the general concept 'to acquire knowledge,' a further alternative is *to make the concept less abstract while keeping it a universal value.* This approach would respond to the problem with a concept as abstract as 'to acquire knowledge' that too many indexes related to educational attainment are acceptable. Examples of such less abstract characterisation of the concept may be, in addition to 'to be able to read and write with understanding' (the definition of literacy) to have information-sorting skills, to be able to construct a logical argument, to be able to operate basic computer soft-ware programmes, to be able to solve basic mathematics, etc. This list of more concrete objectives may suggest a certain rankings of countries. For example, the 1998 Report includes the following observation on the relationship between various important uses of the general capability to acquire knowledge and income levels:

---

[28] There may be relationships between different countries' educational attainment and other social phenomena that may form as justifications for selecting among alternatives. For example, an observation that countries that have achieved high levels of political liberty may not necessarily be the ones that have the highest levels of educational attainment may distinguish between different indicators for education.

> Industrial countries have achieved nearly 100% literacy rates and 85% enrolment ratios. But new surveys show that many people – 18% of adults on average in 12 European and North American countries – though "literate", have such low levels of skills that they cannot meet even the basic reading requirements of a modern society. Another 29% do not have the ability to be trained in skilled employment. Industrial countries may start falling behind the fast-growing developing countries, especially in technical education. Fewer than a third of students in the industrial countries now enrol for applied or natural science – in Norway and the Netherlands only 1 student in 5. But in Chile, China, and the Republic of Korea and South Africa the proportion is 1 in 2 or 1 in 3.[29]

The statement shows that 'technical literacy' does not necessarily correlate with the amount of education. For example, as noted, Netherlands and Norway have a much lower proportion of students enrolled for applied or natural science than Chile, China and the Republic of Korea, although the general educational attainment levels in the former countries are higher than in the latter countries.[30]

Based on such observations, I have compared the three educational indexes used in the HDI between 1991 and 1995 to assess which of the three better reflects this technical literacy factor. For this purpose I prepared the following list ranking these five countries in the three different measures of educational attainment:

Table 2.7 Country rankings according to different educational attainment indexes (Selected comparisons)

Country	1990 Report	1991 Report	1995 Report
Norway	1	5	4
Netherlands	1	16	4
Chile	45	42	55
China	94	92	103
Republic of Korea	37	34	22

[29] UNDP (1998), p. 23.

[30] See the above quote.

This data, although very summary, suggest that in fact the index in the 1991 Report may be the best measure of educational attainment among these three, since it shows less distinctions between Norway and Netherlands on the one hand and Chile, China and the Republic of Korea on the other hand. In the 1991 Report the average ranking of Norway and Netherlands is 10.5 ([5 + 16] / 2 = 10.5), compared to 56 for Chile, China and the Republic of Korea ([42 + 92 + 43] / 3 = 56), or a difference of 45.5 places. In the 1990 Report, the average ranking of Norway and Netherlands is 1, compared to 58.7 for Chile, China and the Republic of Korea ([45 + 94 + 37] / 3), or a difference of 57.7 places. For the 1995 Report, the average ranking of Norway and Netherlands is 4, compared to 60 for Chile, China and the Republic of Korea ([55 + 103 + 22] / 3 = 60), or a difference of 56 places.

Of course, this very limited comparison is problematic without more data. Indeed, some may say that data that would correspond to even a more refined definition of the content of educational attainment are not available. That may be so. But as Amartya Sen notes in his book 'Commodities and Capabilities,' more data should become available if organisations and researchers raise demands for such data:

> Given the limitations of reliable data, it is not easy to make extensive comparisons of the achievements of different countries in the field of extending capabilities and enhancing functionings. One reason why the data tend to be relatively scarce in this area compared with, say, data underlying GNP and GDP estimates, is the lack of demand for such data. There is no reason why it should not be possible to get more comparative data on, say, morbidity or undernutrition, in different countries. The weakness in the *theory* of well-being and living standards has been partly responsible for the underdevelopment of the data base.[31]

[31] Sen (1985), p. 73, emphasis by Sen.

2-3-1-3: Policy usefulness

The third criterion derived from the purpose of the project of quantification is policy usefulness. This is also a criterion for satisfactory econometric models by Morgan as well as a generic criterion for successful models by Boumans.[32] The index-maker's intention that the HDI be suitable for policy uses is expressed, for example, in the following statement by the 1990 Report:

> In any system for measuring and monitoring human development, the ideal would be to include many variables, to obtain as comprehensive a picture as possible. But the current lack of relevant comparable statistics precludes that. Nor is such comprehensiveness entirely desirable. Too many indicators could produce a perplexing picture – perhaps distracting policymakers from the main overall trends. The crucial issue therefore is of emphasis.[33]

The HDI is useful for policy-making to the extent that information obtained from the index can be analysed quickly and easily by policy-makers. Using a limited number of dimensions (for example three, as in the current HDI), the ups and downs of the overall level of the HDI can be easily analysed because changes can only happen through changes in the three dimensions. As reflected in the above quotation, however, if there were many more dimensions to the index, the result would be 'a perplexing picture;' there would be too many more ways in which the overall levels can be affected. Accordingly, in examining an index with respect to its practical usefulness, whether the index measures human development precisely or imprecisely becomes a minor concern.

It is also important to distinguish what we mean by usefulness in this context. Here we mean that the HDI can be useful at the macro level, for example, in evaluating whether the conditions of human development are improving in a

---

[32] Morgan (1988) and Boumans (1999a).

[33] UNDP (1990), p. 11.

particular country. We do not mean that the HDI and its three dimensions can necessarily be useful at the micro level to assess whether a particular policy will provide an improvement in human development. In fact, as developed more fully in Chapter Four, the index is not made for constructing or evaluating policies to improve human development, since the three dimensions of the HDI are not chosen having this aim in mind. Investigations such as how the government should intervene to improve the level of human development of a particular country have to be done by other means.

2-3-1-4: Comparability across space

From its beginning, the HDI was designed for comparison of human development across different countries, that is, comparability across space. The comparability across space is also one of the criteria for a successful thermometer discussed by Chang.[34] He argued that one of the properties that scientists required for the measurement instrument they were looking for was stability – that the chosen instrument can produce consistent results under a wide range of circumstances. Applied to the case of the HDI, this requirement that the index provide comparability over space leads to a number of more specific requirements.

Firstly, the data sets needed to calculate the HDI must generally be available from countries all over the world. The data sets used in the HDI – life expectancy at birth, adult literacy rate, combined first-, second-, and third- level gross school enrolment ratio, and real GDP per capita – are available from most of the member countries of the UN. However, some compromises have been made in an effort to improve the functioning of the HDI. In particular, in the case of the first- or second-

[34] Chang (1996).

level enrolment data, or both, more than 50 countries had to use estimated or provisional values in the 1998 Report. This need to use estimated or provisional figures was a cost that had to be paid to increase the dimensions of measuring educational attainment. (As noted above, in the original HDI, only adult literacy rates were used to measure educational attainment.)

A second specific requirement to provide comparability over space is that the three dimensions of the HDI must be *relevant* for all the countries under investigation. That is, the dimensions have to be relevant as factors that contribute to 'choice enlargement' (the basic definition of human development), regardless of the country to which we apply the index. As noted above, the HDI dimensions are for people (1) to lead a long and healthy life, (2) to acquire knowledge, and (3) to have access to resources needed for a decent standard of living. As the 1990 Report explained, these dimensions were chosen because they are essential at all levels of development and thus can apply to all countries:

> In principle, these choices can be infinite and change over time. But at all levels of development, the three essential ones are for people to lead to a long and healthy life, to acquire knowledge and to have access to resources needed for a decent standard of living. If these essential choices are not available, many other opportunities remain inaccessible.[35]

Finally, at a very practical level, it appears that in order to make multilateral comparisons among countries, we need to use a composite index like the current HDI, as opposed to making vector comparisons of the three different dimensions. As discussed below, in fact the latter type of comparison is essentially unable to produce multilateral orderings of countries' levels of human development.

---

[35] UNDP (1990), p. 10.

Sen offers a vector comparison means of measuring capabilities in the sub-section of his book called 'Incompleteness: Fundamental and Pragmatic.'[36] This means is proposed for cases where there is no agreement on the relative weights to be attached to different dimensions. Because weights assigned to three dimensions in the HDI are not determined with full confidence or with complete agreement, such a means may be more suitable for the HDI's evaluation of human development. In what follows, I will examine whether a 'dominance partial order,' which is one of the alternative ways in which we could measure human development, would be a good alternative to the current HDI.

First, it is important to understand what a 'dominance partial order' is. A 'dominance partial order' is used when we want to order the objects of comparison according to more than one criterion at the same time. A 'dominance partial order' provides ordering of those objects without specifying particular relative weights for each criterion, because the order is determined according to the rule that having more in each criterion is better. The order is 'partial' because often the comparison using the above rule limits the number of objects that can be ordered. According to Sen:

> - - a particular selection of value-objects (in this case, the functionings and capabilities that are accepted as valuable) would yield a 'dominance partial order' even without specification of relative weights. Having more of *each* relevant functioning or capability is a clear improvement, and this is decidable without waiting to get agreement on the *relative* weights to be attached to the different functionings and capabilities.[37]

If we apply the principle of dominance partial order to the rankings of human development, the rule would be as follows: country *A* has a dominant position in the human development ranking compared to country *B* if *A* is at least equal to *B* in all

---

[36] Sen (1992), pp. 46-9.

[37] Ibid., p. 46, emphases by Sen.

three dimensions and greater in one of them. Rankings of countries according to this rule will not produce a complete ordering, but may produce partial orderings that are enough for the particular purpose for which they may be used.

To see whether it is practical to use the 'dominance partial order' approach in the HDI rankings, I have examined what dominance partial orderings we can obtain from the data sets for the three dimensions of the HDI for the countries at the lower scale of development – the lowest 20 countries in the HDI ranking in the 1998 Human Development Report. Because the dimension of educational attainment is measured in terms of two variables (adult literacy rates and combined school enrolment ratio), the actual test of dominance ordering is conducted using four different data sets, specifically *life expectancy at birth, adult literacy rate, combined first-, second- and third- level gross school enrolment ratio* and *real GDP per capita in purchasing power dollars*. The following table summarises the data sets for these 20 countries:

Table 2.8 HDI rankings of the lowest 20 countries from the 1998 Human Development Report

HDI rank	Country	Life Expectancy at birth (years)	Adult literacy rate (%)	Combined school enrollment ratio (%)	Adjusted real GDP per capita (PPP$)
155	Bhutan	52	42.2	31	1382
156	Angola	47.4	42	30	1839
157	Sudan	52.2	46.1	32	1110
158	Senegal	50.3	33.1	33	1815
159	Haiti	54.6	45	29	917
160	Uganda	40.5	61.8	38	1483
161	Malawi	41	56.4	76	773
162	Djibouti	49.2	46.2	20	1300
163	Chad	47.2	48.1	27	1172
164	Guinea-Bissau	43.4	54.9	29	811
165	Gambia	46	38.6	39	948
166	Mozambique	46.3	40.1	25	959
167	Guinea	45.5	35.9	25	1139
168	Eritrea	50.2	25	29	983
169	Ethiopia	48.7	35.5	20	455
170	Burundi	44.5	35.3	23	637
171	Mali	47	31	18	565
172	Burkina Faso	46.3	19.2	19	784
173	Niger	47.5	13.6	15	765
174	Sierra Leone	34.7	31.4	30	625

Source: UNDP (1998), p. 130.

The method I used to obtain dominance partial orderings of these countries using the four factors in the HDI is simple. First I calculated how many bilateral dominance relationships existed in the 10 countries at the bottom of the HDI rankings: that is, how many combinations of two countries we can find in which one country has at least the same level in all four of the basic factors of human development and more in one factor than the other. There are only six such combinations, specifically:

Gambia > Sierra Leone - (1)

Eritrea > Niger - (2)

Eritrea > Burkina Faso - (3)

Mozambique > Burkina Faso - (4)

Guinea > Burundi - (5)

Gambia > Burundi - (6)

I then assessed whether there were any dominance relationships among three country combinations with the character of A > B, B > C, which would make A > B > C. I determined that there is no dominance relationship among three countries.

Then I did the same operation adding 5 more countries from the bottom of the HDI ranking. I found the following two additional bilateral relationships:

Chad > Mozambique - (7)

Chad > Guinea - (8)

From (4) and (7), and (5) and (8), we obtain two dominance relationships ranking three countries, specifically:

Chad > Mozambique > Burkina Faso - (9)

Chad > Guinea > Burundi - (10)

Still, there were no combinations of countries with dominance relationships involving more than three countries using these 15 countries.

Finally, I added 5 more countries from the lowest part of the HDI ranking and repeated the operations. Here, I found 5 more bilateral relationships, specifically the following:

Senegal > Eritrea - (11)

Sudan > Eritrea - (12)

Bhutan > Eritrea - (13)

Sudan > Mozambique - (14)

Angola > Guinea - (15)

From the combinations of inequalities (2) and (11), (3) and (11), (2) and (12), (3) and (12), (2) and (13), (3) and (13), (4) and (14), and (5) and (15), we obtain 8 more dominance relationships involving three countries, specifically:

Senegal > Eritrea > Niger - (16)

Senegal > Eritrea > Burkina Faso - (17)

Sudan > Eritrea > Niger - (18)

Sudan > Eritrea > Burkina Faso - (19)

Bhutan > Eritrea > Niger - (20)

Bhutan > Eritrea > Burkina Faso - (21)

Sudan > Mozambique > Burkina Faso - (22)

Angola > Guinea > Burkina Faso - (23)

There are still no combinations of inequalities with dominance partial orderings of more than three countries.

The result of the examination above can be summarised as follows:

1. Among the countries with the lowest 10 HDI values, there are only 6 bilateral dominance relationships, and none among more than two countries.
2. If we include 5 additional countries (ranked from the 11th to the 15th lowest in the HDI ranking), there are altogether 8 bilateral dominance relationships, and 4 of these 8 bilateral relationships form 2 sets of dominance orderings of three countries. There are no dominance relationships involving more than three countries.
3. If we include 5 further countries to have all of the lowest 20 countries in the HDI ranking, there are altogether 13 bilateral dominance relationships, and 11 of these 13 bilateral relationships form 10 sets of

dominance orderings of three countries, but still there are no dominance relationships among more than three countries.

The general conclusion from the above examination is that if we abandon any relative evaluation of the sub-dimensions of the HDI, the number of countries we can order becomes very restricted. For example, there is a total of 45 bilateral relationships possible for the 10 countries at the bottom of the HDI ($10 \times 9 / 2 = 45$), but, as discussed above, there are only 6 bilateral relationships among these 10 countries ranking them according to partial dominance ordering. This and the other analyses above show as follows:

- Firstly, the HDI rankings are largely dependent on weighting the four variables in the HDI.
- Secondly, rankings that are not too partial (say, rankings with more than three countries) are not easy to obtain amongst these countries with the lowest levels of the basic capabilities.
- Thirdly, in some cases we can usefully compare human development relative to one or two other countries without making any relative evaluation among the three dimensions of the HDI. For example, as noted above, we know that both Guinea and Gambia are dominantly advanced in human development (measured in terms of the three dimensions, four variables) over Burundi. Accordingly, these measures of human development can be sufficient in some cases, for example, for some decisions that international organisations may have to make for resource allocations among these countries.

In sum, in order to obtain a complete ranking of 174 countries with respect to their degrees of human development, it is critical that we use an aggregate index with

relative weights assigned to the components. On the other hand, for some purposes we do not necessary need a complete ordering. Therefore, the decision to use a composite index instead of a set of individual indexes is fundamentally based on whether, as a practical matter, a more complete ordering of the countries' human development is needed.

2-3-1-5: Comparability across time

This aim – that the HDI provides comparability in rankings over time – was added to the human development project four years after the publication of the first Human Development Report in 1990. One of the consequences of adding the desire to make intertemporal comparisons of the HDI was to fix the minimum and maximum values of the denominators when normalising the three dimensions, each of which has a different unit of measurement. Prior to the 1994 Report, the minimum and maximum values of each variable that determined the common denominators used in the normalising process varied from year to year. The change to fixed values was made because fixed denominators are regarded as more suitable for making intertemporal comparison. That is, if the denominators change over time, this might cause an anomaly in which, for example, one country's actual life expectancy increases but their HDI longevity score goes down because the denominator increases due to changes in the life expectancy of another country (either at the bottom or at the top of the scale, for example, because a war in one of the poorest country has worsened its life expectancy).[38]

Note, however, that the interest in fixing the denominators in the normalising process arises essentially from a desire to compare the HDI *values* over time. There is

[38] See Dasgupta (1993), and UNDP (1993), p. 108.

no need to fix common denominators if, instead, the HDI is used for comparisons over time of HDI *rankings*, rather than *actual values*. If denominators are allowed to vary, changes in the HDI value (or the value of one of its sub-dimensions) of country *A* can happen merely because of a change in the performance of a county *B* at the bottom (or top) of one of the sub-dimensions of the HDI. But this change would affect the ranking of country *A* only if HDI values of country *B* go beyond or below those of country *A*. And if that happens, than surely their rankings should be changed, that is, the ordering of the two countries should be reversed. Accordingly, the anomaly explained above is irrelevant if we only seek information regarding the HDI *ranking* of a country over time. It is only when we want to use the index for comparisons over time of the HDI *values* that the fixing of common denominators is required.

The period for which the HDI is designed to be effective in making intertemporal comparisons is determined by the way the fixed minimum and maximum values of each variable are chosen. Those values are chosen by tracing records (and estimates) of 30 years into the past and 30 years into the future. According to the Report, the choice was made, firstly, because there is no internationally comparable data that are reliable for a period longer than 30 years. Secondly, in order to make comparisons over time for the foreseeable future, we need to look not only retrospectively but also prospectively at likely minimum and maximum values.[39]

Another reason to limit the fixing of minimum and maximum values to 30 years from now is that scientific development even within the next 30 years may significantly change both the structure of the world and the scientific tools we use for

---

[39] See UNDP (1993), p. 109.

studying it. For example, having access to and being able to make use of the internet may become critical in many dimensions of our lives (health, education, consumption, resource management and employment), so that it may be appropriate to include new minimum and maximum values, or even new dimension(s) to the measurement of human development, even within a few decades from now.[40]

This discussion of the means of normalising the three dimensions of the new HDI is related to my earlier discussion of the measurement of 'real' temperature by Chang, where a thermometer that is stable across circumstances and time is chosen amongst rival instruments.[41] Marcel Boumans, in his paper 'Representation and Stability in Testing and Measuring Rational Expectations,' points out that standardisation is a way to construct a stable instrument. He explains as follows:

> By testing a model one tries to find out to what extent the model covers the data of the phenomenon, while to be a candidate for a measurement formula the model must represent the whole data range. And among the possible representations the standard model represents the most stable correlation under different circumstances.[42]

Boumans uses as an example a mathematical formula that measures supply elasticities in a rational expectations framework.[43] Applying the 'minimalist overdetermination' strategy suggested by Chang,[44] Boumans chooses a formula that (1) provides the most stable relationship between variables in the model with respect to empirical data, and (2) can be applied to a wide domain of values of the variables. In this section we found that the new HDI based on a revised normalisation method is more 'stable' than the previous index because the former can be used for 60 years

---

[40] On this, see for example, The Economist, June 26th – July 2nd 1999.
[41] Chang (1996).
[42] Boumans (1999b), p. 387.
[43] Boumans (1999b).
[44] Chang (1996).

without causing anomalies in the measurement, whereas the old index was in principle only applicable for one year.

2-3-2: The LSI

In this section I will examine the LSI – which, as discussed above, was developed by Fukuda to be used for policy-making to improve the general level of satisfaction in Japan – in the light of the five purpose-dependent criteria I applied to the HDI, specifically: (1) Universality; (2) Differentiation with the GDP; (3) Policy usefulness; (4) Comparability across space; and (5) Comparability across time.

2-3-2-1: Universality

Since the LSI is designed to measure the satisfaction level of those in a single country, Japan, its ability to provide comparisons among different countries is not relevant as a criterion for a successful LSI. However, one can examine this criterion of universality at a different level, that is, within Japan, as applied to the LSI.

As applied to the HDI, I distinguished between two concepts of universality, that is, the issues of whether the index is universally acceptable and whether it is universally applicable. The former issue was relevant to the HDI since the HDI required support from the different member countries of the UN. However, this issue of universal acceptability is not relevant to the LSI, since this was not dependent on governmental approval or support.[45]

The issue of universal applicability is relevant to the LSI. Specifically, the question of universal applicability as applied to the LSI is essentially whether the

[45] Although Fukuda is a government official, his development of the LSI was included in the privately funded publication not a publication funded by the government.

index is valid for the whole Japan, not for an average individual or for an individual in one town. This question can be answered, first, by examining the nature and quality of the sample used in the public opinion survey conducted by the Japanese government that was the principal basis for the degree of satisfaction level (DSL) data used to develop the LSI. That is, this first issue is whether or not the sample used for this survey is representative of the range of people and groups in Japan. It seems quite likely that this survey is so representative in most respects, since the sample is substantial -- a sample size of 10,000 – and is collected from across Japan. Thus, the LSI appears based upon a sample that is representative of the whole Japan in terms of regions within the country. But the sample is not necessarily representative of the whole Japan in terms of all different age groups, because the sample does not include information about Japanese people under 20.

A second issue related to the LSI's universality is whether the index measures a concept that is relevant for all the individuals in Japan. Here again, it appears that the LSI's focus on satisfaction satisfies this standard. The basic concept of satisfaction, leaving aside the issue of whether or not one believes it to be *the* essential goal of human endeavour, can be regarded as relevant for all the individuals in Japan. So the LSI appears relevant to individuals throughout Japan, and in this sense as well appears to meet the criterion of being universally acceptable across the domain that is relevant, that is the country of Japan.

2-3-2-2: Discrimination from a GDP measure

As with the HDI, an essential purpose behind the construction of the LSI was to find a measure that is different from the GDP. As discussed above, Fukuda discovered that in fact the improvement of the GDP in Japan had not been making the

Japanese people in general more satisfied, and therefore he tried to find a measure of well-being that would better explain the general satisfaction level of the Japanese public. Accordingly, as long as the index is successful in achieving its primary aim, it should give results that need not coincide with those of a measure using GDP.

2-3-2-3: Policy usefulness

This factor of usefulness was at the core of the LSI, since Fukuda's basic motivation was to develop an index that would permit the development and improvement of governmental policies to increase satisfaction levels of the Japanese public.

The LSI should be successful in meeting this criterion for two reasons. Firstly, most if not all of the five areas of life reflected in the LSI were categorised in a way that is easily influenced by government policies. As noted above, these areas are employment, housing, consumption, education and leisure, and in all of these areas the government can implement programmes to affect the areas. For example, even in the area of leisure, the government can improve satisfaction to some extent by policies to encourage fewer working hours. For this reason, the LSI does not include purely private aspects of a person's satisfaction with life, which are not subject to government intervention. For example, there is little, if anything, that a government can do to help people to meet and fall in love, or to encourage a husband and a wife to help one another. Although there are other things that government can do, for example, to encourage good family relationships through provisions of childcare, counselling, or new tax schemes, generally the five areas used in the LSI are more obviously areas where government policies can make a difference.

A second reason why the LSI should be useful for policy-making involves the way the variables were chosen. As discussed above, Fukuda used statistical methods to identify the variables used in the LSI. Accordingly, this process provided empirical evidence that changes in the components of the LSI might be associated with improving the level of 'the degree of satisfaction in life in general' (DSL).

The above analysis of the policy usefulness of the LSI contrasts with that of the HDI. As explained above, the HDI was designed to measure only the main aspect of choice enlargement and a limited number of dimensions was included so that policy makers could easily follow changes in the levels of all the sub-indicators. Accordingly, the HDI can be used as a guide to the degree of human development of a country, but the particular structure of the HDI cannot be used to develop strategies for government policy. In the case of the LSI, however, the index was designed to find important policy variables that it was demonstrated would affect the level of the DSL. Accordingly, the numbers of variables (sub-indexes) did not have to be small in order to be useful for policy purposes. Put another way, the LSI was designed to identify the policy factors that would improve what the index measures, that is the general satisfaction level in the Japanese society, thus making the use of the LSI for policy purposes more direct than that of the HDI.

2-3-2-4: Comparability across space

The LSI is constructed for a single country – Japan – based on particular data sets obtained from the country. Accordingly, the index is not designed to be applied anywhere outside Japan. Further, the index is not constructed to provide analysis between different regions of Japan either, because the LSI does not have sub versions

for regions of Japan. The LSI, therefore, has no capacity to compare or discriminate across space, but nor was it designed to do that.

2-3-2-5: Comparability across time

As applied to the LSI, the criterion that the index provides comparability over time amounts to asking whether the LSI is robust enough not to break down over time, which Fukuda found had occurred with respect to the capacity of real income to measure human well-being or satisfaction. Such robustness requires both a stable index, that is, one that will provide identical results over time if the circumstances remain stable, and a flexible index, that is, one that has the ability to reflect differences in circumstances over time. The LSI was constructed in such as way as to meet both of these robustness requirements. As explained above, the LSI is based upon variables that are regularly updated to correlate with the general level of satisfaction as reflected in yearly public opinion polls. Accordingly, the index is stable in its ability to measure satisfaction in stable circumstances, and is also flexible in that the index can reflect changes over time in both the variables and the degree to which they affect the general satisfaction level.

Chapter three: Theory-dependent criteria for the HDI and the LSI

This chapter discusses the HDI and the LSI in terms of the criteria of successful indexes that are derived from the indexes' relation to theory. I will start by explaining the strong link between the HDI and a well-being evaluation theory called the *capability approach* developed by Amartya Sen. The next section explains that the relativistic and absolutist views on poverty are not incompatible by examining the ideas of Peter Townsend, who is a proponent of the former. In this section I conclude that successful evaluation of well-being requires assessment of both relative and absolute aspect of need. I then show that the traditional well-being evaluation approach (called *commodity-based approach*) is incapable of revealing necessities that are relative to individuals. The following section shows that, on the other hand, the evaluation based on Sen's capability approach is able to reveal such 'relative necessities' of individuals, although the theory is not sufficient in implementing the actual measurement of 'capabilities'. The following section examines the HDI with respect to the capability approach by Sen. Specifically, I will investigate the HDI's link to the capability approach focusing on an aspect that is not considered by the Human Development Report: that is, I examine the index's ability to identify necessities of individuals that are relative to their personal characters. The rest of the chapter is devoted to examining the LSI. As explained in Chapter one, Bentham-like ideas of pleasures and pains do have an effect on the LSI, but the effect is at a very general level. So I will discuss just how close the LSI is to Bentham's original ideas briefly at the end of the chapter.

3-1: Relationship between Amartya Sen's capability theory and the HDI

This sub-section of the chapter is organised as follows: first, I will briefly describe Amartya Sen's capability theory, then I will explain the historical link between Sen's capability theory and the human development project from which the HDI was developed. The third section explains what 'relative necessities' mean and why they are important in evaluating the well-being of individuals. The fourth section explains the limit in the 'commodity-based approach to well-being' (the more traditional method of evaluating well-being) in that the approach does not reveal the relative necessities of individuals. The fifth section shows that one of the strengths of capability theory is its ability to identify the relative necessities of individuals. The section also identifies some problems to be solved before capability theories can be put into a measurement of well-being. The final section examines how effective the sub-indexes of the HDI are in revealing the relative necessities of individuals.

3-1-1: Sen's capability approach to evaluating well-being

Sen's capability approach involves the identification of what Sen calls the 'functionings' of individuals instead of characteristics of commodities, and the examination of people's well-being by evaluating either people's functionings themselves or people's sets of functionings called their 'capabilities.' In this section I will explain functionings, commodities and capabilities and their relation to the concept of well-being.

In his book 'Commodities and Capabilities' Sen defines functionings and distinguishes these from commodities as follows:

> *A functioning is an achievement of a person: what he or she manages to do or to be.* It reflects, as it were, a part of the 'state' of that person. *It has to be distinguished from the commodities which are used to achieve those functionings.* For example, bicycling has to be

> distinguished from possessing a bike. It has to be distinguished also from the happiness generated by the functioning, for example, actually cycling around must not be identified with the pleasure obtained from that act. A functioning is thus different both from (1) having goods (and the corresponding characteristics), to which it is posterior, and (2) having utility (in the form of happiness resulting from that functioning), to which it is, in an important way, prior.[1]

According to this definition, a functioning is a part of the state of a person. Functionings are different from commodities because 'functionings are features of the state of existence of a person, and not detached objects that the person or the household happens to "produce" or "own".'[2] A functioning relates to commodities in that a functioning is achieved by a person's ability to make use of relevant commodities: for example, the functioning 'cycling' is achieved by the combination of a person's ability to cycle and the availability of a bicycle. Because functionings are states of existence of a person, functionings are also described by Sen as 'doings and beings.'[3] I will be referring to this expression often in the following discussion because it is a simple and accurate characterisation of functionings.

The capabilities of a person, on the other hand, reflect 'various combinations of functionings ("beings") he can achieve.'[4] That is, a person's capabilities reflect the range of functionings that the person is able to perform or to be, for example, cycling, swimming, reading, enjoying, being satisfied, etc. As Sen also put it, capabilities reflect or represent 'the person's freedom to lead one type of life or another.'[5] By this he means that a person is free to lead one type of life or another to the extent that his capabilities (that is, his range of functionings) permit a range of possible doings or beings.

---

[1] Sen (1985), pp. 10-1, emphases added.
[2] Sen (1985), p. 15.
[3] Sen (1992).
[4] Sen (1985), p. 14.
[5] Sen (1992), p. 40.

Sen found it useful to clarify the relationships between capabilities, functionings, commodities and their impact on people' s choices using the following notations:[6]

$\mathbf{x}_i$ = the vector of commodities possessed by person *i*,

$c(\cdot)$ = the function (not necessarily linear) converting a commodity vector into a vector of the range of characteristics of those commodities,

$f_i(\cdot)$ = a personal 'utilisation function' of *i*, reflecting one pattern of use of the commodities that *i* can actually make in generating a functioning vector out of a characteristic vector of the commodities that *i* possesses, given *i*'s personal features and *i*'s space and time location,

F_i = the set of 'utilisation functions,' f_i, being any one that person *i* can choose.

If person *i* has the utilisation function $f_i(\cdot)$, then with his or her commodity vector $\mathbf{x}_i$, person *i*'s achieved functionings will be given by the vector $\mathbf{b}_i$, as follows:

$$\mathbf{b}_i = f_i(c(\mathbf{x}_i)).$$

If $v_i(\cdot)$ is the valuation function of person *i*, then the value of the vector of functionings $\mathbf{b}_i$, is given by

$$v_i = v_i(f_i(c(\mathbf{x}_i))).$$

According to the Sen, $v_i(\cdot)$ can be person *i*'s valuation function using either a certain objective standard, or the subjective standard of the person himself.[7] But the evaluation represented by $v_i(\cdot)$ is not the same as an evaluation based only on how happy the person is, regardless of whether happiness is assessed objectively or

[6] The following notations and formulations of them are taken from Sen (1985), pp. 11-4.
[7] Sen (1985).

subjectively.[8] Given that there is a set of utilisation functions available for a person (F_i), the functionings vectors feasible for the person for a given commodity vector $\mathbf{x_i}$ are given by the set P_i ($\mathbf{x_i}$), as follows:

$$P_i(\mathbf{x_i}) = [\, \mathbf{b_i} \mid \mathbf{b_i} = f_i(c(\mathbf{x_i})), \text{ for some } f_i(\cdot) \in F_i \,]$$

If the person's choice of commodity vectors is restricted to set X_i, then the person's feasible functioning vectors are given by the set Q_i ($\mathbf{x_i}$), as follows:

$$Q_i(\mathbf{x_i}) = [\, \mathbf{b_i} \mid \mathbf{b_i} = f_i(c(\mathbf{x_i})), \text{ for some } f_i(\cdot) \in F_i \text{ and}$$

$$\text{for some } \mathbf{x_i} \in X_i].$$

If we want to express a person's feasible functioning vectors in terms of the person's 'freedom of choice,' then we can say that Q_i ($\mathbf{x_i}$) represents a person's freedom of choice over functionings, given his personal features F_i (possible conversion of characteristics into functionings) and his access to commodities X_i. As noted above, 'capabilities' reflect a person's freedom to choose from different types of life that are made possible by different combinations of doing and beings. Q_i ($\mathbf{x_i}$), therefore, represents the 'capabilities' of person *i* given his personal features and his

---

[8] Sen (1985) notes that 'while $h_i(\cdot)$ [the happiness function of person *i* that is related to the functionings achieved by *i*] is also a scalar-valued function - - - , we should not fall into the trap of assuming that the evaluation of how good b_i is (i. e. how high the 'well-being' happens to be) must be given by the corresponding u_i [which is an evaluation function of $h_i(\cdot)$ defined as $u_i = h_i(f_i(c(x_i)))$]. The function h_i just tells us how happy the person is with the functioning vector b_i, and it does not tell us how good that way of living is, or even how good person *i* himself thinks it is. Whether or not happiness is a plausible criterion of the goodness of a life - - -, *valuing* a life and measuring the *happiness* generated in that life are two different exercises.' (12) According to Sen (1985), therefore, if v_i is an objective valuation function, then two individuals with the same *x*, $c(\cdot)$ and $f(\cdot)$ will have exactly the same *v*'s. If on the other hand, v_i is a subjective valuation function, then even two individuals with the same *x*, $c(\cdot)$ and $f(\cdot)$ may have different *v*'s. Some individuals' subjective valuation functions may have only one criterion of 'happiness', but this is by no means the only way individuals assess their own achieved functionings (b_i's). Besides, how to measure happiness may differ from person to person, that is, different individuals may have different *u*'s. An objective standard of evaluating b_i can in principle be happiness only, but for this to happen, the claim that the only criterion of the goodness of life is happiness must be accepted. If v_i is a subjective evaluation function, and if we use it for the measurement of well-being, then necessities that are identified by the measurement are relative not only to personal features, space and time, but also to the view each person has about how to evaluate good in his or her own life. It is only when v_i is an objective evaluation function and if we take the standard of evaluation as given (fixed or absolute) that necessities identified by v_i are relative only to personal features, space and time.

access to commodities; it reflects the range of functionings (doings and beings) the person can achieve.

Given the valuation function $v_i(\cdot)$, it is possible to characterise the values of functionings that the person can possibly achieve, given by the set V_i,

$V_i = [\, v_i \mid v_i = v_i(\mathbf{b}_i)$, for some $\mathbf{b}_i$ in $Q_i]$.

A further important issue is how functionings and capabilities are related to a person's well-being, since the latter is what we are ultimately interested in measuring. On this issue, Sen notes, reasonably, that 'functionings are *constitutive* of a person's being, and an evaluation of well-being has to take the form of an assessment of these constituent elements.'[9] That is, these functionings are part of the person's beings, so that if we want to evaluate the person's well-being, we need to do so in terms of his range of functionings. Therefore, to evaluate a person's *actual* well-being using the above notations, we should use the valuation function, v_i, that represents the valuation of person *i*'s achieved functionings given his or her commodity vector $\mathbf{x}_i$ and personal features f_i.

On the other hand, if we want to measure the person's *capability* for achieving well-being, the appropriate measure is the valuation of a set of functionings that a person can possibly achieve, which is V_i in the above notation. By knowing the person's capability set rather than only his actual use of certain functionings, we can distinguish, for example, whether the person is not eating for reasons other than a lack of food (for example, in order to fast) or just does not having anything to eat, when such a distinction is of direct interest. This is an advantage of the measurement of well-being using a capability set (V_i) rather than achieved functionings ($\mathbf{b}_i$): the

[9] Sen (1992), p. 39.

former allows us to evaluate whether or not there were alternatives available to the state the person actually achieves in the latter. The fact that the capability set is analytically superior does not of course, mean that it is always possible. Faced with the difficulty in practice in measuring the capability set, as Sen notes, 'one might have to settle often enough for relating well-being to the achieved – and observed – functionings, rather than trying to bring in the capability set.'[10]

3-1-2: A historical relationship

There is a clear historical relationship between the HDI and Amartya Sen's capability theory. As Meghnad Desai notes in an article on the HDI, 'it is from this twenty year old literature on inequality and poverty and especially the notion of capabilities that the concept of human development traces one of its strongest roots.'[11] The following examples from the Human Development Reports and the related background papers reflect the strong link between Sen's capability theory and the concept of human development:

> Human Development has two sides: *the formation of human capabilities* – such as improved health, knowledge and skills – and *the use people make of their acquired capabilities* – for leisure, productive purposes or being active in cultural, social and political affairs.[12]
>
> * * *
>
> The three dimensions of the HDI relate to one or many *capabilities that they are expected to capture.* Thus, longevity captures the *capability* of leading a long and healthy life. Educational attainment captures the *capability* of acquiring knowledge, communicating and participating in the life of the community. Access to resources needed for a decent standard of living captures the *capability* of leading a healthy life, guaranteeing physical and social mobility, communicating

[10] Ibid., p. 52.
[11] Desai (1991), p. 352.
[12] UNDP (1990), p. 10, emphases added.

> and participating in the life of the community (including consumption).[13]
>
> * * *
>
> Do they have the *capability* to live long? Can they avoid mortality during infancy and childhood? Can they escape preventable morbidity? Do they avoid illiteracy? Are they free from hunger and undernourishment? Do they enjoy personal liberty and freedom? These are the basic features of well-being which derive from looking at people as the centre of all development activity. *Enhancing their capabilities to function in these elementary ways is what lies at the core of human development.*[14]

One can see a strong link between the capability approach and the human development concept in terms of the importance of freedom of choice. As detailed above, the HDI focuses on human development as a process of enlarging people's choices. That is, a central idea behind the human development project is that the larger the opportunities to choose from, that is, the larger the degree of freedom of choice, the better. Similarly, capability approach naturally leads to evaluating people's freedom of choice. As detailed above, the most significant aspect of the capability approach to well-being is that it does not focus on the amount of materials that individuals possess (but are not necessarily able to make use of), but the range of things individuals can actually do with those materials. As Sen puts it, the 'capability to achieve functionings (i.e. all the alternative combinations of functionings a person can choose to have) will constitute the person's freedom – the real opportunities – to have well-being.'[15]

---

[13] UNDP (1993), p. 105, emphases added.

[14] Anand and Sen (1992), quoted in UNDP (1993), p. 107, emphases added.

[15] Sen (1992), p. 40.

3-1-3: Relative aspects of well-being and absolute criteria for the measurement

One theoretical issue that arises in assessing well-being concerns the extent to which this assessment should be based upon relative or absolute measures of what is 'needed.' As detailed below, I conclude that a proper examination of well-being requires assessment of both relative and absolute aspects of need. Specifically, I define measurements of well-being that successfully take into account 'relative necessities' of individuals as measures that are able to reveal how much an individual with particular personal characters who lives in a particular place and time needs in order to achieve a certain absolute standard of well-being.

Peter Townsend, who is a proponent of a relativistic view of poverty, claims as follows:

> Any rigorous conceptualisation of the social determination of need dissolves the idea of 'absolute' need. And *a thorough-going relativity applies to time as well as place.* The necessities of life are not fixed. *They are continuously being adapted and augmented as changes take place in a society and in its products.* Increasing stratification and a developing division of labour, as well as the growth of powerful new organisations, create, as well as reconstitute, 'need'.[16]

Townsend's comments emphasise that the 'necessities of life' change even for a person with given personal features because of changes in 'society and in its products' as well as changes in the person's 'place.' This circumstance can be regarded as involving changes *relative to time and space.* However, Townsend does not address a different range of necessities that occur even where time and space are given. Specifically, necessities will also differ depending on whether the person is old/young, male/female, fit/disabled, etc. This latter kind of necessities I describe as *relative to personal features.* Accordingly, in order to evaluate the necessities of life we must consider them relative to time, space *and* to such personal features. In the

[16] Townsend (1979), pp. 17-8, emphases added.

following, I assume that when we talk about necessities of life being relative, they are relative with respect to time, place and personal features. Of course, a difficult question is which personal features to include as the basis for differentiating individuals. For example, if we take strength of the will as a relevant personal feature in judging people's necessities, individuals with weaker will who suffer disadvantages caused by the weakness can be viewed as 'needy,' whereas if we do not include such a personal feature as relevant, those with weaker will not be so viewed. The discussion below involves some aspects of the problem regarding which personal features to include.

A further issue is whether this relativistic view is adequate to assess well-being. I agree with Townsend that that there is a relative aspect to well-being, because people's needs are relative to who they are and where and when they live. However, as I show below, a relativistic view on poverty is not incompatible with the recognition of an absolutist component. In fact, I believe that there must be added to this relativistic view on poverty some absolutist aspect in order properly and effectively to evaluate well-being. I believe this is required for two types of reasons. First, although there is a relative aspect to well-being, there is a point at which well-being (or more precisely its lack) can be put in objective terms. Sen (1983) expresses this idea concerning poverty in explaining that there is 'an irreducible absolutist core in the idea of poverty.' That is, there are circumstances that are objectively so desperate that individuals suffer from starvation and malnutrition. Accordingly, even if all the individuals in a particular place and time experience these conditions (and so people are not 'relatively' poorly off), we can agree that they are in poverty. In this sense at least the notion of well-being cannot be seen in purely relativistic terms.

A second reason for the need for some absolute considerations relating to measures of well-being pertains to the need to resolve competing claims for resources. We can see this need in the following example. Let us assume that two individuals from different regions of a country with different circumstances make the following claims of need to the national welfare officer:

1) A (with personal features F_1 , F_2 and F_3) in region R at time T needs X.

2) B (with personal features F'_1, F'_2 and F'_3) in region R' at time T' needs Y.

Let us also assume that the welfare officer wants to act so as to improve the overall well-being in the country, but that the government does not have enough resources to satisfy both claims. In such circumstances, information given in the two sentences above is not enough to resolve which need will be satisfied, because it does not say why A needs X or why B needs Y, or how important it is that A gets X or that B gets Y.

So the two individuals add some information to their original claim to specify what they mean by 'A needs X' or 'B needs Y.' Their revised claims may look like the following:

1') A (with personal features F_1 , F_2 and F_3) in region R and time T needs X in order to H (or to enable A to H).

2') B (with personal features F'_1, F'_2 and F'_3) in region R' and time T' needs Y in order to I (or to enable B to I).

In order to decide which of the two individuals' needs will most affect the overall well-being and thus should be satisfied from the public resources, the officer has to find a common ground for comparison between those claims. That is, the officer has (1) to identify the types of factors or outcomes that are important to well-being – for example, whether the considerations H and I in the above example are

such factors – and (2) to have some method of quantifying the results so that comparisons can be made over time, place and individual circumstances.

Therefore, if we want to make inter-personal comparisons of well-being over a number of people, even if we want to reveal necessities relative to space, time and personal features, we need reference points for comparison; we need to find widely acceptable ideas about the 'needs' that relate to improvements in well-being for individuals regardless of where and when they live and what are their personal characters. This is the second area in which relative necessities are related to absolute criteria for well-being.

3-1-4: The commodity-based approach and its limit

In this section I will explain why one of the most commonly used approaches to the evaluation of well-being across individuals – the 'commodity-based approach' – does not meet the requirement stated above that account be taken of relative aspects of well-being.

The commodity-based approach evaluates the well-being of individuals in terms of the commodities to which they have access. Under the commodity-based approach, commodities are seen in terms of their characteristics, that is, the various desirable properties each commodity has.[17] For example, a certain amount of food is seen as giving the owner access to the properties or benefits of the food, including preventing him from starving, yielding nutrition, creating pleasure, and providing opportunities to socialise with others. One version of the commodity-based approach, for example, is the 'basic needs approach' used by the International Labour Organisation (ILO), which evaluates whether or not a particular society provides basic

[17] Sen (1985), p.9.

needs defined in terms of certain levels of commodities. Broadly speaking, indices based on income figures are also versions of the commodity-based approach, assuming that income is a proxy for how many commodities the income can buy.

A commodity-based approach to the measurement of well-being falls short of meeting the condition required for the type of measurement we are seeking: a measurement that reveals necessities that are relative, and that can be used for inter-personal comparisons of individuals or inter-country comparisons of countries. Under a commodity-based approach, for example, individuals are regarded as equally well-off, regardless of their differences, as long as they have the same amount of commodities. However, this result is too simplistic. For instance, a certain amount and type of food may be enough to prevent a healthy adult from starving, but the same amount and type may not be adequate if the person has a deficiency in his digestive mechanisms, as may occur, for example, in many developing countries. Similarly, that a person possesses a pair of trousers and a T-shirt may be enough for us to know that he can be protected from the weather and can socialise with others without shame if he is a missionary in a South American forest, but not if the person is a banker who works in the City. More broadly, a commodity-based approach may show that two countries have the same GDP, but this may be very misleading if, for example, one of the countries is in a very temperate climate where few resources are necessary to provide heat and protection from the elements, but another country is in a very harsh climate where substantial resources are needed just to provide these circumstances.

Thus, for a measurement of well-being to be effective, it must be able to capture differences in the *use* individuals can make of commodities that result from differences in time, space and personal characters. In order for the measurement to do so, it has to shift its attention from commodities to the relationship between

commodities and individuals, or more specifically, to individuals' achievements in making use of commodities in ways that enable the individuals to satisfy a certain standard of well-being.

3-1-5: The capability approach and its ability to reveal relative necessities of individuals

Now I will explain why the evaluation of functionings and capabilities is able to reveal relative necessities. As the earlier description of the capability approach shows, functionings reflect what a person is able to do with commodities, and capabilities represent the aggregate of such functionings. Because functionings are part of a person's being and some of them constitute his well-being, if we focus on those (well-being) functionings, we can identify and measure differences among individuals in their ability to turn commodities into things that constitute a part of their well-being.[18] What use a person can make of commodities to achieve such functionings depends on his personal features as well as his particular environment. Therefore, if a person's well-being is evaluated in terms of the person's set of functionings, the evaluation will take into account differences in individuals' personal features and other relevant differences. For example, a person who is less capable of making use of a given set of commodities because of his particular personal features would be identified as having lower well-being than those who are more capable. The use of functionings or a capability set therefore captures differences in well-being that are relative to personal features and the surrounding environment, differences that the commodity-based approach to well-being does not reveal.

---

[18] As I noted above, the core idea of Sen's capability approach to well-being is that we should focus on the evaluation of individuals' functionings rather than their access to commodities, because the former is capable of identifying relative necessities that the commodity-based approach cannot.

3-1-5-1: Problems in implementing the capability approach to the measurement of well-being

I find that capability theory itself does not provide answers for the following problems that must be solved before the measurement of well-being in terms of functionings (or sets of them which are capabilities to function) can be practically applied:

(1) Functionings include not only those associated with well-being but also those associated with ill-being, and for some functionings it is not clear to which category they belong;

(2) There are many different states that could be considered as states of well-being, so a choice has to be made about which functionings to use to evaluate well-being;

(3) Depending on what to regard as a reason for a person's not achieving a certain functioning, there are many different necessities that can be revealed, and a decision has to be made about how far 'back' to go in searching for the reason;

(4) The choice of functionings implicitly determine whose relative necessities to take into account and whose not to;

(5) There are problems in making interpersonal comparisons of well-being; and

(6) There is a need to aggregate different functionings when we want to measure overall levels of well-being.

I will explain each problem in turn.

Problem (1): As detailed above, the functionings that underlie the capability approach are constitutive of being and include a range of possible actions or beings,

such as eating, seeing, thinking, moving, breathing, being depressed, having a pain, and worrying. Because they are constitutive of being, functionings include not only those associated with well-being but also those associated with ill-being, such as the last three functionings in the previous sentence. In order to evaluate the well-being of individuals, however, we need to use only those functionings that are constitutive of well-being rather than ill-being, or we need to include negative weights to the latter (that is, for example, to consider someone better off if a functioning of ill-being, such as worrying, declines). Either way, we need a clear distinction between the two types of functionings. However, some functionings are not clearly related to well-being or ill-being. For example, the functioning 'working' can be related to well-being or/and to ill-being. One way to deal with this problem is to summarise functionings into more general ones that are either well-being or ill-being, such as 'being productive,' rather than 'working.'

Problem (2): This problem involves the fact that there is potentially a very large number of functionings defined by the capability approach that could be said to constitute states of well-being. However, obviously we need to identify a finite number of functionings in order practically to use them to measure well-being. This second problem has two aspects. Firstly, we want a finite number of functionings to cover broad aspects of well-being, and secondly we have to decide which aspects of well-being to take into account and which ones to leave out. The answer to the first aspect of the problem is similar to that for Problem (1). Specifically, to be practical, the measure of well-being should include functionings that are general rather than specific, such as 'being fit' rather than the aggregation of functionings such as 'running for 3 km,' 'walking for 30 minutes,' 'lifting a 10 kg object,' etc. The second aspect of the problem has to be solved by a particular index-maker's decision based

on some criteria, such as the UNDP's decision to use the three dimensions for the measurement of human development based on the consideration of universality or Fukuda's decision to use the five areas of life based on statistical results.

Problem (3): When we recognise the need to identify necessities as relative to time, space and personal features of individuals in the capability approach, we confront the issue of how detailed we can be in recognising such relative features. One particular issue is how far back in a logical progression of such personal features or circumstances we should proceed. For example, consider the progression of circumstances that could reflect an individual's achieving the functioning 'taking minimum calories for relevant age and sex.' A particular individual may not be able to achieve this functioning, first, because of a lack of sufficient income to purchase food, but this lack could be not because the family's aggregate income is insufficient but rather because of a second reason, for example, that a family member spends most of the disposable income for drinking; this family member may have started to drink excessively because he was fired from his job, a third reason; he may have been fired from his job because - - -, and so on. Alternatively, a person may not get the minimum intake of calories because he cannot get to town to buy food; he cannot get to town for grocery shopping because he has a problem with his legs; he is disabled because - - -, and so on.

What is the problem in having many possible reasons why individuals could fail to achieve a certain functioning? The problem is that we have to decide which circumstances to use to reveal the individuals' relative necessities. Using the first example discussed above, the individual's relative necessities could be additional income to buy him enough food, if we take only the first reason; the relative necessities would be solving the family member's drinking problem if we take the

second reason; and it would be for the family member to find a new job if we take the third reason. In assessing individuals' relative necessities, we need to decide at which level to focus and how far back to go in looking for the individual's circumstances. Without some narrowing of the 'relativist' considerations, it would be impossible to construct a measure of well-being that could be meaningfully applied on an aggregate basis, for example, to decide on the allocation of scarce resources. We need an independent criterion for this narrowing of relativist considerations.

Problem (4): The fourth problem is that because different functionings are sensitive to different information on relative necessities, a decision to use a particular functioning rather than another for inter-personal comparisons of well-being implicitly involves a decision about whose necessities to take into account and whose not. To use the previous example, the functioning 'taking minimum calories for relevant age and sex' distinguishes people who do not have income to buy food that provides the minimum calories or do not have access to such food for various reasons (such as children of mothers who do not have the knowledge about the minimum calories necessary for children), from those who can buy or have access to such food. However, this functioning – 'taking minimum calories for relevant age and sex' – does not distinguish those who have problems with their digestion mechanisms from those who do not have such problems in determining whether both have an equal amount of energy. In other words, the functioning 'taking minimum calories for relevant age and sex' does not reveal the relative necessities of individuals who have digestion problems and therefore need more than the standard minimum for their age and sex in order to obtain equal energy. Therefore, by using the functioning 'taking minimum calories for relevant age and sex' as an indicator for health, the evaluation is implicitly ignoring the relative necessities of those who have digestive problems.

One solution to this problem would be to use broader functionings such as 'being healthy.' This broader functioning differentiates not only those who have digestion problems from those who do not, but also other characteristics of individuals and the environments in which they live. For example, the functioning differentiates those who have access to clean water from those who do not if the former suffer from illness because of bad quality water; or those who are exposed to air pollution from those who are not if the former contract diseases because of the polluted air. 'Being healthy' distinguishes people who are healthy from those who are not healthy for all sorts of reasons.

There are, however, possible problems in using such broad descriptions. One such problem is that such descriptions may ignore circumstances in which the societies under consideration may regard the individual as responsible. For example, a society may consider the individual responsible for lung cancer caused by heavy smoking, or for heart-related disease caused by consuming excessive alcohol. The fact that 'being healthy' distinguishes unhealthy from healthy people no matter what the reasons behind the ill-health implies the following: if the functioning is used to evaluate the well-being of individuals, it identifies needs for them to be healthy regardless of who is responsible for the individual's states of well-being (or ill-being).

A second difficulty with a broad description of functioning is the difficulty of measurement, since there often are no adequate ways to quantify such broadly defined functionings. In the case of 'being healthy,' for instance, it would be hard to obtain an 'objective' measure for it. Accordingly, it may be necessary for index-makers to sacrifice analytic completeness in taking into account relative necessities in order to obtain practical results. For example, it may be necessary to use a measure such as the

'availability of minimum calories for relevant age and sex' and ignore those with digestive problems, at least in cases where these problems are not widespread.

Problem (5): The fifth problem is related to interpersonal comparisons of well-being. One way of comparing between well-being of different individuals using functionings is to use the function Q, using the above notion by Sen. The function Q represents a set of a person's feasible functioning vectors. The function Q can be useful in the limited case where one individual has the same functionings (or capability sets) as another, plus other functionings that the latter individual does not have. That is, either $Q_A (x_A) \subset Q_B (x_B)$ or $Q_B (x_B) \subset Q_A (x_A)$. In the former case, we may say that *A* is better-off than *B,* and in the latter case we may say that B is better-off than B. Thus, if we were looking at two individuals whose functionings consist of only having access to museums and exhibitions and to sports facilities, if individual *A* gains access to libraries, then we may say that individual *A* is better-off than individual *B*.

Even in this simplified case, however, we may not be justified in concluding that individual *A* is better-off than individual *B*. This conclusion assumes that if an individual has the same set of functionings as another individual plus additional functionings, the former is better off than the latter, *regardless of his aims and goals in life*. But this assumption may not always be appropriate, for example in a case where *A* has all the functionings that *B* does plus more, but A's functionings are all exercise oriented and *A* wants to write a history book. In this case, we may not want to conclude that *A* is better-off than *B* only because *A* has a larger set of functionings than *B*. We need to consider not only the number of functionings available, but their value to the individual, and for this, again, we need to look into the function V.

A more general aspect of this problem involves the fact that it will not often happen that individuals will have completely overlapping capability sets of functionings. In most cases the capability set of functioning for one person (person *A*) will be different from another person's (person *B*), and *A*'s set of functionings will not include *B*'s set of functionings, or vice versa. (That is, neither $Q_A\ (x_A) \not\subset Q_B\ (x_B)$ nor $Q_B\ (x_B) \not\subset Q_A\ (x_A)$.) In this case we cannot make interpersonal comparison of well-being unless we know, using Sen's notation, whether $Q_A\ (x_A) - [Q_A\ (x_A)\ 3\ Q_B\ (x_B)]$ is more valuable than $Q_B\ (x_B) - [Q_A\ (x_A)\ 3\ Q_B\ (x_B)]$. In order to make this type of comparison, it becomes necessary to attach relative values to different functionings. To use Sen's notation again, we need to be able to determine the V_i function for individual i.

As we see below, interpersonal comparison of well-being is a problem that we always have to confront when deciding the function V, which allows for the comparison between $V_i(F_i)$ and $V_j(F_j)$. The problem involves another area where there is a difficulty in including comprehensive information about the differentiating characteristics of individuals, specifically the individuals' different goals and desires. In this case the difficulty involves the fact that the functionings that are the basis of capability theory might be very differently valued by different people, depending upon their aims in life.

If, for example, the function V is determined by letting each individual evaluate his set of functionings, the problem of interpersonal comparisons involves the appropriateness of the assumption that the subjective evaluation of functionings are comparable across individuals. If, alternatively, the function V is determined by a particular index-maker assigning the relative importance of each feasible functioning of each individual, the interpersonal comparison problem involves the index-maker's

ability (or the belief in his ability) to determine those importance factors. Another type of V, of which the HDI is an example, uses a particular set of functionings with a particular relative weight assigned to each. The interpersonal comparison problem concerning this type of function V is that each functioning might be in fact very differently valued by different people, depending upon their aim in life, and still specific weights for a specific set of functionings have to be chosen. For example, consider the functionings of having access to museums and exhibitions on the one hand and having access to a sports facility on the other. The former would be very highly valued by those who want to become artists while the latter would be highly valued by those who want to become athletes.

Problem (6): The final problem in practically applying capability theory concerns the interest in obtaining an aggregate, overall index of well-being using the set of functionings. The problem is that the capability approach itself does not determine how to combine different functionings to make such an overall evaluation of capabilities. The theory does not determine which type of composite measurement to use, for example, whether to make this composite measure using a product of the functionings or an aggregation of them. Nor does it say anything about the relative weights to assign to the functionings under consideration. Here again using Sen's notation, this is another aspect of the difficulty of determining the V_i function for individual i. [19]

[19] Sen is aware of the fact that in order to use functionings and capabilities for measurement, in practice we must face the problem of selecting and weighting them. He says, 'there is no escape from the problem of evaluation in selecting a class of functionings – and in the corresponding description of capabilities. The focus has to be related to the underlying concerns and values, in terms of which some definable functionings may be important and others quite trivial and negligible.' (Sen 1992, p. 44) Much of the discussion in the rest of the chapter is concerned with the reasons for choosing one type of functionings rather than another, and the effects of doing so.

Let me now summarise the conclusions from the foregoing considerations of various problems in practically measuring well-being by reference to the functionings in the capability sets: Firstly, in order to evaluate the well-being of individuals, we need to use descriptions of functionings that are general enough to be states of clearly well-being, not ill-being. Secondly, because there is a large number of states that can be regarded as states of well-being, we have to decide which of them to include in our measurement of well-being, a process that also encourages the use of general functionings. Thirdly, because of practical concerns, we have to make a number of choices in determining how much to take into account different individuals' relative necessities. Thus, we cannot consider all the background circumstances that may have affected an individual's functionings (for example, a drinking family member as discussed in problem 3), nor can we take into account all circumstances that affect these functionings (for example, an individual's digestive disorder as discussed in problem 4), or an individual's different goals and values (for example, an individual's desire to be an artist as discussed in problem 5). In deciding which functionings to use for the measurement of well-being, thus we must implicitly make decisions on which relative necessities to recognise and which to leave out of consideration. Finally, in order to apply the capability theory to measure well-being, we also have to decide upon a method to obtain an overall level of well-being.

3-1-5-2: Problems in implementing a commodity-based approach to the measurement of well-being

The previous section discussed the limits of a capability-based approach to the measurement of well-being. As detailed below, some, if not all of those problems also apply to a commodity-based approach, particularly such an approach based on the use

of individual commodities rather than an aggregate commodity measure such as the GDP.

The first problem relates to some functionings involving ill-being rather than or in addition to well-being. Under a commodity-based approach, an analogous problem arises because some commodities may directly or indirectly produce ill-effects. For example, having access to electricity may be a very important aspect of well-being, but the production of increased electricity may produce more pollution, which in turn produces ill-being in terms of increased health problems, etc. Similarly, having access to more housing may result in deforestation that will have negative consequences in term of the individual's well-being. More directly, production of weapons may on the one hand provide some measure of protection and security and increase the GDP, but may also facilitate or encourage crime and violence, so that an economy producing more weapons – and thus increasing the production of 'commodities' – may be worse-off. A commodity-based approach to calculating well-being must accommodate or at least be affected by these types of considerations.

The second problem involves the large number of functionings that may relate to well-being. This type of problem would also exist in a commodity-based approach using individual commodities, since there is, of course, a large number of different commodities that may be included in an index. This is not such a direct problem for an aggregate commodity-based approach such as one using GDP, at least for a static measure. In order to obtain comparable results over time, however, it is necessary to adjust for inflation, and here these is a problem in selecting particular commodities to include in a 'basket of commodities' to measure the inflation rate, comparable to the problem of identifying a limited number of functionings to measure well-being.

The third and fourth problems noted above involve the issue of how far 'back' to go or how many individual circumstances to take into account in determining the relevant capabilities. There does not seem to be a similar problem under a commodity-based approach to well-being. That is, in this latter approach, the commodities – either individually or in the aggregate – are by definition, the measurement of well-being, and the issue does not arise of whether and how the commodities can be used by individuals. (As noted above, however, this is a basic reason for concluding that commodity-based approaches are inadequate measures of well-being.)

The fifth problem involves the different goals and desires that individuals may have. This problem also exists in a commodity-based approach involving the use of individual commodities, since different commodities may be important to different individuals, and the selection of the commodities can be significant in determining the actual well-being measured by the commodity-based index. At an aggregate level, that is, if we use the income measure for the inter-personal comparison of well-being, the problem relates to the assumption that individuals can be equally well-off with the equal level of income, regardless of different values and goals in life. Using an example noted above, if both the artist and the athlete have the same income, they will be considered equally well-off under the income measure, even if, for example, using sports facilities in general is more expensive than going to art museums.

Finally, the sixth problem involves the difficulty of combining different functionings into an aggregate measure. This is not a problem for commodity-based approaches, which can rely on the market to value the commodities either individually or in the aggregate.

3-1-5-3: The HDI and the practical problems

Capability theory itself does not give solutions to the problems discussed above. We need criteria outside the theory in order to select and attach values to sets of functionings that measure the well-being of individuals in ways that reveal relative necessities. An analogous point was made in Boumans (1997), where he claimed that successful models are the ones that combine various 'ingredients' – including theoretical notions, mathematical concepts or techniques, stylised fact, empirical data, policy views, analogies and metaphors – in such a way that the model meets some criteria. These criteria are not only theoretical, but could be mathematical, statistical or related to usefulness in policy-making.

Let me consider how the HDI deals with the six problems discussed above, specifically, what are the criteria outside the theory used to resolve those problems and what are the answers provided.

Concerning problem (1), the outside criterion used to resolve the problem is the criterion of universality. As detailed above, this is a part of the purpose-dependent criteria for the HDI, requiring that the index use only very general dimensions of human development, specifically, to lead a long and healthy life, to acquire knowledge, and to have access to resources needed for a decent standard of living. These three dimensions are general enough to be clearly regarded by all the member countries of the UN as a part of well-being that relates to human development and not to ill-being.

Regarding problem (2), the criterion of policy usefulness, which is another purpose-dependent criterion for the HDI, provided a solution to the problem of a large number of potential functionings. The criterion of policy usefulness required the index

to have a limited number of dimensions, and thus only three general dimensions are chosen to cover a broad aspect of human development.

Problem (3) is the problem of how far 'back' to search for the reasons for not achieving a particular state of well-being. The HDI deals with this problem by use of the three basic indicators that do not really address the reasons why individuals achieve a particular level of well-being. Accordingly, under the HDI, this is a problem for the index-users (such as international organisations, national or local government) rather than a problem for the index-maker.

Problem (4) involves the need to decide whose relative necessities to take into account. For the HDI the purpose-dependent criteria determined the three dimensions of the HDI, so that, for example, a dimension related to political and civil rights was not included because it was not universally acceptable. As a result, relative necessities of those who do not have the right to vote are not revealed in the HDI results.

Regarding problem (6) concerning the development of an aggregate measure, the previous chapter showed that in order for the HDI to be comparable across nations, which is one of the purpose-dependent criteria for the index, the HDI must be a composite one. As detailed below, the particular relative weights assigned to the three dimensions of the HDI and the particular method of aggregation are determined partly based on statistical criterion and partly on the idea that those dimensions are 'equally important' for human development. However, as I show in the following chapter, the latter reason is misleading because equal weights (based on the idea of equal importance) are assigned after normalising the three different units of measurement for the different dimensions, and therefore, they are not actually weighted equally. As I will argue, a particular assignment of relative weights by the Report does not 'solve' the interpersonal comparison problem. Rather, the way

relative weights are assigned determines the conditions under which the index is supposed to be used.

3-1-6: Capability approach, relative aspects of well-being and the HDI

In this section, I will examine the link between the concept of human development and capability theory from a point of view that is not directly discussed in the Human Development Report nor in its background papers: I will examine the HDI with respect to its ability to identify the relative necessities of individuals. I find such a theoretical criterion is useful in testing the HDI because of the measurement project's close connection to the capability approach to well-being developed by Sen. As I have claimed in chapter one, how important theories are in constructing an index and in what way they are important depends on the motivation of a particular measurement project. The following sections investigate the HDI with respect to the following questions: 1) what are the functionings or capabilities measured by the selected data? and 2) what are the relative aspects of well-being identified by the data?

3-1-6-1: Life expectancy at birth

The Human Development Report's definition of life expectancy at birth used in the HDI is 'the number of years a new-born infant would live if prevailing patterns of mortality at the time of birth were to stay the same throughout the child's life.'[20] The life expectancy figure is derived from a model that defines a relationship between the following variables: mortality rates (q); the number of persons living at the beginning of an age interval (l); the number of person-years (number of persons

[20] See UNDP (1998), p. 219.

multiplied by years) that would be lived within an age interval (for example of five years) by the cohort of 100,000 infants who are assumed to be born on the same day (L); and the total number of person-years that would be lived after the beginning of an age interval by the cohort of 100,000 infants who are assumed to be born on the same day (T). All of these variables are derived from observed age-specific death rates (m). Therefore, these age-specific death rates are the only empirical data on which the life expectancy figure is based. There are several versions of models from which life expectancy is derived, but the basic theory behind the models that are most frequently used is that life expectancy can be defined as the number of years lived by a hypothetical group of infants born on the same day, whose number reduces each year only in accordance with a set of age-specific death rates that are assumed to be fixed at the values of the starting year.[21] Since these life expectancy figures use the past age-specific death rates, they therefore are affected by the upbringings of the past.

Moreover, according to Lucas, the relationship between the age-specific death rates and the overall life expectancy used in the HDI (HDI life expectancy, for short) suggests that a population with a higher HDI life expectancy generally has lower death rates throughout the lower age bands.[22] The result is appealing because it suggests that the HDI life expectancy measure 'can be seen as an indicator - - - of the chance that the members of a population have of surviving to what would commonly be called a "reasonable age".'[23] This measure also correlates highly with the quantity and quality of the provision of goods and services such as nutrition, sanitation, health

[21] See, for example, Shryock et al. (ed.) (1976), chapter 15, and Lucas (1985), pp. 15-6. See Appendix 1 for an example of a model to derive life expectancy.
[22] Ibid.
[23] Lucas (1985), p. 75.

care and education.[24] These relationships imply that if the figure is used as a policy objective, it will encourage governments to improve the provision of those goods both in terms of their quantities and qualities.

The brief description of how life expectancy is derived for the HDI shows that 1) it measures the expected duration of life of a population assuming that a set of age-specific death rates remain constant for each cohort, whereas in fact, such rates are known to be constantly changing;[25] 2) it is an expected duration of life of a cohort of infants who are assumed to be born on the same day with the identical capacity to survive (i.e. empirically collected age-specific death rates are assumed to be the only cause for the number of the cohort to decrease as time goes by); and 3) it is derived from a set of age-specific death rates of a population that existed (or ceased to exist) between time t-α and t and therefore is a demography of that particular population.

Bearing this definition of life expectancy in mind, we can state using the capability approach that the indicator measures the capability to achieve a functioning 'surviving for a certain duration as a member of a population that exists between time t-α and t where α is a time interval, assuming that the age-specific death rates of the population remain constant and that the population can be represented by a cohort.'

3-1-6-1-1: Comparison to other alternatives

Before discussing the relative necessities revealed by *life expectancy at birth*, I will first discuss the reasons why *life expectancy at birth* is more appropriate as a

[24] There may be some delays for this relationship to be observed.

[25] See Lucas (1985), p. 74. This is why Lucas claims that 'to some degree it [the life expectancy indicator] is perhaps best seen as a model not of the real world, but of the concept of mortality.' (Lucas 1985, p. 16)

measure of the capability to live a long and healthy life than other possible alternatives.

Before turning to these comparisons, let me consider the possible objection to using *life expectancy at birth* that the assumption that age-specific mortality rates remain constant is an obviously false assumption about a mortality performance of a population under consideration. However, this by itself does not seem a valid objection. Every estimate is calculated by assuming certain patterns or regularities in the population of interests. Assumptions about the stability in the pattern of behaviour or phenomenon, which are not true in reality, may be used when we do not know better ways to make the assumption closer to reality. In this case, we know that age-specific death rates change constantly, but we do not know what the patterns are for the change or whether they are completely random. In such a case, we may reasonably assume that the age-specific death rates based on data from a particular time-period will apply continuously over time even if we know that this assumption is not an accurate description of reality, simply because we are aware of no better alternative. To criticise an approach only because it uses assumptions that are not accurate descriptions of reality is not constructive. Instead, we should examine whether the assumption and the model used for *life expectancy at birth* are good ones for deriving an estimated value of longevity, and whether that particular value is appropriate for the project of measuring human development. To do so, I will review what alternatives are available to estimate longevity, what are the differences between them and *life expectancy at birth*, and what are the reasons to prefer the latter. Specifically, for these purposes I will examine the following three indicators of the average number

of years a person in a given population lives and compare them with *life expectancy at birth* (LE0) used in the HDI.[26]

(1) Life expectancy indicator 1 (LE1): This is a measure based upon the historical average life span.

(2) Life expectancy indicator 2 (LE2): This indicator uses the *crude death rate* based upon the number of deaths in a year per 1,000 of the midyear population.

(3) Estimation of potential lifetime (EPL): This measure uses an average of each person's conditional life span expectation based on various personal characteristics such as age, class, region of residence and race.

The difference between the LE1 and LE0 is that in the former we obtain a real historical average life span, while the latter assumes that a set of age-specific death rates remain constant for a hypothetical group of infants born on the same day and derives a projected average life span. As we see from the way LE0 is calculated as shown in Appendix 1, LE0 is a summary measure of age-specific death rates based upon age intervals, normally between one and ten years. Because of the way in which LE0 is obtained, the measure is responsive to the improvement of mortality of particular age groups. Accordingly, the LE0 is useful in evaluating the success of policies targeting health conditions relevant to a particular age group. For example, improved medical care for new born babies will immediately be reflected in the LE0 through changes in the mortality rates of the age group of new born babies. By contrast, since the LE1 uses historical data on average life span, the LE1 cannot

[26] The following discussions are heavily influenced by Lucas (1985), but unless indicated, I developed the actual arguments.

reflect such a change in the short to middle term. That is, under the LE1 the results of policies to improve health conditions will be reflected in an improvement in life span only after those who benefited from the policy prove to have lived longer than the previous average. If for example, health policy was targeted toward new born babies, the full magnitude of the effect of this policy on the population's life span would be seen only after all of the babies affected complete their lives. Accordingly, it could take 60 years or more to see the full result of the new policy.

There is another significant difference between LE0 and LE1, that is, the extent to which a target level of longevity is implicit in the measurement. As we can see from Appendix 1, LE0 is calculated as a summation of the proportion of those born in the same year who die between each age interval. It may be said to be implicit in this calculation method that the optimum state is where all the cohort born in the same year are alive at an age interval, that is, where the age-specific death rate is zero, at least for age intervals below some age at which basic biological morbidity may affect the death rates. Put another way, since a higher measure under LE0 results as the age specific death rates for the age intervals decline, it is implicit in the use of this measure that the optimum state exists for a particular age interval when the death rate is zero.[27] This could also be seen to imply a normative idea that LE0 measures the degree to which a society's circumstances allow its members to attain what has be described as an essential right – the right to live as long as biologically possible.[28] On the other hand, LE1, which is an estimated longevity based on how many years people in the past have managed to live, does not reflect any target level of longevity.

---

[27] This implicit assumption is that all individuals within an age interval could survive, that is, that the age-specific death rate can be zero. This assumption may, of course, not be met in many cases, for example, because of death from accidents or diseases that cannot be prevented.

[28] See Lucas (1985), pp. 74-5 and Herrera (1976).

As noted, LE2 uses the crude death rate based upon the number of deaths per year per 1,000 of the midyear population. The difference between LE2 and LE0 is that the former is not appropriate for a comparison over populations that have different age structures. This is because a population which has a relatively higher proportion of old people will normally have a higher *crude death rate* compared to another population which has a relatively lower proportion of old people. Evaluations of longevity based on crude death rates could be controversial because a population with a relatively higher proportion of old people, only because of this characteristic, could be ranked lower than another population with a lower proportion of the old. This lower ranking could result even if the former population has uniformly lower age-specific death rates compared to the latter.[29] LE0, which is derived from age-specific death rates, does not suffer from this deficiency, and therefore is more suitable for comparison over populations that are known to have different age structures.

Finally, the *estimation of potential lifetime* (*EPL*) is an indicator suggested by Desai as an alternative to LE0 for measuring the capability to lead a long and healthy life.[30] *EPL* specifically measures the difference between a person's age today and the person's expected length of life given his/her personal characteristics and other social/economic variables such as age, class, region of residence and race. Since life expectancy is conditional on age more than any other single characteristic, *EPL* is defined as the difference between the current age and the conditional life expectancy (in years), which in turn is defined as the reverse of age specific mortality. Mathematically, the *EPL* of the jth person of age i can be described as follows:

$$EPLj = (L^*ij - Lij) \qquad (1)$$

---

[29] See Lucas (1985), p. 16.
[30] Desai (1989).

In this formulation L*ij is the life expectancy of the jth individual of current age i and Lij is the age of the jth individual of current age i. Under the *EPL,* L*i varies depending on the characteristics of the jth individual other than his/her current age. Specifically:

$$L^*ij = L^*i + Y(Zij) \qquad (2)$$

In this formulation L*i is the average life expectancy of persons of age i, Zj is the vector of characteristics of the jth individual, and Y (Zij) is the individual-specific conditional life expectancy. In order to simplify matters, Y (Zij) is assumed to have a zero conditional as well as unconditional mean.

EPL has the advantage of taking into consideration the effects that certain personal characteristics are likely to have on an individual's life-span, rather than focusing only on what is essentially a combined effect of all sorts of causes, i.e., the overall death rates by age categories used in the LE0.

The method *EPL* uses in deriving an individual's capability to lead a long and healthy life is also in line with capability theory. *EPL* shows the expected years of life given the personal characteristics Zj. In other words, *EPL* tells us the average number of years people who live under conditions Zj can be expected to live, consistent with the type of measures relative to individual characters that we require in capability measures.

There are, however, disadvantages in the *EPL* methodology. For example, *EPLs* calculated for entire populations favour countries with growing populations over those with relatively static populations. Let me explain why. Using equations (1) and (2) above, the aggregate level of *EPL* (***EPL***) is calculated as follows:

$$\boldsymbol{EPL} = \Sigma\,\Sigma\,(L^*ij - Lij)$$

$$= P\,(L^* - L)$$

where L is the average age of the population, that is, $\Sigma\Sigma$ Lij = PL, and L* is the overall average life expectancy of the population, i.e., PL* = $\Sigma\Sigma$ L*ij. Alternatively, ***EPL*** can be expressed as a function of population (P), average life expectancy (L*) and the 'age gap,' which is defined as the difference between the overall life expectancy of a nation (L*) and the average age of the population (L) relative to L*. That is,

$$EPL = P \bullet L^* \bullet T$$

Where $T = (L^* - L)/L^*$

Under a given level of age-specific mortality rates in two countries, the average age of a population (L) will decrease in a growing population compared to a population that is static or declining. (That is, as the number of births increases in the country with a growing population, the average age of the population will decline more rapidly than in a country with a lower birth rate.) This makes both L*-L and the 'age gap' (T) higher in a growing population compared to a static or declining one. For this reason the aggregate EPL will favour countries with growing populations compared to those with static populations. The Human Development Report cites this 'natalist' bias as a reason for not opting for the *EPL* in the HDI used for international comparisons.[31]

In sum, an analysis of the use of the LE0 – that is, the life expectancy at birth – compared to the three alternatives confirms that LE0 has the following benefits:

1) it is useful for policy evaluations in the short to medium term;

[31] See UNDP (1993), p. 105.

2) it provides a measure against an implicit standard where the optimum state exists where individuals live as long as biologically possible, arguably an essential right;

3) it enables comparisons over populations with different age-structures;

4) it emphasises improvement in the population's chance of surviving to a reasonable age; and

5) it does not have a 'natalist' bias by favouring countries with a growing population.

Point 3) definitely fits into the purpose of the HDI because different countries have very different age-structures. Regarding point 1), since human development was developed as an alternative policy objective to the GDP growth, it is important for governments to see the effect of their policies to improve general health conditions not too far in the future. The second and forth points are each consistent with the UN's goals of providing basic human rights to all. Finally, the fifth point is consistent with the Report's reluctance to make any (implicit) value judgement between a growth in population and an improvement in average life span.

The above is not to say, however, that the LE0 is superior to all the alternative measures in all respects. In particular, as noted above, from the point of view of incorporating the essence of capability measures (i., e., its ability to take into account necessities of individuals relative to their personal features, time and space), *EPL* is more appropriate to measure the capability to lead a long and healthy life than LE0. The principal difficulty in the *EPL* is its 'natalist' bias, as discussed above. Accordingly, it would be appropriate for the makers of the HDI to consider whether this bias overweighs the benefits of the *EPL* in term of its consistency with the capability measures.

3-1-6-1-2: Identification of relative necessities

In order for a capability measurement to identify relative necessities, as discussed earlier in the chapter, we need a common (or absolute) criterion relative to which we reveal necessities of different individuals or groups of individuals. Such a criterion implicit in *the life expectancy at birth* is to live as long as (biologically) possible, because the age-specific death rates, from which we derive the life expectancy of a population, regard the state where no one dies between any age-interval (zero death rate) as the optimal state. We can say that an explicit criterion set by the HDI is a life expectancy of 85 years, because it is the maximum in the scale that measures the capability to live a long and healthy life. Therefore, relative necessities revealed by the life expectancy in the HDI are necessities of a population in general to be able to live until 85 given its characteristics.

There are a variety of reasons behind a population's not being able to achieve the life expectancy of 85 years, such as inadequate provision of nutrition, bad hygienic conditions, war, a high proportion of suicide, and other diseases or accidents. Similarly, the necessities for the population to achieve the target level also vary: provision of nutritious food, a clean water supply, a stable political environment, adequate mental-health care, solutions to private problems, effective medical treatment etc. The capability measured by life expectancy does not discriminate between these different reasons for achieving a particular level of life expectancy. This is one of the problems discussed above that is inherent in the measurement of capability, that is, that the measurement does not itself dictate the extent to which we identify relative necessities. That is, it does not discriminate between various sorts of reasons behind a population's poor performance on the longevity measure during a

certain time and at a certain place, and therefore does not provide guidance on what needs should be met in order to see improvements.

What about necessities relative to features specific to each individual? Could such necessities be identified by the capability measurement using the life expectancy figure used in the HDI?

Improvement in health conditions and conditions relevant for survival at an individual level contribute to an improvement in life expectancy in the following way: by improving the probability of surviving between now and α years later (the time interval in the data on age-specific death rates used to derive life expectancy figures) and thus improving the performance of the relevant age-specific death rate. Improvement of the death rate for any age interval should contribute to longer life expectancies. Therefore, there is a very large number of ways in which individual improvements in the probability to survive contribute to an improvement in the overall life expectancy of the population. This implies that in theory, at the individual level, there are no fixed criteria relative to which the necessities of each individual are identified in terms of the overall target, such as a life expectancy of 85 years.

In practice, however, there may be such criteria, not at an individual level but at the level of a group of individuals who share common characteristics that are recognised as resulting in poor performances in life expectancy. As I have mentioned earlier, in fact, there exists evidence that a population with a higher life expectancy generally has lower death rates throughout the lower age-bands, and also that there is a strong correlation between life expectancy and the provision of goods such as healthy food, sanitation, health care and education. Because of such evidence, low overall life expectancy could be seen as an indication that the mortality rates in the

younger age-groups need to improve and that improvements need to be made in providing proper nutrition, sanitation, health conditions and education.

3-1-6-2: Adult literacy rate and the school enrolment ratio

The *adult literacy rate*, according to the definition provided by the Human Development Report, is the percentage of people aged 15 and above who have the ability, with understanding, both to read and to write a short, simple statement on their everyday life.[32] This test for literacy, if constructed properly,[33] therefore is a measurement for 'the capability to read and write with understanding.' This test for literacy applied to an individual measures whether a person can read and write with this understanding. This capability measure, however, does not measure whether the individuals are in fact reading and writing. That is, the person who is literate could choose not to use his ability to read and write or may not have the resources to read or write. In this case, the literacy test measures the maximum proportion of the population that is in fact reading and writing at a functional level.

A population's literacy rates, analogous to the life expectancy figure, can be seen as a summary of a population's overall achieved level of functional literacy, since this summary assigns equal weight to each individual in the population. If we see the population's literacy performance as reflecting all the possible individual states with respect to reading and writing, a literacy rate can be seen as a measure of a

---

[32] See UNDP (1998), p. 218.

[33] As the 1990 Human Development Report notes, sometimes there are difficulties in controlling the quality of the literacy measurements in practice. (UNDP 1990, p.112) Therefore, the measurements and comparisons of literacy across societies are reliable only if the quality of the measurements is internationally well monitored. The 1998 Human Development Report also reports evidence suggesting that literacy is not necessarily a good measure of being able to read and write with understanding in the modern era. The Report says that 'many people – 18% of adults on average in 12 European and North American countries – though "literate", have such low levels of skills that they cannot meet even the basic reading requirements of a modern society.' (UNDP 1998, p. 23)

set of functionings in the population and therefore, a measure of a capability set in a formal sense of the term.

But the result from the literacy test is not the same as 'the capability of acquiring knowledge,' which the Report wants the HDI's educational index to measure. The capability to read and write with understanding is an important condition for the capability to acquire knowledge in civilised societies, but being able to read and write with understanding is not sufficient for a person to be able actually to acquire knowledge. I propose that in order for the literacy rates to measure a person's ability actually to acquire knowledge, the person needs (1) materials (books and other publications), (2) informative interactions with others, and (3) some guidance on how to accumulate and organise knowledge, *in addition to* the ability to read and write. Therefore, under my proposal, in order for literacy tests to measure 'the capability of acquiring knowledge,' we need to determine whether these three conditions also exist.[34]

3-1-6-2-1: Comparison to other alternatives

As explained above, the HDI has used different means of measuring 'the capability of acquiring (or being able to acquire) knowledge.' Specifically, since the 1996 Report the HDI has used the *combined gross school enrolment ratio* and the literacy rates, and in the earlier Reports, the educational dimension was measured firstly by the adult literacy rate only and then by the combination of the literacy rate and the mean years of schooling. Let me now consider whether this most recent

[34] Here I am talking only about knowledge that requires reading and writing for its acquisition; therefore, other types of knowledge, such as how to cycle or use a pair of scissors, are not included.

measure – the combined gross school enrolment ratio and the literacy rate – is superior to the other two alternatives used earlier in terms of capability measures.

The combined gross school enrolment ratio as detailed by the Human Development Report is calculated as the number of students enrolled in a level of education – whether or not they belong in the relevant age group for that level – as a percentage of the population in the relevant age group for that level. Note that the numerator of this ratio is not limited to the members of the particular age group. One alternative to this, which is called the net school enrolment ratio, would be only to include students who are enrolled at school in the relevant age groups. However, I believe that the ratio used in the current HDI – the gross figure – is better in terms of measuring a dimension of capability to reflect relative aspects of well-being. Specifically, in counting those who are enrolled at school later or earlier than the standard age, the gross figure takes into account individual differences regarding their learning speed or the right timing for school education. That is, given that different individuals in a society may be able to learn more effectively at an earlier or later age than the norm, the gross figure is able to measure differences based upon the educational experience of these individuals.

In any event, the first comparative issue I consider is whether the combination of the literacy rate and *the combined gross school enrolment ratio* is a better measure for the capability of acquiring knowledge compared to the literacy rate only. I believe that it is better in light of the above discussion on the three conditions that must exist for literacy effectively to measure the capability of acquiring knowledge. That is, a calculation of the proportion of people from the relevant age group who are provided with school education could be a good indicator for the degree to which society meets these three conditions. Specifically school education provides a measure of whether

the individuals have been provided with (1) materials useful for building knowledge, (2) guidance about the method of learning things, and (3) opportunities to interact with others who have similar interests and who are at the similar levels of education. Therefore, *the combined gross school enrolment ratio* should complement the literacy rate in measuring the capability to acquire knowledge.

The next issue is whether the combined gross school enrolment ratio is more suitable for the purpose of the HDI than *the mean years of schooling*, as previously used in the HDI. I believe that the combined gross school enrolment ratio is preferable, for reasons that are similar to the preference for the use of the life expectancy of birth compared to the average life span as discussed above. Specifically, the combined gross school enrolment ratio is expressed in terms of a ratio that implicitly assumes that a 100% combined school enrolment ratio among the school age population is the optimum target level of educational attainment. Accordingly, the closer the ratio for this proportion is to 100%, the nearer we can say that the proportion is to this optimum level. However, since *the mean years of schooling* is expressed in terms of number of years, this measure does not have any such implicit target. As discussed earlier in the comparison of different measures of life expectancy, it is advantageous for the HDI to have such a target since this target encourages a society to educate all the individuals in the school age. This encouragement occurs since the target level of 100% is reached only when everyone in the school age population is enrolled in school.[35]

---

[35] Note that since the combined gross school enrolment ratio includes in the numerator individuals enrolled in school who are not in the particular age group, it is possible to have a ratio greater than 100%, and to this extent the measure is not as effective in providing an implicit target. Use of the net school enrolment ratio, discussed above, would be more effective for this purpose. However, as discussed above, there are the advantages to using the gross ratio, and these advantages appear to me to support use of the gross ratio than the net ratio.

Still, we need to keep in mind that both of these measures are only indicators for the three conditions noted above, not the direct measurement for them. That is, the fact that students are registered in schools does not necessarily mean that they have been provided with books to read, people to interact with, or appropriate guidance for building knowledge. There is, in fact, evidence that in some societies schools have very poor facilities and students do not get proper guidance from teachers.[36]

3-1-6-2-2: Identification of relative necessities

Let me now consider how the use of the combined gross school enrolment ratio relates to the concept of *relative necessities* as expressed above. As noted, relative necessities refer to the fact that an individual's ability to utilise particular resources – and by extension a society's ability – depends on the personal characteristics of the individuals, and such abilities can be measured in terms of individuals' functionings. In this case, the relative necessities are measured in reference to whether the individual is 1) literate; and 2) has enrolled at school whether or not the person belongs or belonged to the relevant age group (assuming the person is 15 or over), respectively. By the current HDI measure, only those who have achieved both functionings are regarded as 'capable of acquiring knowledge.'

Let us first consider what this circumstance means at an individual level. In this case it is apparent that an individual may lack either or both of these functionings of being able to read and write and of being or having been enrolled at school for a variety of reasons. For example, the former – not being able to read and write – may result because 1) the individual is not provided with appropriate school education; 2) the individual is disabled and does not have the physical ability to read and write;

[36] See, for example, Bruce Fuller (1986).

or 3) the individual is not interested in learning. Some major reasons for the latter – not being or having been enrolled at school – may be 1) the individual lives in a rural area with no means to commute to school; 2) the individual's parents will not allow him to go to school, but instead require him to do something else (usually to work); or 3) a war is going on and the civil services such as schools are paralysed. Because the literacy test and the combined school enrolment ratio do not discriminate between these various reasons, these measures do not provide complete measures of the relative necessities in terms of different individuals' functionings.

Further, the use of the gross school enrolment ratio also implicitly assumes that individuals who are registered at school have functionings that allow them to take advantage of school enrolment (as a proxy for the provision of educational opportunities and materials). However, this does not reflect the needs of those who are registered and yet are not actually provided with education at school, for example those who are registered and yet do not go to school because they are bullied or they are forced to work.

So far, I have discussed the implications for the relative necessities if the literacy tests and school enrolment ratio are applied at the individual level. Of course, the actual figures for the dimension of education in the HDI are summaries of these measures over a population. That is, the adult literacy rate measures the proportion of people who are literate in the population, and the combined school enrolment ratio measures the proportion of the people in the population of relevant age who are enrolled at schools. As described above, educational attainment in the HDI is an aggregate of the literacy rate and the combined school enrolment ratio, with two-third and one-third weights, respectively. The maximum level for this educational dimension in the HDI is 100%, which consists of a 100% literacy rate and a 100%

combined school enrolment ratio. Unless everybody in the population is both literate and enrolled at school (if they belong to the relevant age group), educational attainment will not reach its maximum level of 100%. Accordingly, although both literacy rates and combined school enrolment ratio are summary measures of a population, they do reflect the necessities of different individuals to be literate and enrolled at school. That is, the HDI measurement of the 'capability to acquire knowledge' does adjust for those in the population under consideration who have not yet achieved either or both of these functionings by causing a lower than target level of educational attainment.

A further advantage the literacy rate and school enrolment ratio have over other summary measures such as average income or average life span is that the literacy rate and school enrolment ratio are expressed in terms of proportions, and as a result they are able to indicate distributional aspects of the educational attainment. Accordingly, unlike these other summary measures, the number of people who are not literate or who are not enrolled at school cannot be averaged out by the number of people who are literate and who are enrolled at school. (In this sense the educational attainment measure is similar to the measure of the life expectancy at birth, which also has provides some distributional aspect of well-being. This is because the life expectancy at birth is aggregated in such a way that a population with a higher life expectancy at birth generally has lower death rates throughout the lower age bands. As discussed below, as an aggregate measure, the real GDP is much more problematic because of its inability to take into account distributional aspects.)

As with the analysis as the individual level, a problem with this capability-measurement of educational attainment using literacy rates and the school enrolment ratio is that the measurement does not discriminate between different reasons why the

population has not achieved the target of 100% educational attainment. Any reason – from the government not providing the population with appropriate primary teaching to the populations' not being willing to learn how to read and write – is a reason for some proportion of the populations' being incapable of acquiring knowledge. If the measurement is to be used to allocate resources, it would, however, be very helpful to know why the population has a low level of educational attainment.

3-1-6-3: The real GDP per capita in purchasing power parity dollars

The per capita gross domestic product (GDP) is normally used to estimate the average income of the people in a country. Even this correspondence is not without problems,[37] but in this section, I will examine the GDP figure as a measurement of a dimension of capability, assuming that the GDP per capita is an appropriate estimation of the average income of the population in a country. (In this section I will sometimes refer to the HDI's 'income figure.' This term refers to the per capita real GDP figure adjusted to account for inflation and other factors noted in this section.)

Under this assumption, the GDP per capita measures the degree to which the people in the country are capable of purchasing (and consuming) goods and services in the particular country. More precisely, the GDP per capita data reveal the maximum functionings (among a set of them) an average individual in the country could achieve in terms of purchasing and consuming goods and services. I say that this data reveal the maximum such functionings because the average GDP figure assumes that the average individual has the capability to use all his income for purchasing goods and services for himself. There may, however, be reasons why this average individual may not have this capability, for example because of existing debts

[37] See, for example, Morgenstern (1963), and Dornbusch, Fischer and Startz (1998).

that must be repaid or because the distribution system in the country does not provide access to many goods and services.

There are, however, more basic problems with the use of the GDP per capita in the HDI. These problems concern the fact that the HDI income figure is supposed to measure *the degree to which people in the country are capable of having access to resources needed for a decent standard of living*, [38] not the capability of the average individual to purchase and to consume goods and services. The use of the GDP per capita figures is much more problematic in the case of the former goal compared to the latter, for a number of reasons.

First, as other commentators have noted, not all the needs to maintain a decent standard of living can be purchased. Some such needs, for example, clean water and clean air or public playgrounds for children, cannot be purchased by an individual consumer, but have to be supplied publicly.[39] There are other needs that may be regarded as necessary for a decent standard of living that simply cannot be satisfied materially; participation and communication in the local community may be examples of such needs. These problems may not be crucial if we can reasonably assume that countries will provide public goods necessary for a decent standard of living, that people do not pay disproportionally high taxes for these goods, and that people's non-material needs are satisfied, so that the GDP can measure only the rest of the needs for a decent standard of living. However, these assumptions are not likely to be

---

[38] The 1990 Human Development Report uses 'having access to resources - - - ' and 'having command over resources- - -' interchangeably. I will use the former expression in this paper.

[39] This does not mean that the citizens of the country do not bear the expense in a collective sense. However, the expense they bear is normally much less than the value they enjoy from using these public goods. More importantly, a country's income level does not correspond systematically with the country's provision of public goods. A country's provision of such public goods depends on government tax scheme, the size of the population, the public's preference for public goods and the availability of mechanisms through which such preferences can be reflected.

reasonable for many countries. Further, there is likely to be a bias in the countries for which the assumptions are or are not reasonable. Specifically, it is likely that these assumptions are reasonable for developed countries, which have effective political systems to encourage the provision of public goods and non-monetary aspects of a decent living standard, but will not be reasonable for under developed countries with very limited political rights. To the extent that this bias exits, the use of the GDP per capita is inconsistent with the universality goal of the HDI, that is, to provide a measure that is universally applicable.

A second problem in using the GDP per capita for the measurement of command over resources for a decent standard of living relates to the concern that the measure does not take into account relative necessities. To some extent the HDI has attempted to adjust for such necessities. Thus, the HDI uses the real GDP figure instead of the nominal one, and thus takes into account the fluctuations in purchasing power caused by inflation (the change in prices over time). The index also makes adjustments (in dollars) to account for the degree of openness of the economy, the relative size of tradable and non-tradable sectors, and the possible overvaluation of exchange rates.[40] These adjustments aim at using the income figure for international as well as inter-temporal comparison of the degree to which people in different countries have access to resources needed for a decent standard of living. However, these adjustments do not take into account differences in personal features or differences in climates and cultures that may give rise to different levels of income needed to achieve a similar standard of living for different individuals or group of homogeneous individuals. Even the adjusted real GDP per capita implicitly assumes

[40] See UNDP (1993), p. 106.

that individuals in countries with the same amount of income will be able to satisfy their needs to the same extent. For example, this measure assumes that *A* who is young, fit and lives in a temperate climate and *B* who is ill, needs expensive medical treatment and lives in an extremely cold climate can achieve a decent standard of living to the same degree with the same amount of income. This implicit assumption is a problem from the point of view of the capability theory, because the income figure ignores necessities relative to the individual's personal features and surrounding environment.

One way of making the GDP indicator more sensitive to the necessities relative to different countries as well as to the actual capability sets people enjoy in a particular society is to adjust the GDP for different personal features and environmental factors such as the following: 1) climate conditions; 2) the proportion of the population that is disabled; and 3) the degree to which the society has infrastructures permitting the free movement of goods and people. The last type of adjustment is important in the measurement of capabilities using indicators that evaluate individual access to resources. That is, for example, in order for an individual to be actually capable of 'moving around,' the fact that a disabled person has a wheelchair is not sufficient. We also need to know whether there are wheelchair ramps and the accommodations that actually enable the person to move around. Because indicators about individuals (or an average individual) such as the GDP per capita do not tell us how advanced a society is in terms of the infrastructures which facilitate the individuals' capabilities, we need to compliment such individual indicators with social indicators.[41]

---

[41] I would like to thank Patrick Feng (a visiting scholar at the Tinbergen Institute, Amsterdam, 1999) for pointing this need out.

A third problem with the adjusted real GDP figure is that it is an average one, and therefore it cannot really measure the proportion of individuals in the country who have achieved the income level necessary to enjoy a decent standard of living. One possible way to provide such a measure is to disaggregate the GDP data according to regions, genders, or ethnic groups, assuming that there exist some regularities between income levels and these classifications. Results obtained from regional disaggregation and gender disaggregation are in fact presented in the Human Development Report.

An alternative way of solving at least partly the problem of using an average income figure for a measurement of capability to reveal the relative necessities of individuals is to use distribution-adjusted income. In fact the 1994 Human Development Report calculated income-distribution-adjusted HDI for each country by multiplying income (adjusted for by the diminishing marginal utility assumption) by 1 minus the Gini coefficient. An example using this data for Sri Lanka and Botswana can serve to explain how this adjusted figure relates to the capability issue. Sri Lanka and Botswana have similar HDI rankings and are both categorised as 'medium human development' countries in the 1992 Human Development Report. But Sri Lanka ranks much higher than Botswana if income is adjusted for distributional inequality.[42] These rankings may imply, together with other information, that in Sri Lanka, a higher proportion of people have access to resources that allow them to enjoy a decent standard of living than in Botswana. This inference is likely to be true especially because both countries' average income levels are not very high; (If the average income of a country is very high, even people at the bottom of the distribution may be

[42] See UNDP (1994), p. 107. Sri Lanka's HDI (1992) and income-distribution-adjusted HDI (1992) are 0.665 and 0.510 respectively, and those of Botswana's are 0.670 and 0.374 respectively.

said to have achieved a decent standard of living.) Accordingly, a capability measure that gives Sri Lanka a higher ranking compared to Botswana is superior to the one that ranks them equally.

There is another reason why a distribution-sensitive income figure may be superior to the income figure currently used in the HDI in terms of capability theory. Specifically, the use of an average income (distribution-insensitive income) measure such as used in the current HDI does not take into account necessities that are relative to what others in the same community have. Recall, for example, Adam Smith's well known concern with the well-being of individuals in terms of their access to resources enabling them to appear in public without being ashamed. The quotation from Townsend at the beginning of this chapter also provides a similar message: what are considered necessities differ according to what others in the same society have, demand and produce.[43] If we take the view that what constitutes a decent standard of living is relative to time and place, as Townsend does, then a capability measure that aims at revealing relative necessities of individuals should take into account the distribution of income. That is, where the income distribution is relatively flat, individuals in the society will have a greater relative sense of well-being than in a society of comparable average GDP per capita where the income distribution is more skewed.

So far I have ignored the diminishing marginal utility assumption applied to the income figure used in the HDI. The actual income figure in the HDI has been scaled so that the income level beyond the world average is discounted progressively

[43] Townsend (1979).

as the level gets higher.[44] Here, I would like to examine whether this adjustment in the income figure solves any of the problems discussed so far about the use of the per capita GDP to measure the capability to have access to resources.

In the 1998 HDI, the real per capita GDP in purchasing power parity (PPP) dollars was adjusted by applying progressively lower weights for amounts beyond a poverty line set as the world average income. Diminishing value beyond this poverty level income is expressed by applying a modified version of the Atkinson formula. Specifically, for any per capita income that lies between the poverty level and twice the poverty level, the Atkinson parameter (the elasticity of the marginal utility of income with respect to income) is taken to be one-half; for any income between two and three times the poverty level, the parameter is taken to be two-thirds, and so on. This approach can be expressed as follows:

$$W(y) = y \text{ for } 0 < y < y^*$$

$$= y^* + 2(y - y^*)^{1/2} \text{ for } y^* \leq y \leq 2y^*$$

$$= y^* + 2(y^*)^{1/2} + 3(y-2y^*)^{1/3} \text{ for } 2y^* \leq y \leq 3y^*$$

Now recall two different problems addressed above concerning the use of the average per capita GDP to measure a dimension of capability sets. One problem is that in using an income figure for the capability measure, we assume that the same amount of income will satisfy needs of different individuals to the same degree. The second problem is that the average data do not reveal the proportion of individuals in

[44] The 1999 Human Development Report uses a new method of representing the diminishing marginal utility assumption. The new method does not have a threshold income below which there is no discount and beyond which there is a progressively larger discount, as was the case previously. Instead, the 1999 HDI income index applies the diminishing marginal utility assumption throughout the income scale by taking the logarithm for every income level. But the points I make in the following discussion still apply for this new method adjustment in income data.

the country who have not achieved the level of income necessary to enjoy a decent standard of living.

The diminishing marginal utility adjustment in the income figure used for the HDI implies that countries with higher average incomes obtain marginally less of an increase in their standard of living than those with lower incomes when they both receive the same amount of additional income. This characteristic of the adjustment implies as a policy matter that if resources are allocated among countries in order to make the most efficient use of resources, the resources will go to the country with the lower income level. Accordingly, I conclude that the adjusted income data is superior to the non-adjusted data because the former can capture at least one aspect of relative well-being, that is, the fact that those who earn less income are likely to benefit more from a given increase in income in terms of their ability to achieve a decent standard of living.

This aggregated approach to the adjustment for the diminishing marginal utility is not, however, the optimum way to make such an adjustment in order to reflect relative necessities. For this purpose it would be better to apply the diminishing marginal utility adjustment to the individual data from which we obtain the average country figures, not to the averaged country figures. This intra-country adjustment would essentially allow the income adjustment used in the HDI to reflect the relative necessities *within* a country as well as between countries. Thus, this intra-country adjustment would allow the index to distinguish between two countries with the same average income level but very different income distributions. Assuming that the diminishing marginal utility applies within the country, then the most efficient allocation of resources would likely be to provide the resources to the country with the higher proportion of poor people compared to the country with the more equal

income distribution, at least if one assumes that the resources would actually be provided to the poor people in this country. Although this result could have the perverse effect of encouraging or rewarding countries with an unequal income distribution, the result would be analytically more consistent with a focus on the relative necessities.

The above analysis on the benefits of the diminishing marginal utility adjustment does not, however, apply to necessities relative to personal features and surrounding environments. The adjustment, for example, is not sensitive to the fact that those who are physically disabled need more resources for a given level of well-being than others who are not disabled, nor is the adjusted income sensitive to the fact that those who live in a cold climate need more resources than those who live in milder climate. Therefore, although the adjustment captures relative needs of individuals measured by income, such an adjustment does not capture relative needs that are not reflected in the income figure but are also relevant to a measure of the command over resources to maintain a decent standard of living.

3-1-6-4: Summary of the subsection

The discussions in the subsections of 3-1-6 have shown a number of important points relative to the HDI's consistency with capabilities and relative necessities. First, the discussion showed that the indicators and proxies chosen for each of the three dimensions of the HDI are not necessarily the best alternatives with respect to the HDI as a measure of essential capabilities. *Life expectancy at birth* and the aggregation of *the adult literacy rate* and *combined gross school enrolment ratio* do have advantages over other measures of health and education as measures of capability, but *the estimation of potential lifetime* as a measure of health is better than

life expectancy at birth strictly in terms of a capability measure. The capability dimension for a decent standard of living is clearly better measured by a distribution adjusted GDP compared to the HDI's current income figure – the adjusted real GDP per capita ($PPP).

Second, the selected data – life expectancy at birth, literacy rates and combined school enrolment ratio, and adjusted real GDP per capita in purchasing power dollars – do not measure essential capabilities (sets of doings and beings that are the components of well-being) unconditionally. In order for them to measure such capabilities of interests, various implicit assumptions have to be made. In the case of life expectancy at birth, it is assumed that a particular set of age-specific mortality rates applies throughout a cohort's life. In the case of literacy rates, assumptions are made on access to books, guidance and interaction with the world in general. In the case of per capita GDP, it is implicitly assumed that everyone is able to achieve the same level of capability with an equal amount of income. The capability measure is successful if the conditions described by those assumptions are met in a society where the measurement is used. The awareness of such conditions is important in my account of successful indexes where an instrument for measurement is supposed to have a 'capacity' to measure a concept under investigation. As I explained in chapter one, Cartwright's notion of 'capacity' indicates that a model or a measurement instrument reveals its ability to inform us about facts about the world only when a set of conditions are met.

As discussed above, however, it is very difficult for these assumptions to be true, especially for the whole population under investigation. The questions, therefore becomes whether the assumptions can be seen as good abstractions of the reality (which is the case in the life expectancy at birth) or what proportion of the population

meets the condition described in the assumption (which is the question relevant for adult literacy rates). I have shown that the implicit assumption behind the income figure is not only a bad abstraction of reality, but is also inconsistent with the main goal of the capability approach: that is, to apply a measurement that can take into account relative aspects of well-being. For the HDI to be consistent with theory, thus, what we expect the income figure to measure has to be modified.

Third, the three indicators all potentially suffer from the fact that they are summary measures of a population, that is, they are incapable of revealing relative necessities at individual levels. But the three indicators have different ways of solving at least part of this problem of summation. The measurement of educational attainments, because it sets 100% as the target, solves the problem thus: unless everyone in the population is capable of acquiring knowledge, the 100% target will not be achieved. The income figure, because of the adjustment for diminishing marginal utility, is capable of encouraging international organisations to identify and improve the states of countries that are less well-off. In order for this result to be achieved for differences *within* a country, the diminishing marginal utility adjustment has to be applied to disaggregated income data for different groups of individuals within the country. Regarding the life expectancy figure, because of a correlation between this figure and other natural and social phenomena, life expectancy is capable of identifying those who possess or suffer from phenomena such as poor nutrition, sanitation, health conditions and education.

Finally, a common problem with the three indicators is that they do not discriminate among the reasons for achieving particular levels of the capability to live a long and healthy life, the capability to acquire knowledge, and the capability to maintain a decent standard of living. This is one of the problems inherent in the

capability measurement, as I have discussed in the previous section of the chapter. If the index were to be used for resource allocation, independent criteria for discriminating among different reasons for the shortfalls would be necessary.

3-2: The LSI and theory

The principal purpose of this chapter is to discuss the HDI in terms of the theory-dependent criteria for a successful index. To provide some useful comparison to this analysis, I will briefly examine the relationship of the LSI to the theories developed by Jeremy Bentham for measuring pleasures and pains in his book 'An Introduction to the Principles of Morals and Legislation' first published in 1789.[45] This analysis provides a contrast to the purpose-dependent analysis for the LSI since the LSI index-maker was less interested in implementing a particular theory into the measurement than were the index-makers of the HDI. As noted above, the HDI's index-makers were rather directly and expressly influenced by Sen's capability theory in developing the HDI. By contrast, the index-maker for the LSI did not have such express theoretical purposes. Instead, the LSI index-maker was influenced by a number of general theoretical sources, including the 'utility' theory developed by Bentham.[46] Accordingly, the analysis of the LSI in terms of theory-dependent criteria is less significant for the LSI.

[45] The edition I used is Jeremy Bentham An Introduction to the Principles of Morals and Legislation, edited by J. H. Burns and H. L. A. Hart (1996), pp. 11-2.

[46] Other general theoretical sources of the LSI include B.W. Heady's studies on well-being based on an integration of theories from psychology and economics and R. Veenhoven's evaluation of happiness based on psychological and sociological understandings of the concept. See, for example, Heady (1993) and Veenhoven (1993).

3-2-1: The general relationship of Bentham's utility theory to the LSI

Bentham's utility ideas are of course well known. According to Bentham, the principle of utility 'approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interests is in question.'[47] By 'utility' he means the 'property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happiness of mischief, pain, evil, or unhappiness to the party whose interest is considered.'[48] An object is said to promote the interest of an individual member of a community 'when it tends to add to the sum total of his pleasures; or, what comes to the same thing, to diminish the sum total of his pains.'[49] Further, for Bentham the interest of the community is 'the sum of the interests of the several members who compose it.' An action is said to conform to the principle of utility for the community 'when the tendency it has to augment the happiness of the community is greater than any it has to diminish it.'[50] Finally, regarding the government's role, Bentham claims that 'a measure of government - - - may be said to be conformable to or dictated by the principle of utility, when in like manner the tendency which it has to augment the happiness of the community is greater than any which it has to diminish it.'[51]

One can see from this brief summary that Bentham's utility theory bears a relationship to an underlying premise of the LSI. Specifically, the LSI is premised on the view that a good government policy is the one that improves the general level of

---

[47] An Introduction to the Principles of Morals and Legislation (1996), pp. 11-2.
[48] Ibid. p. 12.
[49] Ibid. p. 12.
[50] Ibid. pp. 12-3.
[51] Ibid. p. 13.

satisfaction of a society. As summarised above, the LSI uses a public opinion poll to determine the 'degree of satisfaction level' (DSL), and the LSI is calculated so as to relate various measures to the results of this poll. This view is in harmony with Bentham's idea that the government's role is to act so that the sum total of happiness increases relative to the sum total of pain for all the individuals in a community.

3-2-2: The LSI and Bentham's method of measuring pleasures and pains

A deeper level of comparison between Bentham's approach and the LSI can be developed by examining the methods of measuring pleasures and pains that Bentham provides in the later chapter of his book.[52] According to Bentham, one should evaluate the government's act in terms of its general tendency to promote or to diminish the interests of a community by first looking at the degree of satisfaction of individuals. According to Bentham, a person's degree of pleasure or pain is determined by the following four factors:

1. The intensity of the pleasure or pain;
2. The duration of the pleasure or pain;
3. The certainty or uncertainty of the pleasure or pain; and
4. The propinquity or remoteness of the pleasure or pain.

Further, according to Bentham, in order to evaluate the effect of a government action in terms of pleasures and pains, we need to take into account the following three additional factors:

1. The fecundity of the pleasure or pain, or the chance that a pleasure or a pain will be followed by sensations of the same kind;

---

[52] Ibid. pp. 38-41. Fukuda (1995) also introduces the method.

2. The purity of the pleasure or pain, or the chance of a pleasure or a pain not being followed by sensations of the same kind; and

3. The extent of the impact of the pleasure or pain, or the number of persons to whom a pleasure or a pain extends.

According to Bentham, in order to evaluate a proposed government action with respect to its effect on a community as a whole, we begin with any one person whose interests seem most immediately to be affected by an act, and evaluate the following for this person:

1. Each distinguishable *pleasure* that appears to be produced by the act in the *first* instance,

2. Each *pain* that appears to be produced by it in the *first* instance,

3. Each *pleasure* that appears to be produced by it after the first, and

4. Each *pain* that appears to be produced by it after the first.

Bentham would have us sum up the values of all the pleasures on the one side, and those of the pains on the other. If the sum balances on the side of pleasure, the result indicates the 'good' tendency of the act with respect to the interests of the particular individual; if the sum balances on the side of pain, the results indicates the 'bad' tendency.

Bentham would then have us repeat the same process for all the individuals in the community whose interests are affected by the government's action. He would then count the number of individuals to whom the 'good' tendency prevails and the number of individuals to whom the 'bad' tendency prevails as a result of the action, and if the former is larger than the latter, the result indicates the general 'good' tendency of the act. If, on the other hand, the latter is larger than the former, the result

indicates the general 'evil' tendency of the act with respect to the community in question.

There is some degree of similarity between Bentham's approach and that of the LSI. Thus, as with Bentham's approach, the LSI attempts to determine the overall level of satisfaction by cumulating the satisfaction levels of individuals. The LSI uses the opinion poll result of the 'degree of satisfaction level' (DSL) to measure improvement in the LSI. The DSL is the proportion of people within the whole population who state that they are satisfied (in general with their lives), and it uses the method of counting the number of persons who are satisfied in general and compares that with the number of persons who are not satisfied in general.

There are, however, more fundamental differences or questions between Bentham's approach to measuring pleasure and pain and that employed by the LSI. An initial difference or question requires some clarification in the precise significance of the four factors that Bentham identifies as determining an individual's degree of pleasure or pain – that is, intensity, duration, certainty or uncertainty, and propinquity or remoteness.[53] Specifically, it is not clear whether Bentham means that as a normative matter pleasures and pains *should be* evaluated according to those four criteria, or whether he means that as a descriptive matter based upon scientific conclusion, sensations of pleasures and pains are determined by those four factors. If Bentham's approach is a normative one, then the people asked the question 'How satisfied you are in your life in general?' should be informed in advance that they should take Bentham's four criteria into account if the results are to be consistent with Bentham's approach. Since they are not so informed, if this normative approach is

[53] Criteria relevant for such evaluations are six if the purpose is to assess government actions. However, as I point out shortly, DSL is not designed for the purpose.

assumed then there would be an inconsistency between the DSL results and Bentham's approach. If, on the other hand, Bentham's approach is a descriptive one based on his understanding of how human sensations of pleasure and pain are determined, then it is possible that the DSL's reliance of subjective judgement of individuals will approximate the results of an analysis based on Bentham's approach. That is, if all individuals asked to state whether they are generally satisfied with their life do this by implicitly calculating the balance of their pleasures and pains according to Bentham's four criteria, then the results could be consistent with the results of his analysis. However, a determination of whether in fact Bentham's four criteria are descriptive as a matter of fact is beyond the scope of this thesis.

A second difference or question relates to the temporal aspects of Bentham's approach. As noted above, each person's response to a particular act is to be evaluated in terms of the pleasures and pains produced in the first instance and also after this first instance. However, we also do not know whether the people who answer the questionnaire actually calculate their general levels of satisfaction (dissatisfaction) in this inter-temporal manner suggested by Bentham, that is, to weigh up pleasures and pains arising from the current state as well as its effect on the future pleasures and pains.

A third area of difference concerns the extent to which the index focuses on specific rather than general actions. As noted above, Bentham's approach focuses on evaluating whether a particular governmental action produces a positive result in terms of the balance of individual results. However, the DSL is designed to measure a community's state of satisfaction in general at a particular period in time, not to evaluate the impact a certain government action has on the community in terms of people's satisfaction.

A fourth difference involves the quantitative measurement used in the DSL. That is, the DSL provides information not only on whether the general level of satisfaction is positive or negative, but also the degree to which it is so. And the numerical representations of DSL are necessary to find (using econometric methods) important policy variables that improve the satisfaction level in general of a society, which in turn constitute the LSI. By contrast, Bentham's analysis is based upon whether on balance the number of individuals whose pleasure is improved by a particular action is greater or less than the number whose pleasure is reduced, without making rankings or numerical comparisons.

Chapter four: Condition-dependent criteria for the HDI and the LSI

This chapter examines the HDI and the LSI with respect to their relation to concerns involving the real world. The examination derives from my view that a successful measurement instrument or index is like a successful model in that it must be applied under conditions where relevant causal capacities in the world can display themselves. Recall the relationship between causal relationships, models and the real world phenomena discussed by Cartwright.[1] Causal relationships require not only that capacities exist in the world but also that certain conditions be met in order for these capacities to be realised. Models are blueprints for nomological laws in that models describe specific kinds of situations in which the particular set of capacities will be realised in a repeatable way.

Indexes are created in order to measure or quantify concepts. But there is no one to one relationship between a number in an index and the concept being measured, and numbers can represent concepts only under particular conditions. For an index to be successful, therefore, we must determine the conditions under which the numbers will convey information that actually measures the concept, and indexes have to be used only when such conditions are actually met.

This chapter will discuss the situations in which measurement instruments or indexes are and should be used. I divide the discussion into two categories: (1) the quality and analytic acceptability of the data and of the aggregation of the data used in the index; and (2) the conditions of the countries or societies in which the index is used for policy purposes.

---

[1] Cartwright (1989).

The first category involves not only the 'practical limits' to the data used in the index but also some 'analytical limits' that were not discussed in the previous chapter. Practical limits involve difficulties in obtaining reliable data. For example, some countries have problems collecting reliable information on adult literacy rates because literacy tests are not conducted under proper supervision, and this is a practical limit. 'Analytical limits ' are problems of a data set that remain even assuming that the data can be practically collected. I have already discussed some analytic limits. For example, as noted, even if countries have reliable testing procedures to evaluate adult literacy rates, there remains an analytical limit in measuring a functioning 'to acquire knowledge;' just being able to read and write is not sufficient to achieve this functioning. However, there are other analytic problems in the HDI and in the LSI not discussed previously, particularly the problems of the use of a composite index in the HDI and the difficulty in using public opinion polls for the LSI.

The second category of discussion examines whether the two indexes are used for policy purposes in situations for which they are appropriate. Since the LSI has not yet been used for actual policy making in Japan, and the various ways in which the HDI may be used are not fully reported, there is a limit to this discussion. I will therefore focus on the analytic advantages and disadvantages of the use of these indices for policy purposes.

4-1: Quality of data and analytic problems with the approaches used in the HDI and the LSI

4-1-1: Quality of data

The HDI is designed to permit international comparisons of more than 170 countries in the world, including countries whose statistical systems are not yet developed. Accordingly, the four sets of data – life expectancy at birth, adult literacy rates, combined first-, second-, and third-level gross school enrolment ratio, and adjusted real GDP per capita (PPP$) – were chosen in part because most countries can produce these indicators. Still, even these indicators have data quality problems.

First, as the Report notes, there are sometimes difficulties in controlling the standard of the definition of literacy. Because different countries use different languages and different ways of writing, the HDI uses the general standard 'to be able to read and write with understanding.' Because 'with understanding' leaves room for interpretation, however, literacy tests could be made easier or more difficult for people to pass. In addition, a literacy test may be given without proper supervision, thus undercutting the reliability of the results.

A second data quality problem involves the use of estimates rather than actual data. Thus, even for these basic indicators in the HDI, for some countries it has been necessary to use estimates, including estimates produced either by the Human Development Report Office or by other organisations such as UNESCO. For example, the 1998 Human Development Report indicates that (1) 54 out of the 174 countries used figures estimated by UNESCO for their combined school enrolment ratio, and (2) 18 out of the 174 countries used 'provisional' figures for their real GDP per capita (PPP$).

Yet the success of the HDI in improving the quality of data should also be acknowledged. The process of creating and reporting the index has encouraged individual countries to produce better data sets. The Report claims that 'the publication of the *Human Development Report* is beginning to put pressure on all countries to improve their data systems and analysis, especially their social statistics.'[2] Also, as I write this dissertation, improvements have been reported in the quality of data, not only for individual countries but also for international organisations that collect and analyse data sets from all over the world.[3]

The LSI, on the other hand, uses data sets that are not available from all over the world. Instead, the LSI uses the results of annual opinion polls on the general level of satisfaction and the requirements for government action, which not all countries in the world produce. The index also relies on a large set of social and economic indicators from which only 12 are selected (using regression models) as variables that are significant in explaining satisfaction levels of different area of life. The LSI can afford to choose such non-basic indicators because the index is designed only to measure and analyse circumstances in a single country - Japan. The advantage of using data sets that come from one country (and the reasonable assumption that those data sets cover all of Japan with equal quality) is that they are more likely to reflect a consistent set of definitions and collections compared to data sets from many different countries. Application of a definition of literacy, for example, is likely to be more consistent within a country because of a shared culture and understanding.

---

[2] UNDP (1992), p. 21.

[3] The 1999 Human Development Report explained that 'this year's HDI is based on improved life expectancy data from the United Nations Population Division and revised data on adult literacy and combined gross primary, secondary and tertiary enrolment rations from UNESCO. Data on purchasing power parities (PPP) have been updated by the World Bank following the more comprehensive 1997-8 surveys by the International Comparison Programme (ICP).' (UNDP 1999, pp. 128-9)

Moreover, the LSI uses data that appear to be generally adequate. We can see this from the following facts on the data involved in the LSI:

1. The relevant data sets are available since 1975, and the regression analyses are made using data between 1975 and 1993.
2. The sample size of the opinion poll is 10,000.
3. The opinion poll is conducted once a year in May for those who are 20 and over.

The period for over which the time series analysis is conducted may be considered reasonably long (18 years) and the sample size large enough (10,000) with reference to normal econometric analysis. However, there is some concern that because the entire sample is taken only once a year in May, which is known to be a period that a number of people suffer from seasonal depression, the level of DSL may be underestimated. Moreover, there is a further concern with the use of subjective evaluations, which tend to be volatile and can be easily affected by events in the respondent's life that happened immediately before answering the questionnaire. We do not want the data to be affected by such temporary factors when our purpose is to find policy variables that improve satisfaction levels in life in general. Accordingly, in order to eliminate the possible underestimation of the level of DSL the sampling should be done in a time of the year other than May, and to minimise effects from temporary factors a larger sample size should be used.

4-1-2: Analytic problems with using the composite approach in the HDI

As I have noted in chapter two, using a composite index rather than vector comparisons without combining the three dimensions was critical in making

multilateral comparisons of human development. This sub-section examines the analytic difficulties in using such a composite index.

4-1-2-1: The weighting of the dimensions

First, in order to construct a composite index with three dimensions, we need to decide how to weigh the three dimensions, which involves deciding on their relative values. That is, we have to provide answers to questions such as, 'how many years more of life expectancy, how much more proportion of educational attainment and how much more income should be treated as equivalent to increasing human development by a certain fixed degree?' Answering this type of question is a problem because we often are not certain of the relative values or we do not want to make relative evaluations of different aspects of well-being. And yet, if an index is an aggregate one, the results require the use of relative weights.

In the case of the HDI, the Report explains that ideally they would like to use a 'meta production function' of human development to determine the weights. A 'meta production function' of human development would require a measure of human development independent from the HDI as an explained variable. The different HDI dimensions would be used as explanatory variables, and each would have a correlation co-efficient describing the degree to which the variable explains changes in human development. The method would solve the problem of weights by econometrically deriving the relationship between this measure of human development and the HDI dimensions.

The problem is, of course, that such a measure of human development independent from the HDI does not exist. Given this situation, the Report assigns equal weight to the three dimensions and provides or implies two kinds of

justifications for doing so, one conceptual and one empirical. The conceptual rationale is that these dimensions are equally important in improving human development.[4] The empirical rationale is that there exists a phenomenon which may be considered as an essential aspect of human development, and an econometric method shows that the phenomenon is explained well by assigning equal weight to the three variables.[5] Both justifications, however, have significant weaknesses.

The principal weakness of the conceptual justification involves the fact that in order to create an index with variables with different units, the variables have to be normalised (to obtain a common unit). Thus, depending on which common unit is used, what is implied by 'equally important' – that is, what results in an equal change in the HDI – is different.[6] For example, in the 1998 HDI, a 6 year increase in life expectancy, a 15% increase in the adult literacy rate, a 30% increase in the combined school enrolment ratio and a $3,990 (PPP) increase in income (if the original level of income is below $5,990 (PPP)), equally result in a 0.1 increase in the HDI.[7] However, as I described in chapter two, the units for each dimension were chosen based on fixed

---

[4] Desai (1991) notes that 'equal weighting is supposed to reflect the equal importance of these three variables.' (p. 355) This conceptual rationale is not expressly stated in the Report. But I regard the rationale as one of the justifications for the equal weights used in the HDI because as I understand, Lord Desai has been one of the principal designers of the HDI. Desai (1991), however, also adds to the claim quoted above that 'the equal weighting is not strictly true since the income variable is truncated and then concavified.' (p. 355) In the discussion below, I will explain the weakness of the conceptual justification of equal weight that is related to the weakness noted by Desai. Specifically, I will focus on the fact that sub-indicators are transformed (in order to normalise) before they are aggregated with equal weights.

[5] See Technical notes to the UNDP (1993).

[6] The Report raises this problem in the Technical notes to the UNDP (1993).

[7] These figures are derived from the following information about fixed minimum and maximum values for each indicator:

Life expectancy at birth: 25years and 85 years

Adult literacy: 0% and 100%

Combined gross school enrolment ratio: 0% and 100%

Real GDP per capita (PPP$): 100$ and $40,000 (PPP$), where the world average income of $5,990 (PPP$) in 1995 is taken as the threshold level beyond which a progressively higher discount rate is applied to income levels.

minimum and maximum values of each variable, not on how many additional years of life, how much of an increase in the school enrolment ratio, how much of an increase in the literacy rate, and how much of an increase in the per capita income (in purchasing power dollars) have an equally important impact on human development. So the fact that particular changes in the four variables result in the same change in the HDI does not show that the change in human development is equal in any meaningful sense.

Both Sen and the Report deal with the conceptual problems raised by the weighting of the factors in the composite index, albeit in somewhat different ways.

Sen argues that the necessity to weigh the variables should not be an obstacle to creating a measurement of capabilities and functionings. He claims:

> It is certainly clear that some types of capabilities, broadly conceived, are of little interest or importance, and even the ones that count have to be weighted *vis-à-vis* each other. But these discriminations constitute an integral part of the capability approach, and the need for selection and weighting cannot really be, in any sense, an embarrassment.[8]

Sen's stance is that because human development is multifaceted, selecting and weighting is a necessary part of its measurement. He even claims that 'the need for selection and discrimination is neither an embarrassment, nor a unique difficulty, for the conceptulization of functionings and capabilities.'[9] Sen justifies his claim by referring to the method taken to derive an income measure:

> The varying importance of different capabilities is as much a part of the capability framework as the varying value of different commodities is a part of the real-income framework. Equal valuation of all constitutive elements is needed for neither. We cannot criticize the commodity-centred evaluation on the ground that different

---

[8] Sen (1992), p. 45.

[9] Sen (1992), p. 44. Notice that Sen uses the word conceptualization, which indicates his awareness that the process of obtaining a measurement for an abstract idea involves refining and restricting the definition of the idea that can be expressed as conceptualization.

> commodities are weighted differently. Exactly the same applies to functionings and capabilities.[10]

To some extent Sen's appeal to the analogy of a commodity-centred evaluation is persuasive. It is certainly true that in order to quantify a concept such as 'total input' or 'total output,' which we use to obtain national income measures, we need to assign relative values to different commodities that constitute the total input or output. Using the current valuation of all constitutive elements is a way of assigning such relative values. Moreover, in order to obtain *real* income measures, which Sen refers to, we need to know the changes in price levels over time, and for this purpose it is customary to calculate price levels by selecting a basket of commodities representing the range of products in an economy. This 'basket' approach is used rather than the tracing of the prices of all the commodities that exist in a society both for practical reasons and because not all commodities (and the change in their prices) are regarded as relevant for a group of individuals whose income levels we are interested in. In addition, there would be multiple counting if the price index includes all the commodities. (For example, including both an increase in the price of crude oil and an increase in the price of petrol would be such a double counting relative to the real income levels of the public in general.)

However, I believe that there is not a precise analogy between a commodity-centred index and one measuring capabilities such as the HDI, and for this reason Sen's argument does not appear to be sufficient to defend the relative weighting of the sub-dimensions of the HDI. Specifically, the analogy between a real income measure based on commodities and capability measures is not adequate for Sen's argument for two reasons.

---

[10] Ibid., p. 46.

Firstly, in the case of income measures, we have a good idea about what *are* the goods and services produced in an economy, simply because most of them are exchanged in the market and we have a record of their market prices.[11] On the other hand, we have less confidence in our knowledge of what *are* the functionings that constitute 'human development' because it is something that someone or some people have to decide and there is not yet a consensus on such a decision.

Secondly, in an income measure, the relative values of the commodities are in principle determined by market prices.[12] 'Total input' or 'total output' is obtained by adding total commodities and services, which are valued in terms of their market prices. But for functionings and capabilities we do not have a valuation mechanism equivalent to the market. Accordingly, the relative valuations of the dimensions of the HDI are not already given as in the case of real income measures.

For these reasons, capability measures in the form of composite indices are potentially more problematic than indices of aggregate income, and therefore we need more caution in interpreting the results of the former indices.

As indicated above, lacking a 'meta-production function' for human development, the Report also provides an empirical justification for assigning equal weights to the three dimensions of the HDI. Specifically, the Report uses a so-called Principal Component Analysis (PCA) to show that the variables in the HDI can be

---

[11] As is well known, however, not all products are exchanged in the market, and therefore some are omitted from the record. Examples are goods that are exchanged in 'underground' markets, goods that are self-supplied, or goods that are publicly supplied.

[12] I say 'in principle' because some components of the income measures do not have market prices. For example, values for self-administered haircuts and the services of the police force or the government bureaucracy have to be determined in alternative ways. Also, some productive activities such as a spouse's household services are simply attached zero value in the current income measures because they are not exchanged in the market.

most comprehensively explained by assigning equal weights to the three variables.[13] Thus, the Technical Notes to the 1993 Report explain that a PCA conducted by Tatlidil showed that by assigning approximately equal weight to the three variables and also assuming a linear relationship among these three variables, approximately 88% of the total variance in the HDI could be explained, while the use of other weights resulted in a much lower explanation of the variance.[14]

The problem with the PCA analysis, however, it that it supports the reasonableness of weighting the three dimensions of the HDI equally on the *assumption* that these three dimensions are the total measure of human development. That is, if in fact there are other dimensions of human development not reflected in these three dimensions (and thus not included in the overall data set), then it is possible that some other weighting including these other dimensions would be preferred under the PCA analysis. As discussed above, there certainly are such additional dimensions. On the other hand, if we accept the limitation to these three

---

[13] The PCA primarily seeks to reduce the number of dimensions in a data set with a large number of interrelated variables (variables with common causes, variables that are causes of other variables, or variables that correlate with one another) by transforming the data set into new data sets called the principal components. (In the case of the PCA applied for sub-dimensions of the HDI, there is no reduction of the number of dimensions, but the three dimensions are simply transformed into new data sets that have a particular character, which the original data sets did not possess.) The characteristics of the transformation are 1) that the principal components are uncorrelated with one another, and 2) that they are ordered (the first principal component, the second principal component and so on) so that the first few components reveal most of the variations present in all the original variables. Each time the transformation is conducted to obtain a new principal component, the PCA results indicate econometrically 1) how strongly the variables are correlated with the overall variance in the original data, and 2) how much each variable accounts for the total variance of the original data. Each time we move to the next stage of the transformation (lower order of principal components), the new data set does not contain the variables that were previously accounted for, so that the PCA results for the new principal component explain only the remaining sources of variation.

[14] Specifically, the 1993 Report included a table showing that when the life expectancy component is weighted at 0.969, the adjusted income component is weighted at 0.916 and the educational attainment component is weighted at 0.925, 87.769 percent of the variance in the HDI data is explained. (See UNDP 1993, p. 109.) See also Tatlidil (1992).

dimensions, then it does appear from the PCA that the equal weighting of the three dimensions is reasonable.

4-1-2-2: The HDI and the policy trade-offs

A further analytic difficulty with using the HDI composite index involves the fact that although the HDI has a structure HDI = A + B + C, it is misleading to think that equivalent improvements in the values of A, B or C will necessarily lead to an equal improvement in human development. This is so principally because the HDI was not constructed to provide guidance on the trade-offs among the variables in the index. Instead, the HDI was created to provide a rough measure of the state of human development as an alternative to a development index based only on income levels, to encourage individuals and governments to focus on development in a more comprehensive way. As I have shown in chapter three, the particular dimensions included in the HDI – the average life expectancy at birth, an educational attainment index and the adjusted real GDP per capita in PPP$ - were determined for a variety of reasons, including universality, comparability, practicality (how easily the information can be known to policy-makers), and reasons related to capability theory, but not for the purpose of providing policy trade-offs among the variables in the HDI.

The Report deals with the issue of whether one can or should interpret HDI = A + B + C as expressing 'trade-offs' between the different factors expressed in A, B, and C that could be used for policy purposes. The Report points out that this is not possible in the following discussion:

> It would be tempting to interpret the relative coefficients as trade-offs, but a note of caution should be introduced. Superficially, it would be easy to say that one extra year of life expectancy is "worth" $150 of income, but these are not choices open to an optimizing economic agent. Take a poor country with a per capita income as high as $1500 (only 17 of the 65 countries with low human development in 1992 had

> income this high). An extra year of life expectancy (above a median value of about 50 years) would be the same as 10% growth in real per capita income. Neither of these two outcomes is likely in the short run, nor are they independent of each other in the real world. Thus, it would be wrong to interpret the coefficient as reflecting a "menu of policy choices".[15]

That is, the Report states that the relative coefficients do not reflect trade-offs between different aspect of human development for policy purposes for the following reasons: 1) as a matter of fact, neither an increase in life expectancy of a year nor a 10% growth in income (which would produce an equal increase in the HDI) are likely to happen in the short-run among countries with $1500 real per capita income (PPP), no matter what policies they take; and 2) as a matter of fact, the three aspects of human development measured in the HDI are not independent from one another in the real world, so, for example, government actions to improve per capita income may also improve the country's performance in the longevity and educational attainment dimensions. For these reasons the Report concludes that governments should not see the coefficients as alternative policy choices when they attempt to improve the levels of human development.

---

[15] UNDP (1993), p. 110. As was shown above, each variable in the HDI is normalised by taking the difference between the minimum and maximum values as a common denominator. Minimum and maximum ranges for the four variables are very different from one another, and therefore the actual effective weights are not equal as they appear from the formula HDI = A + B + C. The 'relative coefficients,' which are the actual effective weights, are derived subject to ranges of the following scales: 36.6 (= 78.6 – 42) years for life expectancy, 80.8 (= 99 – 18.2) percentage points for literacy, 12.2 (=12.3 – 0.1) years for mean years of schooling and $5,074 dollars for adjusted income for 1992. Taking these ranges into account, a one-year improvement in life expectancy, a one percentage-point increase in adult literacy, a one year improvement in mean years of schooling and a $1 increase in per capita income would represent the following changes in the HDI:

One unit change in	*Changes in HDI*
Life expectancy	1/108
Literacy	1/365
Mean years of schooling	1/108
Income	1/15,222

Note that 'relative coefficients' are dependent on which maximum and minimum range we choose in order to normalise the variables in the HDI.

Upon examination, the Report's first point does not seem very persuasive. It is true that in the very short-run, say, within a year, countries with low human development are not likely to achieve a 10% growth in income. Based on the income data of 1989 and 1990,[16] for example, in fact no low human development countries with per capita income levels higher than 1500 (PPP$) achieved income growths as high as 10%. The maximum growth rate achieved in this period among the 17 such countries was by Lesotho, whose per capita income improved by only 5.9% from 1646 (PPP$) to 1743 (PPP$). However, in the other dimensions, these countries were able to obtain improvements equivalent to a 10% growth in income under the HDI – that is, a 1 year increase in life expectancy or a 0.08 improvement in educational attainment index for a country whose per capita income is about 1500 PPP$ – even in the very short run. Thus, for educational attainment, 14 of these 17 countries (all except for Bolivia, Lesotho and Swaziland) achieved more than a 0.08 improvement in educational attainment between 1990 and 1992.[17] With respect to longevity, three countries – Bolivia, Cameroon and Lesotho – among the 17 countries improved their life expectancy by one year or more over the 2 years between 1990 and 1992.

Accordingly, this first reason the Report gives for not regarding the relative coefficients as trade-offs – that in the short-run the dimensions of human development in the HDI do not improve by one year of longevity, or by 150PPP$ of per capita income (both of which are equivalent to approximately an increase of 0.027 in the

[16] These data are obtained from the 1992 and the 1993 Human Development Reports.

[17] The Human Development Reports for 1993 and 1994 which I used for this analysis still used an educational attainment index consisting of the adult literacy rate and mean years of schooling, with two-third weight on the first and a one-third weight on the second. A 0.08 point increase in educational attainment could be achieved, for example, by a 8% improvement in the adult literacy rate and a 2-year's increase in the mean years of schooling.

HDI) – is not correct. At least in the case of longevity and of educational attainment such improvements can and do occur.

There is another analytic difficulty with this first point in the Report. Specifically, even if large improvements in any of the measures were not likely in a year, the coefficients could still be useful as 'a menu of policy choices' with respect to smaller changes. For example, a 5% increase in income, 0.5 additional years of life expectancy, and a 0.04 point improvement in educational attainment are equivalent under the HDI model, and as I have shown, such improvements do happen in reality. Using the same data sets as above, in fact I find that 10 out of the 17 countries noted above (countries with low human development but have income levels higher than 1500 PPP$) achieved more than a 5 % increase in income within two years, 12 among 17 such countries achieved more than 0.5 years increase in life expectancy within a year, and all the 17 countries achieved more than a 0.04 point improvement in educational attainment within a year. Policy-makers could use these as reflecting 'a menu of choices' to set priorities for the short-term policy goals.

The Report's second reason for rejecting use of the HDI factors in making policy choices – that there are likely interactions between the three aspects of human development – is more persuasive; that is, analytically it seems likely that such interactions will occur. For example, investments in the educational sector may not only improve literacy but also create more skilled labourers, which should contribute to the improvement of the country's per capita income, which in turn may improve the general health conditions of the population. In such cases, the HDI coefficients do not provide independent trade-offs for a government aiming to maximise human development with a given amount of resources. There also may be cases of negative correlations. For example, a government's attempt to improve per capita income by

investing more in fast-growing industries may lead to reduced investment in education and health sectors. In this case an effort to improve the country's development by improving the income component of the HDI may result in offsetting or partially offsetting declines in the other components. For this reason, the Report is correct in saying that governments should not regard the relative coefficients implicit in the HDI as reflecting a 'menu of policy choices' to achieve improved development.

In short, since the three components of the HDI are not independent of one another, governments aiming to improve their country's level of human development should not use the coefficients reflected in the HDI to determine their policy strategies.

This is not to say that the HDI cannot be useful at all in the area of policy. The ways in which the HDI can be useful for policy purposes are discussed in Section 4-2-1, below.

4-1-2-3: Limitation in the elements included in the HDI

A further analytic problem with the composite approach of the HDI is that it plainly does not capture all of the elements that are relevant to a comprehensive notion of human development. For example, as discussed above in chapter three, largely for political reasons, the HDI does not include any measure of the political and civil freedom that citizens in the country enjoy. The significance of this omission was noted in the recently released Human Development Report 2000,

> Even a composite indicator such as the human development index, while a broader measure of progress than gross national product, does not pretend to measure civil and political rights. Czechoslovakia had ranked higher in the human development index than in gross national product, indicating a fairer distribution of economic resources than that in many other countries at the same income level. But the index does

not measure the political dimension of rights – an area in which many one-party states were seriously deficient.[18]

In addition, as discussed in connection with the purpose-dependent criteria, notwithstanding the richness of human development, for practical reasons the number of dimensions in the HDI has to be limited. Accordingly, policy-makers cannot rely simply upon actions to improve the HDI since this would ignore many important areas of human development that the HDI does not include. Indeed, the 1991 Human Development Report notes this limitation in the HDI, stating as follows:

> Last year's Report went beyond defining human development by proposing a way to measure it. The human development index (HDI) combined national income with two social indicators – adult litracy and life expectancy – to give a composite measure of human progress. It was fully recognized then, as now, that the *concept* of human development it much broader than its *measurement.*[19]

The tension between the advantages of having some human development index measuring social dimensions and the analytic recognition that the index is certainly incomplete as a measure of the many aspects of human development was noted by Sen, one of the designers of the HDI. Thus, in an article included as a special contribution to the 1999 Report, Sen explained as follows:

> The HDI, which is inescapably a crude index, must not be seen as anything other than an introductory move in getting people interested in the rich collection of information that is present in the Human Development Report. Indeed, I must admit I did not initially see much merit in the HDI itself, which, as it happens, I was privileged to help devise. - - - *Why give prominence, it was natural to ask, to a crude summary index that could not begin to capture much of the rich information that makes the Human Development Report so engaging and important?*
>
> This crudeness had not escaped Mahbub [ul Haq, the originator of the Human Development Report] at all. He did not resist the argument that the HDI could not be [anything] but a very limited indicator of development. But after some initial hesitation, Mahbub persuaded

[18] UNDP (2000), p. 63.

[19] UNDP (1991), p. 15, emphases by the Report.

himself that the dominance of GDP (an overused and oversold index that he wanted to supplant) would not be broken by any set of tables. People would look at them respectfully, he argued, but when it came to using a summary measure of development, they would still go back to the unadorned GDP, because it was crude but convenient. As I listened to Mahbub, I heard an echo of T. S. Eliot's poem "Burnt Norton": "Human kind/Cannot bear very much reality".

"We need a measure", Mahbub demanded, "of the same level of vulgarity as GNP – just one number – but a measure that is not as blind to social aspects of human lives as GDP is."

- - - The crude index spoke loud and clear and received intelligent attention and through that vehicle the complex reality contained in the rest of the Report also found an interested audience.[20]

As Sen notes, the HDI is 'a crude summary index' that cannot 'capture much of the rich information' relating to human development. However, he and the author of the Human Development Report concluded even this 'crude summary' serves better than the GDP in monitoring whether a society is moving in the right direction in terms of its development.

4-1-3: Analytic problems with the LSI

As explained above, the LSI relies in part upon correlations between an opinion poll on overall satisfaction levels (DSL) and the respondents' identification of areas in which government improvements can be made. Various objective indicators are then developed for each of five basic areas – leisure, price, education, housing and employment – and correlations are developed between combinations of these objective indicators and the DSL. The result of this process is the identification of a series of objective indicators that is most closely correlated with the public opinion poll data on overall satisfaction levels. There are, however, a number of analytical problems with the resulting LSI.

---

[20] Sen's contribution in UNDP (1999), p. 23, emphases added.

One analytical problem involves questions of whether in fact a particular government has the capacity to improve the areas in which people express strong requirements for governmental action (as revealed by the public opinion poll). Specifically, in order for this data correlating satisfaction in life with requests for government intervention to be effectively used, government authorities have to have the credibility and the power to take actions to improve the objective indicators, such as the total working hours and unemployment rates in the area of employment. In some cases, thus connection is not present, for example, where the government has a low credibility among the public, and government intervention is associated with worsening, not improving living standards.

A second analytical problem concerns one implicit assumption of the DSL, that is, that the satisfaction level of one's life in general is the aggregation of the satisfaction levels of the five different components of leisure, price, education, housing and employment. However, the general satisfaction question does not ask the respondents to assess their overall satisfaction in terms of only these five areas, and there may be many individuals whose overall satisfaction is based largely on areas outside these five areas. In response to this problem, the general satisfaction question could be restructured as follows: 'How satisfied are you in your current life in general? Please rate your satisfaction in the following different aspects listed and base your general answer on those partial assessments.' This formulation of the question would essentially define 'a degree of satisfaction in general' based upon the satisfaction levels in the different identified aspects of life. But this restructuring of the question would create its own difficulty. Specifically, such a question would lose the virtue of being an overall subjective measure of happiness, which is the original aim of the LSI, and would limit the usefulness of the data from respondents as to

which different aspects of life were important to their judgements about their happiness (assuming the respondents in fact limited their responses by the indicated aspects). On balance, therefore, it appears best to avoid this type of criteria for the overall question on satisfaction and rely on the statistical correlations to identify the separate components such as the five components used in the LSI.

Finally, there is a more fundamental and somewhat complicated analytic concern with the use of a public opinion poll asking people about their satisfaction levels of life in general. This concern relates to two different possible uses of the information revealed in data from such introspective questions.[21] The first use involves the evaluation of the *individuals' attitude toward an object or an idea.* The second involves the *assessment of the value of an object or an idea.* In the former case, the focus of the measurement is to provide information on the state of *a person,* whereas in the latter the focus is on providing information on the conditions in the *outer world* that affect a person. An individual *A*'s answers to the questionnaire used in the opinion poll can analytically be used for both purposes; the answer reveals person *A*'s attitude toward his current life in general, and can also be seen as representing the value of the world currently surrounding *A*.[22]

What about the poll used for the LSI? If used for the first purpose, the polling data tell us that *N* number of people answered that they were satisfied in general about their life conditions, and *M* number of people answered that they were not satisfied, etc. In short, the result can be used for the first type of use to categorise and count the

---

[21] See Thurstone (1929).

[22] An analogy may be helpful to see the distinction. Assume a group of individuals is asked how much satisfaction they receive from eating apples and from eating oranges, and that the overall results show that apples are rated as 5 and oranges at 7 out of a possible 10. The first type of information we learn from this is that on average individuals are more satisfied eating oranges than apples. The second type of information would be that oranges are more important than apples.

number of people who had the same opinion towards life conditions that surround them.

There is, however, a particular problem in using the polling data for the second purpose, that is, to assess conditions in the outer world that affect individuals. This difficulty relates to the fact that the individuals responding to the poll are facing different states of life. It would be misleading to aggregate the results of the polling over all individuals and call the sum an evaluation of 'the value of the current state of life in general,' given that there is no single object of measurement among the individual respondents.

Therefore, the aggregate direct results of the opinion poll are most appropriately used for the first purpose, that is, as an evaluation of the mental states of individuals. This focus on the mental states of individuals does not presuppose the existence of a common concept or criterion of 'satisfaction' or 'satisfactory life,' and recognises the fact that individuals are evaluating different states of affairs using various different criteria for judging their degree of satisfaction. As long as the opinion poll results are used to show the proportion of people who answered that they were satisfied or unsatisfied in their life in general, there are no assumptions made about the conditions of these evaluation that are difficult to meet in reality. The only assumptions made may be that people answer the questions honestly.

4-2: Conditions under which the HDI and the LSI can reasonably be used for policy purposes

The above discussion allows us to consider the further issue of whether and under what conditions the HDI and the LSI can be useful for policy purposes, which purposes could include 1) constructing recipes for policy intervention by the

governmental or international organisations to improve the level of human development or satisfaction and 2) framing policies that focus on certain areas of development in the hope of improving the level of human development or satisfaction.

4-2-1: The HDI

One possible policy use for the HDI would be to use the HDI as a guide to the choice of policy by viewing the components of the HDI as 'trade-offs' for alternative methods of improving human development in a particular country. That is, a government or international organisation could target programmes that would improve the score that the country would obtain on the HDI in an effort to improve the level of human development in the country.

As developed above in Section 4-1-2-2, however, the drafters of the Human Development Report have explained that the three components of the HDI should not be used as a guide to policy 'trade-offs' made to improve human development, and I concluded that at least one of their express concerns was reasonable. That is, the national governments or international organisations cannot reasonably use the HDI for constructing policy recipes to improve the level of human development. The HDI was not developed for this purpose, and it appears likely that the three components are interrelated either positively or negatively.

This is not to say that the HDI cannot be used by international bodies or individual countries in order to improve the level of human development. Although it does not appear reasonable to use the HDI as 'trade-offs' to develop the detailed recipes for policy intervention, the HDI can be helpful in the policy area in at least two respects.

First, the HDI can be useful as a means of policy evaluation, that is, the HDI can be used after the fact to provide some measure of whether the policy strategies utilised by a government are in fact improving human development. This policy evaluation process could allow government planners and international organisations to more effectively evaluate not only whether a particular policy programme is effective, but to allow for some comparisons of the effectiveness of different policy approaches over time or among different countries based upon the HDI results (assuming the relevant socio-economic conditions remain stable over time or over the countries involved in the comparisons).

Second, the HDI could be effectively used to identify areas of development that the government should focus on in order to improve the country's level of human development. This would not involve using the HDI to provide detailed recipes for policy intervention, but rather would result in using the HDI in a more general way to identify areas of improvement in human development. That is, for example, a country could realise that its score on the longevity component of the HDI is weak, and focus its development efforts on programmes to improve longevity by reducing infant mortality, etc. (keeping in mind the possible impact of these efforts on other areas of human development).

Even these last two uses of the HDI for policy purposes, however, are limited by some conditions, including the following:

First, the HDI's policy usefulness is limited by the extent to which the data upon which the HDI calculations are made can be reliably collected. As noted above, however, some of the factors – such as the literacy standard of 'being able to read and write with understanding' – are subject to varying possible interpretations, and in other cases the HDI calculations rely upon estimates provided by the Report or by

other organisations such as UNESCO rather than actual data for the countries involved.

Second, for a variety of reasons – including political considerations – the HDI does not include factors that are arguably important for human development, such as political freedom. The exclusion of such factors limits the usefulness of the HDI as a measure of human development, particularly in those situations where these excluded factors are significant. For example, in countries with repressive political systems, the HDI would provide an overstated calculation of the actual state of human development.

Finally, although there is some econometric analysis (that is, the PCA) to support the equal weighting of the three dimensions of the HDI, this analysis is limited to an explanation of these components and does not, of course, reflect any considerations not reflected in these dimensions.

4-2-2: The LSI

The relationship of the LSI to policy development is fundamentally different from that of the HDI. Specifically, as explained above in Chapter Two, the LSI, unlike the HDI, was expressly developed to provide a recipe for policy intervention to improve the satisfaction level in the Japan population. Accordingly, the factors in the LSI were designed to allow policy-makers to construct specific policies in this area. However, as with the HDI, the LSI's usefulness for policy purposes is limited by a number of conditions, including the following:

First, the LSI's usefulness is limited by the extent to which the Japanese government can take effective actions to improve conditions in the five areas identified generally and in the related objective indicators specifically. To the extent

that the government is incompetent or corrupt and cannot improve these conditions, the LSI will not be a very useful tool for policy purposes.

Second, the LSI's usefulness is limited to the extent that individuals in Japan obtain significant degrees of satisfaction in areas other than the five areas identified of leisure, price, education, housing and employment. For example, none of these areas deals directly with matters of personal freedom, freedom of worship or spiritual dimensions of life, and these could be important aspects of overall satisfaction. In his recent book Robert William Fogel emphasises the importance of spiritual dimensions in constructing policies essentially in America, but his observations would also apply to other nations with comparable levels of material well-being such as Japan. Fogel writes:

> In a world in which all but a small percentage are lacking in adequate nutrition and other necessities of life, self-realization may indeed seem like mere ornament, but not in a country where even the poor are rich by past or Third World Standards. That is the case in America today since the poverty line is at a level of real income that was attained a century ago only by those in the top 10 percent of the income distribution. ---- Failure to recognize the enormous *material* gains of the last century, even for the poor, impedes rather than advances the struggle in rich nations against chronic poverty, whose principal characteristic is the *spiritual estrangement* from the mainstream society of those so afflicted. Although material assistance is an important element in the struggle to overcome spiritual estrangement, such assistance will not be properly targeted if one assumes that improvement in material conditions naturally leads to spiritual improvement.[23]

A final and most fundamental limitation of the LSI's usefulness for policy purposes involves, firstly, the important distinction between a correlation and a causal relationship, and secondly, the nature of causal relationships as 'capacities' rather than 'necessities.' As explained above, the LSI was developed through establishing a correlation between the various objective indicators developed for each of five areas

[23] Fogel (2000), p. 3, emphases by Fogel.

and the general satisfaction levels as measured by opinion polls in Japan. Thus, in the area of housing, correlations were found with three objective indicators, 'budget for utilities,' 'gini-coefficient for land assets,' and 'total areas of parks in cities.' However, a problem in using these correlations for policy purposes is that such relationships developed from regression models may only involve correlations but not causation.

The second aspect of this fundamental limitation is that even if we know that the relationship is a causal one, according to Cartwright, causal relationships are 'capacities' that exist in nature, and because they are capacities rather than necessities, the particular causal relationship does not always result. Certain conditions have to be met for the relationship to actually result.

For example, we may know that an increase in the total area of parks in cities causes an improvement in the satisfaction level because of housing factors, *ceteris paribus*. However, a causal relationship with overall satisfaction may not always actually result. Suppose that an increase in the total areas of parks is accompanied by an increase in land prices because of a reduction in the supply of residential property with no change in demand. If the degree of dissatisfaction caused by the land price inflation exceeds the degree of satisfaction caused by an increase in the total areas of parks, the positive causal relationship between the total areas of parks and the satisfaction in the area of the housing aspect of life will not be useful for policy purposes.

In other cases the statistical relationships or correlations between satisfaction levels and the objective indicators may reflect only the presence of a common cause. For example, in the LSI, the basic indicator for the education area was found to correlate with 'education fees,' 'the number of students absent from high schools' and

'the number of youth crimes.' It seems likely, however, that an increase in the number of students absent from high schools does not *cause* the satisfaction levels in the education sector to fall. The two indicators are likely to correlate because they have a common cause, such as 'the number of students per teacher' or 'teaching quality.' That is, as the number of students per teacher increases or the quality of teaching declines, less personal attention or less effectual teaching to the students causes some students to stop attending school. In this case, as a matter of policy the government should be seeking to affect the common cause of the two variables that correlate.

The above examples show that in order for the Japanese government to use the relationships between objective indicators and satisfaction levels in the LSI, we need to know that either of the following circumstances exists:

1. We need to know that there is a direct causal relationship between the objective indicators and the satisfaction level, and that all relevant conditions are satisfied so that the causality works; or
2. We need to know the common cause for the objective indicators and the overall satisfaction level data, or other causal structures that give rise to the correlation, and that conditions are satisfied so that a change in a cause actually gives rise to an expected effect – in our case an improvement in the satisfaction level of a particular area of life.

Only when the government is confident that either of the two types of conditions is met, can it use the objective criteria to improve the satisfaction level of a particular area of life.

* * *

The chapter examined the ability of the HDI and the LSI to measure their respective concepts - human development and the level of satisfaction in life in

general – and to use these measurements for policy purposes. There were two aspects of this examination, one involving the quality of data currently in use and analytic problems and the other involving whether the index is used under appropriate conditions. Such appropriate conditions depend on 1) the purposes of the measurement, 2) the techniques used in the index (e.g. principal component analysis or regression models), and 3) the knowledge frontier (e.g. the availability of an independent measure of human development or knowledge about various causal relationships in the world). From the discussion of this chapter, I conclude that the probability that an index is actually able to measure accurately a particular concept of interest and to be useful for policy purposes will be increased by the following factors:

1. Collection of better and larger sets of data,
2. Use of the index for purposes that are consistent with the way it is designed, and
3. Improvement of our knowledge frontier.

Conclusion

This dissertation developed my views on proposed criteria for successful indexes and examined two indexes that measure human values, the HDI's measurement of human development and the LSI's measurement of satisfaction. A philosophical foundation of the dissertation is that there is no one-to-one relationship between a concept of human value and an index that quantifies the concept. Therefore, no index can be *the* measure of concepts such as 'human development' or 'satisfaction level of life in general.' On the contrary, I conclude that there are a variety of indices that are non-trivially different from one another, any one of which, if used properly, can successfully measure a human value. The success of an index in measuring a human value depends on comparisons to criteria that are relative to the purposes for which the index is made.

The view I have developed on the relationship between concepts, indexes and the phenomena in the world is analogous to the view on the relationship between theories, models and phenomena by Cartwright, Morgan, and Boumans, as well as Chang's views on the relationship between concepts, measurement instruments and real world phenomena.[1]

Cartwright claims that theories are true only in models, which are the blueprints for sets of conditions under which 'capacities' in the world can display themselves in regular behaviours. Each model focuses on a specific set of capacities that is assumed to give rise to a particular real world phenomenon, which often is a product of a number of capacities operating at the same time. The model describes the situation under which the particular arrangement of capacities gives rise to the phenomenon in question.

Morgan finds that early econometricians were not searching for models that verify a particular fundamental economic theory, but were looking for models that would successfully explain empirical data. Their success was measured by a wide range of concerns that included usefulness for policy and the capacity for further developments of theories.

Boumans argues that satisfactory models are those that combine various factors that include theory and data in ways that meet one or more criteria, such as theoretical, mathematical and statistical requirements or usefulness for policy.

Chang's study on the measurement of temperature proposed a criterion of comparability (or stability) in the measurement results, to solve the circularity involved in selecting an instrument to measure concepts such as temperature.

A philosophical position shared by these four people is that a theory cannot produce *the* model or *the* measurement instrument that represents a phenomenon in the world. My view on abstract concepts, indexes and the real world was developed from this philosophical position: I believe that even where a theory is available to define an abstract concept such as human development or satisfaction, there is no single corresponding measure for the concept.

So how can we tell good models or good measures from bad ones? I sought solutions in the same direction as Cartwright, Morgan, Boumans and Chang did, that is, (1) to introduce external criteria and (2) to make the success of an index conditional (as Cartwright does in making a success of a model in displaying causal relationships conditional). More specifically, I urged that indexes must be examined in the light of the detailed background and motivations behind the particular measurement project, and that indexes have to be used for the purposes for which they

[1] Cartwright (1983), (1989), (1997), forthcoming, Morgan (1988), Boumans (1999a) and Chang (1996).

are designed. Indeed, for this examination of the HDI and the LSI I have put more emphasis on the purposes and motivations of the measurement project than the four people who influenced me for the following reasons:

1. Since there are so many different ways in which a concept of human development or satisfaction can be quantified even using a theory that defines the concept, it is particularly important to identify who makes the index and for what reason in order to select a particular measure for the concept.
2. The purposes of the project can dictate which particular aspects of a relevant theory are important in selecting and examining the index (e.g., the importance of examining the index's ability to reveal the relative necessities of individuals in the HDI)

In short, given that purpose-dependent criteria play an important role in the creation of the HDI and the LSI, in order to examine and criticise the indexes in a constructive manner, we need to take into consideration the backgrounds of each particular measurement project that has led to the selection of the specific components and structure of the index.

There are several benefits from using criteria of successful indexes that are dependent on their purposes. Firstly, we realise that a number representing a human value could be a product of conflicting requirements imposed by concept-independent concerns. Such conflicts are especially prominent in the case of the HDI: considerations of universality required the HDI to use only broad and general definitions of human development, but in order for the index to discriminate among countries more specific definitions for each dimension of human development were needed. Also, in order for the HDI to give a complete ordering of more than 170

countries we need a composite index, but we do not have strong justifications for assigning the relative weights to the three dimensions of the index.

Secondly, by knowing the purpose for which the index is made, we can identify the range of applicability of the index. For example, the LSI is deigned to be used for making policies to improve the general satisfaction level in Japan, while the HDI is created to give a rough indication of countries' levels of development in a more comprehensive sense than the mere growth in income.

Finally, criticisms and discussions about the index will be more specific and relevant if we take into consideration who makes the index and for what purposes. Examining the HDI, for example, without taking into account that it is made by the UN is likely to create criticisms that are not useful, such as the claim that the index is too simple to capture the rich concept of human development. More interesting and relevant questions arise only after acknowledging why the index is so simple, because then we can focus on how to avoid the misuse of the simple index and also how to improve the index given the constraints.

Postscript

In this thesis, I have developed my own account of successful indexes and examined two examples – the Human Development Index and the Life Satisfaction Indicator – according to the criteria. The investigation was based on the view that criteria for successful indexes are relative to particular projects of quantification. The nature of those projects define many things: the indexes' relationship to theories, the purpose for which the indexes are made, as well as the index-makers' ideas about what are appropriate objectives for individuals and societies. I did not investigate the more fundamental question of whether enhancing 'human development' or improving 'satisfaction' are good for humans. This fundamental question is concerned about the philosophies of index-makers, which I took as givens in my account. In this section I would like to give a cursory summary in response to the crucial question: Are human development and increased satisfaction necessarily good for humans? The question should be part of any serious treatment of measurements of human development or satisfaction, but it is a complex one that needs a detailed treatment. I outline here some of the areas to be included in a treatment of this question. (A more complete discussion of the question will be one of my future research projects.)

Human development has been defined as enlarging choices. And Amartya Sen's capability approach to well-being discussed in chapter three, which provides the theoretical foundation of the HDI, was based on the idea that widening of options is a good. So the first question to be asked is about the truth of the claim that choice enlargements are good for individuals. I find two problems or questions in this claim:

1. Is *any* additional choice good for individuals? In other words, are having more options to choose from always better than having fewer?

2. Is having choices sufficient as a human good?

In a simplistic approach to this first question we can respond that having more options must be better than having fewer options, because those with more options can choose to live with fewer options (assuming that options include 'cutting options'), but those with fewer options have but to live with them. For this reason we may say that even choices such as taking drugs, which itself may not be a good, will add to the goodness in person's life because that person has larger options in life to choose from. But some may object to the claim by saying that some options simply do not add goodness for a person's life, or could even worsen it. In fact, although Sen generally claims that 'choosing may itself be a valuable part of living, and a life of genuine choice with serious options may be seen to be – for that reason – richer,' he does not forget to add that 'this is not to say that every additional choice makes a person's well-being go up - - -.'[1]

But by distinguishing *freedom of choice* from *achieved well-being of a person*, we can maintain a positive relationship between freedom and advantage. I use an example given by Sen, specifically the case where a person comes across a crime scene.[2] Here, for the sake of the argument, we assume that the person still has a choice of not getting involved in the crime scene and also that the person's general capability to achieve well-being (for example his ability to make rational decisions) is not affected by encountering the event.[3] By actually opting for the newly added choice (of trying to prevent a crime through getting involved in the crime scene), the

[1] Sen (1992), p. 41.

[2] Sen (1992).

[3] If we do not assume these conditions, then a person's *freedom to achieve well-being* goes down, as Sen points out.

person's *well-being* may well decrease if he gets injured. Still, the person's *freedom of choice* increases, in a sense that the person can now choose to prevent a crime.

But there are also cases where having more options is not better than having fewer because the very fact of having many options has a negative influence on the person. For example an article in the Financial Times July 24th 1999 describes difficulties (and sufferings) that today's highly educated young professionals face in making decisions about which course of life to follow. This may be considered as a counter example to regarding choice enlargement as necessarily a human good. Sen is aware of the possibility 'that increased freedom might be *disadvantageous* to the person by forcing on the person the necessity to spend time and effort in making lots of choices that he or she would rather not have to make.'[4] Still, he claims that this is not a conflict between freedom and advantage in general, because the freedom to achieve the preferred form of life is still enhanced by a loss of options for those who would rather have fewer alternatives to choose from.[5]

Regarding the second question, having choices themselves cannot be sufficient as a human good unless humans are capable of making use of the range of choices. In order to view choices themselves as good for humans, therefore, we implicitly assume that individuals are autonomous beings and are actually capable of making use of choices for whatever goals they have in their life.

Another problem we may have in regarding choice enlargement as good in itself is that the subject himself may not be aware of all the choices he has in his life. Can we say that a person's well-being is larger than he thinks it is because he has a talent which he is not aware of? Fukuda makes this point.[6] He is critical of Sen's idea

---

[4] Sen (1992), p. 62, emphasis by Sen.

[5] Sen (1992), p. 63.

[6] Fukuda (1995).

of evaluating well-being in terms of capabilities because people may not be aware of all their capabilities, and to evaluate their well-being based on something that the subject does not even know of is not appropriate. Such unknown capabilities of a person could be brought to the person's attention by outsiders, but for Fukuda outsiders pointing out the person's capabilities and claiming that the person can do better is equally awkward.[7]

Fukuda, on the other hand chooses 'satisfaction' as a basic human good and as an object of improvement. As the discussion in chapter three has shown, Fukuda's choice is based on his sympathy towards Bentham's idea of 'utility' defined in terms of pleasures and pains. If we take satisfaction as human good, neither of the problems I raised against the choice enlargement approach to human development arise: any negative aspect of having additional options, if there is such an aspect, would be taken into account when an individual evaluates his or her satisfaction level; and the subjective evaluation of satisfaction comes only after capabilities are used and give rise to such mental states.

However, treating satisfaction as a human good has its own problems. Sen raises some of these problems.[8] Firstly he argues that because of our tendencies to learn to adjust our desires to the realistic ones and also to be affected by the consideration of practical possibilities, those who have very few resources and little prospect of obtaining more may learn to be satisfied with their situation, while those who are already affluent and expect to earn more may never be satisfied. Sen's argument against any subjective evaluation of well-being is not so much a criticism of

[7] Ibid., pp. 64-5.
[8] Sen (1985).

satisfaction or happiness as a human good but the consequence an evaluation on this basis could bring if used for distributive purposes. If satisfaction is used as an indicator to equalise well-being of a society, those who are already affluent could end up getting more and those who have very little may not receive any because the former tend to be less satisfied than the latter. One defence Fukuda may provide is that the distributive problem is less severe in Japan, which is known to be a country where the variance of the income level is relatively small. If the LSI type of indicator is used for policy-making in countries where income distributions are more unequal, the criticism Sen makes would be more serious.

Sen also provides an argument directly against treating satisfaction or happiness as the main objective of improvement in life. His claim is that happiness or satisfaction, even if is it is good in our lives, is only one aspect of goodness in life.[9] Under Sen's view, a measurement of satisfaction, no matter how good the measurement instrument is, cannot identify the whole of human good. This view applies even from the point of view of *individual* well-being, because even for a subject himself, satisfaction may not be the only element that makes his life a good one. This disagreement on the importance of satisfaction or happiness is a significant and highly contested area, and one I look forward to exploring in a future research project.

[9] See Sen (1985), p. 12.

Appendix 1: A method of deriving life expectancy

m: age-specific death rates;

q: mortality rates;

l: the number of persons living at the beginning of an age-interval;

L: the number of person-years that would be lived within an age interval by the cohort of 100,000 birth assumed;

T: the total number of person-years that would be lived after the beginning of an age-interval by the cohort of 100,000 birth assumed.

$${}_nq_x = 1 - e\,[-n \cdot {}_nm_x - an^3 \cdot {}_nm^2{}_x]$$

where n is size of the age interval, ${}_nm_x$ is the central death rate of the age interval between x and x + n, a is a constant, and e is the base of the system of natural logarithms. The constant a used in the Reed and Merrell model is $a = 0.008$.

$$l_{x+n} = (1 - {}_nq_x)l_x$$

$$T_x = -.20833\, l_{x-5} + 2.5\, l_x + .20833\, l_{x+5} + 5\, \Sigma_{\alpha=1} l_{x+5\alpha}$$

If the age intervals are 5-year, and

$$T_x = 4.16667\, l_x + .83333\, l_{x+10} + 10\, \Sigma_{\alpha=1} l_{x+10\alpha}$$

If the age intervals are 10-year.

The life expectancy (e_x) is computed as the ratio of T_x to l_x.

References

- Anand, S. and Sen, A. (1992) 'Human Development Index: Methodology and Measurement' Background paper for Human Development Report (1993), the UNDP, New York.
- Bentham, J (1789) An Introduction to the Principle of Morals and Legislation, edited by J. H. Burns and H. L. Hart (1996), Clarendon Press, Oxford.
- Boumans, M. (1999a) 'Built-In Justification' in Models as Mediators; Perspectives on Natural and Social Science, edited by Mary S. Morgan and Margaret Morrison, Cambridge University Press, Cambridge, pp. 66-96.
    - (1999b) 'Representation and Stability in Testing and Measuring Rational Expectations' in Journal of Economic Methodology, 6:3, pp. 381-401.
- Cartwright, N. (1983) How the Laws of Physics Lie, Clarendon Press, Oxford.
    - (1989) Nature's Capacities and Their Measurement, Clarendon Press, Oxford.
    - (1997) 'Models: The Blueprints for Laws' in Philosophy of Science, 64.
    - 'The Limits of Causal Order: from Economics to Physics,' in Fact or Fiction?, edited by U, Maki, forthcoming.
- Chang, H (1996) 'Spirit, Air and Quicksilver: the Search for the "Real" scale of Temperature,' Text for monthly seminar for the British Society for the Philosophy of Science, 13 May, 1996 at the LSE.
- Dasgupta, P. (1990) 'Well-Being in Poor Countries' Economic and Political Weekly, August 4: 1713-20.

- (1993) An Inquiry into Well-Being and Destitution, Clarendon Press, Oxford.

- Desai, M. (1989) 'Potential Lifetime (PLT): A Proposal for an Index of Social Welfare' in Towards a New Way of Measuring Development, Caracas: Office of the South Commission.

- (1991) 'Human Development - Concepts and measurement' in European Economic Review, Vol. 35, 350-7.

- (1995) 'Measuring Political Freedom' in LSE on Freedom, edited by E. Baker, LSE Books, 195-225.

- Dornbusch, R., Fischer, S. and Startz, R. (1998) Macroeconomics 7th ed., Boston, McGraw-Hill/Irwin.

- The Economist, June 26th – July 2nd, 1999.

- Fogel, R. W. (2000) The Fourth Great Awakening – & the future of egalitarianism, The University of Chicago Press, Chicago.

- Frank, R. H. (1997) 'The Frame of Reference as Public Good,' The Economic Journal, 107 (November), 1832-47.

- Fukuda, K. (1995) Nihon wo Yutaka ni su ru Houteishiki [The equation that enriches Japan], Nihon Hyoron-sha.

- Fuller, B. (1986) 'Raising School Quality in Developing Countries' in World Bank Discussion Papers: 2, The World Bank, Washington, D. C.

- Heady, B. W. (1993) 'An economic model of subjective well-being: integrating economic and psychological theories' in Social Indicators Research 28.

- Herrera, H. (1976) Catastrophe or New Society, International Development Research Council.

- Jolliffe, I. T. (1986) Principal Component Analysis, Springer-Verlag, New York.
- Kolm, S. C. (1994) 'The Meaning of "Fundamental preferences"' in Social Choice and Welfare, 1994, No. 11, 193-8.
- Lucas, H. (1985) Life Expectancy as an Integrating and Demographic Data: A Summary of Six Country Studies, Development Centre of the Organisation for Economic Co-operation and Development.
- Morgan, M. (1988) 'Finding a satisfactory empirical model' in The Popperian legacy in economics: papers presented at a symposium in Amsterdam, December 1985, Cambridge University Press, Cambridge.
    - (1990) The History of Econometric Ideas, Cambridge University Press, Cambridge.
- Morgenstern, O. (1963) On the accuracy of economic observations 2nd ed., Princeton University Press.
- Ng, Y. K. (1997) 'A Case for Happiness, Cardinalism, and Interpersonal Comparability,' The Economic Journal, 107 (November), 1848-58.
- Oswald, A. J. (1997) 'Happiness and Economic Performance,' The Economic Journal, 107 (November), 1815-31.
- Sen, A. K. (1983) 'Poor, Relatively Speaking' in Oxford Economic Papers, Vol.35, March, No. 1, pp. 153-69.
    - (1985) Commodities and Capabilities, Amsterdam, North-Holland.
    - (1992) Inequality Reexamined, Clarendon Press, Oxford.
- Shryock, H. S., Siegel, J. S. eds. (1976) The Methods and Materials of Demography, Academic Press, New York.

- Tatlidil, H. (1992) 'A New Approach for Human Development: Human Development Scores,' IDS, Sussex. Mimeo.
- Thurstone, L. L. (1929) 'The Measurement of Psychological Value' in Essays in Philosophy T. V. Smith and W. K. Wright ed., The Open Court Publishing Co., Chicago.
- Townsend, P (1979) 'The Development of Research on Poverty' in Department of Health and Social Security Social Security Research: The definition and measurement of poverty, London, HMSO.
    - The United Nations (1998) Basic Facts about the United Nations, United Nations Department of Public Information, New York.
    - The United Nations Development Programme Human Development Report 1990 - 2000.
- Van Praag, B. M. S and Frijters, P. (1997) 'Choice behaviour and verbal behaviour: a critical assessment of their relevance for practical policy,' a presentation for the Tinbergen Conference, Empirical models and policy making, May 1997.
- Veenhoven, R. (1993) Happiness in nations, Rotterdam: Risbo.