1

What is national wellbeing and why measure it?

> not all the calculators of the National Debt can tell me the capacity for good or evil, for love or hatred, for patriotism or discontent … at any single moment in the soul of one of these its quiet servants.
>
> Charles Dickens, *Hard Times*

This book is about social progress: its definition and its measurement. In particular, it is concerned with the overall wellbeing or quality of life of the people in a nation at a point in time, how this has changed (i.e. 'progress') and whether it is sustainable. In exploring this topic, the book seeks to address three primary questions:

What is national wellbeing?

Why should national wellbeing be measured?

How should national wellbeing be measured?

Underlying those three primary questions lie others, including:

What is individual wellbeing?

What is wrong or inadequate with existing measures of progress?

How do we *measure* national wellbeing, rather than just describe the state of the nation when we measure specific aspects of wellbeing?

How do the current and future states of the environment, including stocks of natural resources, fit into our understanding of wellbeing?

All of these questions are interwoven. For example, the definition of wellbeing and its method of measurement are two sides of the same coin, and the meaning of *national* wellbeing depends both on what is meant by *individual* wellbeing and on how one *aggregates* individual

The Wellbeing of Nations: Meaning, Motive and Measurement, First Edition. Paul Allin and David J. Hand.

values to produce a national value, as well as whether there are aspects of national wellbeing which are distinct from individual wellbeing, which should somehow be included as part of the definition.

Formally, perhaps we should take the opening three questions in the order given above: start by defining the concept, then give the motivation for measuring it and then describe how to measure it. That, however, would make for rather dry reading. It would rely on the forbearance of the reader, who would have to plough through the definition before getting to the reason for reading the book and then to how it was to be done. Better, we think, to begin with the motivation, so that the pressing need for the ideas and tools described in the book serve to drive the reader on. Once the motivation has been established – once we can see the need for such an exercise – then we can dig down into precisely what it means and how we might go about it.

For this reason, we begin in the next section with the motivation: what is wrong with current approaches, current tools and current strategies for gauging the state of society and whether it is advancing or regressing. We then move on to discuss the nature and aspects of individual wellbeing and how to measure the wellbeing of individuals, before embedding this in the larger context of national wellbeing. The measurement of national wellbeing certainly involves aggregating the wellbeing of the individuals in the nation, but it also involves other aspects, such as higher level societal properties which are not evident at the level of individuals as well as other factors which may influence wellbeing and permit improved measures. As the UK Office for National Statistics report *Measuring National Well-being: Life in the UK, 2012* put it (Self *et al*., 2012, p. 3 and see Chapter 7 below): 'The well-being of the nation is influenced by a broad range of factors including economic performance, quality of life, the state of the environment, sustainability, equality as well as individual well-being.'

It may be that the individuals within a population appear to be fine, while the larger picture shows something rather different (we are reminded of the parable of the turkeys, congratulating themselves on how wonderful life is, as Christmas approaches). Conversely, and as does appear to be reflected in some of the measures, aspects of individual wellbeing may not show things in so positive a light as do some aspects of national wellbeing. For example, individual anxiety about crime may increase even while measures of actual crime rates are decreasing; likewise, increased longevity might be taken as a sign of national wellbeing, even while the ailments associated with advancing age may lead to lower individual wellbeing scores. Such discrepancies, when they occur, need to be explored and explained. (And, for the last particular example, this is one reason why Self *et al*., 2012, look at 'healthy life expectancy' rather than simply 'life expectancy'.)

Before delving into the details, however, some introductory comments are appropriate.

First, the measurement of wellbeing is a big subject. This book is very much our perspective on it: other authors will doubtlessly have different views of what is important, and will place their emphasis on different topics.

Second, and related to the first point, the book will not seek to provide an exhaustive list of relevant measures. Apart from the sheer size of such a task, it is probably impossible, simply because the area is a dynamic one, experiencing growth and change. Any list we produced would be outdated by the time this book was published. We limit ourselves to taking stock of current developments and drawing in depth on some of them.

This dynamic growth is a consequence of a third characteristic of the area: it is currently the focus of a huge amount of research attention. On the one hand, this means that governments and other actors are exploring how best to apply such ideas in policy formulation, while on

the other hand it also means that the very concepts themselves, along with how to measure them, are mutable and are still being refined and polished. Some people suspect that this refinement will take a long time, pointing out that the system of national economic accounts has been refined over decades. Waiting for the refinement to be completed (if it ever is) would appear to be out of the question.

Fourth, this book is primarily about *measuring* wellbeing. It is not about the policy implications of the results, although we do recognise that policy uses of any new measure need to be taken into account in designing and delivering it. Likewise, it is only in part about what influences wellbeing – the fact that, for example, marital status can have an effect – though inevitably there is some discussion of such matters: causes and effects may be indicators of the extent of wellbeing and thus might be used to improve measures of wellbeing, via regression or calibration models.

The history of measurement in general is one of gradual encroachment, as concept after concept succumbed to the advance of quantification. But it has been a slow and painful advance. Almost every step forward in measurement technology has faced opposition from those who thought attempts to measure or quantify some attribute were impossible or meaningless:

> Such pretensions to nicety in experiments of this nature are truly laughable! They will be telling us some day of the WEIGHT of the MOON, even to drams, scruples, and grains - nay, to the very fraction of a grain! - I wish there were infallible experiments to ascertain the quantum of brains each man possesses, and every man's integrity and candour: - This is a desideratum of science which is most of all wanted. (Harrington, 1804, p. 217)

The history of attempts to measure wellbeing can be traced back as far as one likes. Writing 130 years ago Edgeworth remarked 'hedonism may still be in the state of heat or electricity before they became exact sciences' (Edgeworth, 1881, p. 98). It is perhaps an indication of the difficulty of 'hedonimetry', as Edgeworth calls it, that it is only relatively recently that sound underpinning models and theories for measuring wellbeing have been developed.

Measuring wellbeing is also characterised by its multidisciplinary nature. Psychologists, sociologists, economists, statisticians, medical researchers, ecologists and others all have something to say about how it should be done. In part this is because they have different potential uses for such a measurement in mind, but in large part it is because the issue is fundamentally multidisciplinary. Wellbeing may be a characteristic of an individual, but it is at least partly a reflection of social interactions and is influenced by external forces.

1.1 Motivation: Why measure wellbeing?

United Nations Resolution 65/309 says:

> '*The General Assembly …*
>
> 1. *Invites* Member States to pursue the elaboration of additional measures that better capture the importance of the pursuit of happiness and well-being in development with a view to guiding their public policies;

2. *Invites* those Member States that have taken initiatives to develop new indicators, and other initiatives, to share information thereon with the Secretary-General as a contribution to the United Nations development agenda, including the Millennium Development Goals;

3. *Welcomes* the offer of Bhutan to convene, during the sixty-sixth session of the General Assembly, a panel discussion on the theme of happiness and well-being;

4. *Invites* the Secretary-General to seek the views of Member States and relevant regional and international organizations on the pursuit of happiness and well-being and to communicate such views to the General Assembly at its sixty-seventh session for further consideration'.

The UK Office for National Statistics report *Measuring National Well-being: Life in the UK, 2012* (Self *et al*., 2012, p. 3) said:

'In particular, having a more complete picture of national well-being will lead to:

- better understanding of policy impacts on well-being;
- better allocation of scarce resources via more informed policy evaluation and development;
- comparisons between how different sub-groups of the population are doing, across a range of topics;
- more informed decisions on where to live, which career to choose, based on well-being information for that area/organisation;
- assessments of the performance of government;
- comparisons between the UK with other countries'.

Implicit in these two extracts is the fact that a (perhaps the) key role of a government is to ensure the wellbeing of those it governs. That in itself probably provides sufficient answer to the question posed in this section's heading.

Only if one can measure wellbeing can one tell if the government is succeeding and, even more, if progress is being made. Furthermore, government policies ought, above all, to be evidence based, and usually that evidence will be quantitative, so that *measures* of wellbeing are critical. Such measures, when developed for groups, at more than the individual level, will allow one to monitor change (are things getting better or worse?), to compare groups (do the sexes have different degrees of wellbeing? Do different ethnic or social groups progress differentially?) and generally to investigate the impact of policies (have changes to education systems enhanced overall quality of life?). There are then other general questions which require answers. For example, what explains the frequently observed U-shaped distribution of wellbeing (with greater wellbeing being observed in younger and older age groups)? What explains the weakness of the correlation between income and wellbeing over time? Why in many countries are aggregate levels of happiness much the same as they were at the end of the Second World War, despite the dramatic growth in income per capita since then? And so on.

From the time of the Second World War to the present, economic measures, typically in the form of national economic accounts, have taken the primary place in monitoring progress. Measures such as gross domestic product (GDP) and gross national product (GNP) are headline indicators. But economic indicators focus on just one aspect of life. They do not touch on health, for example (other than through the cost of health service provision). There is concern in some quarters that standard measures of economic performance and progress are not really suited to the policy decisions which they are being used to inform. In 2009, the economists Joseph Stiglitz, Amartya Sen and Jean-Paul Fitoussi, wrote in the preface of their report on measuring societal wellbeing, commissioned by Nicholas Sarkozy, the then French President:

> We see the world through lenses not only shaped by our ideologies and ideas but also shaped by the statistics we use to measure what is going on, the latter being frequently linked to the former. GDP per capita is the commonly used metric: governments are pleased when they can report that GDP per capita has arisen [sic], say, by 5%. But other numbers can tell a very different picture. In Russia, declining life expectancy suggests there are underlying problems, even if GDP per capita is rising. So, too, in the United States, most individuals saw a decline in income, adjusted for inflation, from 1999 through 2008 - even though GDP per capita was going up - providing a markedly different picture of performance. Such a disparity may arise when income inequality increases at the same time that income increases. (Stiglitz *et al*., 2010, p. xix).

There are other similar examples: Egypt showed a per capita GDP increase from $4760 to $6370 between 2005 and 2010, and yet the proportion of survey respondents classified as 'thriving' declined from 29% to 12% (OECD, 2013, p. 27).

An example of the disconnect between crude economic indicators such as GDP and more general measures of wellbeing is given in the United Kingdom by current intensive debate about whether a third runway should be built at Heathrow. From a narrow economic perspective the answer is clearly yes. The United Kingdom is suffering, relative to other European countries, from its limited capacity for long-haul flights to other parts of the world, especially the Far East. But from the perspective of the wellbeing of people in West London and surrounding villages living under the flight paths of 1300 wide-bodied jets taking off and landing every day, economic measures only scrape the surface of what is meant by wellbeing. Likewise, they fail to tap into the environmental impacts of increased air travel.

So GDP misses central aspects of what people regard as important. GDP was not designed to be an overall measure of wellbeing, so it is not surprising that it is now judged inadequate in that respect. More generally, as we shall see, perhaps *economic* measures themselves are inadequate or insufficient. After all, contentment, happiness, quality of life etc. are influenced by more than mere financial wealth or income. The richest person, suffering constant pain from an incurable disease, may well not rate their wellbeing as very high. Likewise, exhaustion of natural resources may lead to a short-term benefit, but it will mean the consequent enhanced quality of life will not be sustainable.

Furthermore, some things which increase economic activity, and hence GDP, are best regarded as decreasing wellbeing. A famous comment made by Robert Kennedy (which we will quote in Chapter 2) illustrates this, referring to such things as jails, napalm and nuclear

warheads as contributing to GNP but hardly enhancing national wellbeing. Sitting in traffic jams consumes fuel and hence adds to GDP, but is hardly beneficial to society.

Another reason for the divorce of measures such as GDP from the wellbeing of any one person is that the former are necessarily *aggregate* measures. Any measure at the national scale must be an aggregate measure, but it is not, a priori, clear that the same concepts or ideas should apply at the national level as at the individual level. To take a trivial analogy: we can talk about the variance or the skewness of a set of values (as, e.g. wealth distributions are typically skewed, with long right tails), but we cannot talk about the variance or skewness of individual values. This example, of skew wealth distributions, is particularly relevant to wellbeing, as there is evidence that *inequalities* in the distribution of resources across a population have a negative impact on perceived wellbeing. Compelling evidence for this is adduced by Richard Wilkinson and Kate Pickett in their book *The Spirit Level* (Wilkinson and Pickett, 2010). As Stiglitz *et al.* (2010, p. xi) remark: 'One of the reasons that most people may perceive themselves as being worse off even though average GDP is increasing is *because they are indeed worse off*' (their italics). In general, it is entirely possible for the majority of people to experience a decrease in some variable while the average still increases.

In fact, even in the context of straightforward economic indicators, GDP has peculiarities as a measure of a country's 'economic wellbeing'. It excludes revenue earned from overseas production but paid to a country's residents, and it includes income paid to non-residents. It also ignores internal household work, such as childcare, and, at a more fundamental level, a level which is increasingly being recognised as of fundamental importance, GDP takes no account of sustainability: are finite natural resources being consumed, so that current 'economic progress' is being achieved at the cost of a future loss?

Most important of all, however, is the straightforward limited scope of GDP, in that it focuses on economic measures and fails to tap aspects which are regarded just as central to wellbeing, such as security, health, social networks, freedom to pursue what interests one, leisure time and so on.

Crime will decrease wellbeing overall in society. This is apparent even in such measures as GDP, where fraud has a negative impact. For example, the collapse of the Kabul Bank in Afghanistan as a result of fraud cost the country 5% of its GDP (BBC, 2012). But keeping such gross examples aside, GDP completely fails to tap into other aspects of crime which arguably have a much more direct adverse effect on individual wellbeing. So-called 'quality-of-life' crimes, including low-level things such as broken windows, littering and vandalism have a tiny economic impact but a large wellbeing impact. As we stress throughout the book, different aspects of wellbeing are interlinked in a complex way, so that minor quality-of-life crimes may have a larger impact through the subliminal messages they send about the extent to which people (including the authorities) care.

Chapter 6 discusses the shortcomings of GDP in more detail, but the fact is that GDP was never intended to be used as a global indicator of *wellbeing*. Its dominant role as an indicator of economic progress gives a very one-sided view of a multifaceted concept. This is a point to which we will repeatedly return, especially when we consider whether it is in fact possible to measure wellbeing by a single index, or whether something more is needed – a 'dashboard' of indicators, perhaps.

Having said all the above, GDP does have some attractive properties. We should learn from this when we try to develop more appropriate measures of wellbeing. For example, GDP adopts a monetary numéraire, translating *everything* into financial units. This means one can

balance different attributes against each other. We shall say more about this when we discuss *conjoint measurement* in Section 1.7.

The report by Stiglitz *et al*. (Stiglitz *et al*., 2010) makes five recommendations relating to GDP to improve its relevance as a measure of living standards.

Recommendation 1: When evaluating material wellbeing, look at income and consumption rather than production.

Recommendation 2: Emphasise the household perspective.

Recommendation 3: Consider income and consumption jointly with wealth.

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

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

These recommendations will be discussed in detail in Chapter 6, but here we merely remark that recommendations 1 and 3, and to a certain extent 4, might be regarded as a manifestation of the Micawber philosophy (Dickens, 1850): 'Annual income twenty pounds, annual expenditure nineteen pounds and six, result happiness. Annual income twenty pounds, annual expenditure twenty pounds ought and six, result misery'. (This refers to Great Britain's pre-decimalisation currency, in which there were 240 pence in a pound.) Recommendation 2 is concerned with the distinction between the (almost) individual level rather than the aggregate national level and recommendation 5, with the fact that economic measures are not everything. With the exception of 5, these recommendations still refer only to economic performance. Wellbeing goes beyond that.

The Stiglitz report goes on to cover aspects beyond the economic, making a further seven recommendations.

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

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

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

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

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

Recommendation 11: Sustainability assessment requires a well-identified dashboard of indicators. The distinctive feature of the components of this dashboard should be that they are interpretable as variations of some underlying 'stocks'. A monetary index of

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

Recommendation 12: The environmental aspects of sustainability deserve a separate follow-up based on a well-chosen set of physical indicators. In particular there is a need for a clear indicator of our proximity to dangerous levels of environmental damage (such as associated with climate change or the depletion of fishing stocks.)

Recommendation 6 needs to be read in conjunction with a list of seven aspects of life presented by Stiglitz *et al.* as the key dimensions to be taken into account, ideally simultaneously, in measuring quality of life. Six of the seven dimensions are actually mentioned in the recommendation: they all need improved measures. One of these, the environment, is linked to sustainability, the subject of Recommendations 11 and 12, because it is meant to address future conditions as well as the present condition of the environment. The other dimension of quality of life, material living standards, covers income, consumption and wealth, which are also the subjects of Recommendations 1–5. So, while there is no precise definition of wellbeing here, the recommendations do cover economic performance, quality of life, the environment and sustainability. We will discuss the recommendations in detail in Chapters 4 and 6.

Finally, as a last illustration, the Belgian Federal Planning Bureau (2005) describes three areas of capital that can be used and developed.

1. *Human capital: comprising the standard of living (material well-being), health (both mental and physical) and knowledge/capacities (what individuals know and are able to do).*
2. *Environmental capital: including both natural resources (water, air, land and mineral resources) and the biosphere with all its biological diversity.*
3. *Economic capital: subdivided in physical and technological capital (equipments, buildings, infrastructure, and intangible assets including software and technology patents) and net financial assets.*

Only one of these three is primarily about economics.

1.2 What is individual wellbeing?

Wellbeing has been the subject of a very considerable amount of research, stretching back decades, especially within the psychological and medical communities, and has been defined in various ways. For example:

> Wellbeing … comprises objective descriptors and subjective evaluations of physical, material, social and emotional wellbeing, together with the extent of personal development and purposeful activity, all weighted by a set of values. (Felce and Perry, 1995)

> Well-being is a complex construct that concerns optimal experience and functioning (Ryan and Deci, 2001).

> Well-being can be understood as how people *feel* and how they *function*, both on a personal and a social level, and how they *evaluate* their lives as a whole (Michaelson *et al.*, 2012).

Other, closely related terms are also used, including such things as an individual's 'quality of life' and 'life satisfaction'. Indeed, Easterlin (2003) opens his discussion paper on 'building a better theory of well-being' by saying 'I take the terms wellbeing, utility, happiness, life satisfaction and welfare to be interchangeable'.

Fayers and Machin (2000, p. 3) comment:

> It is clear that QoL [Quality of Life] means different things to different people, and takes on different meanings according to the area of application. To a town planner, for example, it might represent access to green space and other facilities. In the context of clinical trials we are rarely interested in QoL in such a broad sense, but are concerned only with evaluating those aspects that are affected by disease or treatment for disease. This may sometimes be extended to include indirect consequences of disease such as unemployment or financial difficulties.

The word 'happiness' also often crops up. It appears, for example, in the extract from the UN General Assembly resolution quoted at the start of Section 1.1 – though note that it is invariably in the context of the phrase 'happiness and well-being' rather than in isolation. This suggests that happiness is different from wellbeing. However, any differences are often lost in popular media accounts, which frequently characterise the measurement of wellbeing as the measurement of happiness. The New Economics Foundation report, *Measuring Well-being: A Guide for Practitioners* (Michaelson *et al*., 2012, p. 6), has this to say about happiness (based on a well-established understanding of wellbeing in psychological terms):

> It is worth pointing out that well-being is not exactly the same as happiness. **Happiness** often refers to how people are feeling moment-to-moment and does not always tell us about how they evaluate their lives *as a whole* (although it can do), or about how they *function* in the world. **Well-being is a much broader concept** than moment-to-moment happiness: it includes happiness but also other things such as *how satisfied people are with their lives as a whole*, and things such as *autonomy* (having a sense of control over your life), *purpose* (having a sense of purpose in life).

Wellbeing and quality of life are different from 'standard of living', which is taken to refer primarily to economic aspects – income, wealth and general material conditions such as quality of accommodation, access to health care and so on. It is clear from this that quality of life and wellbeing are broader concepts, since they also cover such things as social life, the environment and cultural activities and also subjective happiness.

Things are then further complicated by the existence of specialised variants, such as Activities of Daily Living scales, Health Related Quality of Life (HRQoL) and others. Activity of Daily Living scales measure the extent to which people can perform personal activities and normal self-care functions. They cover such things as dressing, bathing and personal hygiene, feeding oneself, organising things at home and so on. One might argue that a high state of wellbeing would be reflected by a high score in such areas. HQoL scales often

cover similar things, though they may be relative to what the patient might expect, given their condition. That last point leads to a potential complication in wellbeing measures in general: the response to a question such as 'overall, how satisfied are you with your life?' might well be qualified (perhaps not explicitly) by 'under the circumstances (of your condition) . . . '.

None of the above should be surprising. Wellbeing is an intrinsically complex and multi-dimensional concept, with different parties naturally having interests in different aspects, so to expect a single universally agreed definition would be to expect too much. But it can lead to problems. Fayers and Machin (Fayers and Machin, 2000, pp. 3–4) say that 'In the absence of any agreed formal definition of quality of life, most investigators circumvent the issues by describing what *they* mean by quality of life and then letting the items (questions) in their questionnaire speak for themselves'. As we discuss in Section 1.4 and Chapter 4, defining the concept in terms of its measurement procedure is fine. What is less fine, however, is producing a superficial definition which fails to meet the basic criteria of a measurement or indicator. These issues are discussed in Section 1.5. More generally, if one does have multiple descriptors, each addressing some aspect of wellbeing, then in some circumstances one can sidestep summarising them into a single number, instead of reporting a profile: painting a picture rather than a single score.

Wellbeing, especially in the psychological sense of flourishing, is in some sense the opposite of clinical conditions such as depression. There has been a huge amount of work on measuring depression – see, for example, Dunn *et al.*, 1993; Santor *et al.*, 2006. The latter notes that 'since 1918, more than 280 measures of depressive severity have been developed and published'. It is perhaps interesting that relatively little of the work on wellbeing makes explicit reference to the work on depression, despite the complementary relationship between the two. For example, Table 1 of Santor *et al.*, 2006, lists symptom domains covered by various depression scales. This list includes sleep, irritability, anxiety, hopelessness, suicidal tendency, concentration, energy/fatigue, worthlessness, agitation, withdrawal, no interest in others and so on, all of which researchers in wellbeing will recognise. More generally, Kahneman *et al.* (2003, preface) note that, in the context of hedonic psychology – the psychology of 'pleasurable and unpleasurable states of consciousness' – 'another characteristic of past research is the remarkable accentuation of the negative. Textbooks that do not mention pleasure or well-being at all devote many pages to the clinical phenomena of anxiety and depression'.

A high level informal definition of wellbeing might be that it is 'what matters to people', and indeed the phrase 'What matters to you?' was adopted as the strapline for the study on wellbeing policy in the United Kingdom and the measuring national wellbeing programme of the UK Office for National Statistics (see Chapter 7). This informal definition drives home the fact that wellbeing is something to be hoped for, aspired to and aimed at.

Section 1.5 discusses the basic principles of measurement. We shall see that measurement procedures have two aspects – sometimes termed their *representational* and *pragmatic* aspects. Representational aspects are those concerned with preserving the relationships between the objects under study, when mapping from the system being measured to a numerical representation. Pragmatic aspects are those concerned with deciding what characteristics are relevant, and how they should be captured within the measurement index (the 'items' in the terminology of Fayers and Machin, though pragmatic aspects go much further than the mere choice of questions in a questionnaire).

Wellbeing measures typically have a large pragmatic component, and so to a large extent their definition and measurement procedure are one and the same; the specification of the

way wellbeing is measured is also a definition of what is meant by wellbeing. The large pragmatic component inevitably means that not everyone will agree on a definition – perhaps because they have different intended uses for the indicator in mind or because they have different philosophical bases for what they mean by measurement. For example, even if you and I agree that having a supportive social network and being healthy are key contributors to wellbeing, we might disagree on the relative weight to be assigned to them in our overall index.

In summary, in general we should expect there to be multiple measures of wellbeing: different measures, that is, different *definitions* of wellbeing, will be suitable for different purposes – just as different measures of inflation are appropriate according to whether one is seeking a macroeconomic indicator or a cost of living indicator.

1.3 Aspects of individual wellbeing

We have stressed that individual wellbeing has multiple aspects. The potential constituents of individual wellbeing can be classified in various ways. In this section we shall look at some useful distinctions. We shall discuss how the constituents might be combined to yield an overall measure in the next section.

McAllister (2005) classified constituents into five types. 'Most researchers agree about the domains that make up wellbeing: physical wellbeing; material wellbeing; social wellbeing; development and activity; emotional wellbeing. The elements can be paraphrased as physical health, income and wealth, relationships, meaningful work and leisure, personal stability and (lack of) depression. Mental health is increasingly seen as fundamental to overall health and wellbeing. These elements are sometimes viewed as "drivers" of wellbeing.'

In contrast, the Organisation for Economic Co-operation and Development (OECD) study *How's Life: Measuring Well-Being* (OECD, 2011a) and the OECD *Compendium of Well-being Indicators* (OECD, 2011b) identified two pillars of wellbeing. The first was *material living conditions*, covering such things as income and wealth, jobs and earnings and housing. The second was *quality of life*, covering such things as health, work/life balance, education, civic engagement, social connections, environmental quality, personal security and subjective wellbeing. Then, in addition to these two, they added *sustainability*, necessary if current levels of wellbeing are to be maintained over time. Data availability and other issues meant this last one was not discussed in detail in the *How's Life* report, although in other work the OECD does report on sustainable development and green growth indicators and how these might be used within the broader picture of national wellbeing (e.g. Strange and Bayley, 2008). It is this broader picture of national wellbeing in which we are particularly interested, seeing future and current wellbeing as integral to 'how a country is doing'.

We see immediately from this that the OECD definition does not treat quality of life as synonymous with wellbeing, but rather as a subcomponent of it. Since subjective wellbeing is regarded as a component of quality of life, at first glance it might look as if a rather circular definition results, but in fact overall wellbeing includes other, objective aspects, noted above.

A less high-level classification of the constituents of wellbeing, focussing more on the quality of life aspects, is into cognitive and affective components, with the affective component being further partitioned into positive and negative aspects. This is a quite common distinction. Thus, for example, the *Commission on the Measurement of Economic Performance and Social Progress* (see Stiglitz *et al.*, 2010) asserted that 'subjective well-being (encompasses)

three different aspects: cognitive evaluations of one's life, positive emotions (joy, pride) and negative ones (pain, anger, worry)'. Particular, measures, (e.g. the Positive and Negative Affect Schedule, Watson *et al.*, 1988) focus on specific aspects.

And digging down yet further, the New Economics Foundation guide to measuring well-being (Michaelson *et al.*, 2012), focuses very much on subjective wellbeing, and identifies three aspects for practitioners to consider:

– how people feel, covering emotions such as happiness and anxiety;
– how people function, covering such things as 'sense of competence' and 'sense of being connected to those around them';
– how people evaluate their lives, covering such things as satisfaction with life, and comparison with the best possible life.

We can see from the above that how wellbeing is defined – and hence how it is measured – depends very much on the level at which it is approached. A high-level perspective will include things such as the freedom to express political opinion. A low-level perspective might focus very much on subjective wellbeing. There also is a clear link to the notions of exogenous and endogenous variables in economics, with the former being external things which influence wellbeing, and the latter internal aspects of wellbeing. In the next section we draw a related distinction between 'influences on' and 'consequences of' when constructing measures of wellbeing. This is a distinction which has not been widely recognised, but could have an important impact on the quality of the measurement system which results. Traditionally, in wellbeing measurement, all related variables have been treated in the same way.

An important distinction underlying notions of wellbeing is that between *hedonic* and *eudemonic* wellbeing. Rather than being alternatives, however, they might properly be regarded as complementary aspects of a larger whole.

The hedonic perspective might be seen as the 'naive' perspective, and is certainly the viewpoint that popular media often take. This perspective regards wellbeing as a degree of happiness. In contrast, the eudemonic perspective might be characterised as regarding wellbeing as a fulfilled potential and meaning in one's life.

Notions of hedonic wellbeing can be traced back to the ancient Greeks, in a balance of pleasure against displeasure. Utilitarianists such as John Stuart Mill and Jeremy Bentham held that the proper aim of actions should be to maximise 'utility', defined in terms of positive happiness and negative suffering. For example, Bentham's *An Introduction to the Principles of Morals and Legislation* begins

> Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do . . . By the principle of utility is meant that principle which approves or disapproves of every action whatsoever according to the tendency it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words to promote or to oppose that happiness. I say of every action whatsoever, and therefore not only of every action of a private individual, but of every measure of government. (Bentham, 2009)

Nobel Laureate Daniel Kahneman and his co-editors, in their book *Well-being: The Foundations of Hedonic Psychology*, defined 'the new field of hedonic psychology' as being about

'what makes experiences and life pleasant and unpleasant' (Kahneman *et al.*, 2003). The perspective is very much a subjective one.

Like hedonic wellbeing, notions of eudemonic wellbeing can be traced back at least as far as the ancient Greeks, with, for example, Aristotle characterising simple (hedonic) happiness as suitable only for beasts. In contrast, eudemonic wellbeing arises when one's life is most in accordance with one's beliefs and attitudes, when there is personal development and a sense of achievement or fulfilment. Descriptions such as 'positive functioning', 'fully functioning', 'personal growth', 'meaningfulness', 'acting with integrity' and so on are often associated with eudemonic happiness.

As we mentioned in Section 1.1, a key motivation underlying the recent interest and work on measuring wellbeing is the intention to use such measures as indicators of the effectiveness of public policies. This will only be possible if wellbeing can be influenced and changed by such policies. That raises a number of issues.

The obvious one is whether wellbeing is a psychological trait or state. A trait is fairly constant over time, whereas a state is something which depends on circumstances and experiences. Happiness, for example, will probably usually be regarded as a state. If wellbeing is a trait, then, while measuring it might be of scientific interest, it would be of limited value for policy decisions. In contrast, if wellbeing is a state that could be influenced by external events and experiences, then it clearly does have policy implications. The distinction is tied to the notion of homeostasis.

A homeostatic system is one which, when perturbed by shocks, tends to revert to its original condition – an equilibrium state, one might say. If a state of wellbeing (whether low or high wellbeing) is such a homeostatic equilibrium, then measurements of it would not be expected to change much over time (at least not in the aggregate), so that again wellbeing would be of purely scientific interest, rather than policy interest. Indeed, a classic study of lottery winners and accident victims suggested that this homeostatic notion might be true, with the happiness of both groups later returning to their pre-incident levels (although this result has recently been criticised by Helliwell, CSLS, 2011). This phenomenon has also been termed *the hedonic treadmill* (Brickman and Campbell, 1971).

We must also ask ourselves whether different people are likely to have different basic degrees of wellbeing. For example, it does not seem unreasonable to suppose that someone predisposed to depression will have a lower degree of wellbeing than others. In a meta-analysis, DeNeve and Cooper (1998) found that subjective wellbeing and various personality traits were correlated, with (perhaps unsurprisingly) extraversion and agreeableness being positively correlated and neuroticism negatively correlated.

Then, even if measured wellbeing does (or can) change over time, we need to be sure we can identify the cause of the change, and determine that it is not merely due to measurement characteristics or other kinds of change. The Flynn effect illustrates the dangers. This is the phenomenon, discovered by James R. Flynn, that IQ scores have increased from generation to generation over the past century (Flynn, 2012). One explanation is that successive generations are becoming more intelligent, and not in a minor way – the changes have been described as massive and for some populations are much larger than one standard deviation of the IQ distribution. But this seems unlikely on other grounds – the evidence of history shows that our great grandparents were not idiots. A far more convincing explanation is that people are becoming more skilled at the kinds of abstract reasoning which IQ tests typically measure. Any observed changes in measured wellbeing would not be attributable to the same cause (increased skill in abstract reasoning), but we need

to be confident that they do not have some cause other than the policy change we may be investigating.

This sort of phenomenon is related to issues of cultural difference. Apart from measuring the impact of public policies, one of the ways that wellbeing measures will be used is to compare groups. Nationally, these will be social, ethnic, gender, age and other groups within a country, but the measures will also be used internationally to compare nations. The difficulty is, as countless cross-cultural comparisons have found in the past, that such comparative statements are fraught with problems. At the very simplest level, translations of questionnaires can change the meaning, and we know, even at the most basic of levels, that question wording can have a huge impact on responses. Moreover, as Christopher (1999) has pointed out, definitions of wellbeing are necessarily culture dependent. The problem of identifying equivalent constructs across different cultures may well defy effective solutions.

Another way of classifying aspects of wellbeing is into 'objective' and 'subjective' aspects. An objective aspect will be something external to the individual but which is thought to have an impact on their wellbeing or be indicative of their state of wellbeing. A subjective aspect will be (for example) a simple self-reported state of someone's wellbeing. Note, however, that it is not the self-reporting which makes the indicator subjective, but rather the fact that it is reporting on an internal (subjective) aspect of their condition.

Yet another classification is into those which represent some kind of predictability, choice or control and those which do not. For example, if one feels one has made one's own choice about education, daily personal activities, use of leisure time, political voice and so on, then one is likely to feel more at ease with oneself than if one feels these choices were imposed upon one. Being told what to do all the time is not conducive to psychological wellbeing. The notion that one is able to control and influence things, rather than being subject to the whims of fate, is something which will recur in our consideration of wellbeing. The feeling of being unable to control things may be one reason why unemployment is generally associated with poor wellbeing beyond the extent to be expected as a consequence of loss of income.

Likewise, one can distinguish between aspects which represent security and those which do not. For example, feelings about health, crime, a safe place to live, good nutrition, strong social networks and so on will all impact the degree of wellbeing.

Various objective measures have been identified as central to wellbeing. They include the following.

- *Education*: An educated workforce is necessary for a modern technological society, and so education and economic progress are related. But more than this, we have already mentioned the importance of choice and control, and education gives people greater choice and control. Education breeds opportunity. A higher level of education is also associated with better health, lower unemployment and greater social engagement, though clearly the causal links may not be straightforward. On the other hand, there are subtleties. For example, there is the complication that perhaps education can promote inequality. It might lead to a greater dispersion in income over a population. Such things are discussed below, in the light of evidence that societies which have higher degrees of inequality may have lower aggregate measures of wellbeing (Verme, 2011). This illustrates the importance of measuring the shapes of distributions, not merely their means or medians. Alternatively, increased education might lead to increased dissatisfaction if appropriate jobs are not available.

Human capital measurement, as quoted from the Belgian Federal Planning Bureau document at the end of Section 1.2, invariably focuses on the labour market benefits of education and training. This has the result that someone unemployed or retired has no human capital. Of course, there is a wider concept of human capital, defined in terms of contributions to society and to one's own fulfilment and wellbeing.

– *Personal activities*: Again, having the freedom to pursue one's own interests represents an aspect of choice and control, and so is likely to promote wellbeing.
– *Political voice and governance*: Stiglitz *et al.* (2010, p. 78) stress 'political voice [as] an integral dimension of the quality of life'. Once again, this is closely related to notions of control, rather than being subject to the (potentially arbitrary and unpredictable) whims of someone else. It permits individuals to feel they have a say, at a fundamental level, in how their lives proceed. This opens up important cross-cultural and cross-national issues: if political voice is to be included as a contributory factor in a wellbeing measure, how should countries with different political structures be compared? These ideas are also closely related to trust in public institutions, showing why responsible and accountable regulatory authorities are so important.
– *Security*: This is almost too obvious to need mentioning, but a feeling of safety – from crime, war, accidents, economic crises and so on – is conducive to higher degrees of wellbeing. Security is, of course, also related to health – the feeling that one is safe from disease or illness. As far as crime is concerned, we must distinguish between the actual level of crime and the perceived level of crime. For example, 'when people have a perception of crime as a problem in an area, the experience of crime or worrying about crime does not offer much additional purchase in understanding their quality of life' (Christmann and Rogerson, 2004). Indeed, anxiety about crime can go up while true crime rates go down. Economic security and the anxieties which can arise from it, are particularly pertinent at present, in the straitened economic conditions consequent on the banking crisis. This includes things like financial insecurity, employment insecurity (unemployment, as we have noted, impacts wellbeing to an extent beyond the obvious financial hit), as well as long-term insecurity as people fear what may await them in retirement.
– *Social networks*: Strong social support networks induce benefits at several levels. At the basic level, there is a correlation between strength of social network and quality of life, and there is considerable evidence that social support is associated with better health and wellbeing in general. Of course, there may be issues of causal direction. At a higher level, clearly a larger social support network means that one is more likely to be in employment, and possibly have a higher level of education. The positive impact of being married (at least, for men) is one aspect of this and, in general having an intimate partner confers some protection from the potentially detrimental consequences of adverse life events. All these things are correlated.
– *Environmental conditions*: Immediate environmental conditions have an impact on our everyday quality of life. The contrast between living next to a busy motorway and a tinkling stream illustrates this. But there is also a larger impact on health through pollution or on cost of living due to droughts impacting food prices, for example. Moreover, unsustainable costs or resource consumption incur a debt to the future. It means that future quality of life will be proportionately lower. This is why more recent thinking includes sustainability having an impact on wellbeing. One can see particular

difficulties here, because of the timescales involved, and the need to determine likely impacts in the future.

- *Material living standards*: This is the aspect which has received most attention in the past, so perhaps it is unnecessary to say much about it here. We will return to it below.
- *Health*: In Chapter 2 we quote the World Health Organisation definition of health, as contained in its constitution: 'Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity'. Again we have something of the circular in the definitions: Measures of wellbeing often include health as an aspect, but the WHO definition of health includes wellbeing. Despite this, health clearly has a central impact on wellbeing.

We have already mentioned that all these various aspects are correlated. Blunt description fails to capture the subtle causal relationships, and indeed to tease these out would be a vast project. But it also fails to identify the fact that there are probably synergistic relationships. Being both unemployed and ill is likely to have a greater impact than the mere 'sum' of the parts. At the least this means that one cannot simply combine the results of instruments measuring the distinct aspects of wellbeing, but have to look at things from a holistic perspective: both the measures and their interactions need to be studied.

1.4 How to measure individual wellbeing?

1.4.1 Basics of measurement

In a detailed discussion of the foundations of measurement, Hand (2004, p. 23) pointed out that we view the world through the spectacles of measurement. Modern technology, modern civilisation, is built on quantitative characterisation of the objects and attributes about us. It is such quantitative characterisations which ensure that our bridges do not fall down, our food distribution networks function, transport systems get us to work and our television, telephone and other communications systems do what they were intended to. It is such quantitative characterisation which enables us to measure how well we are doing, and to set targets and assess our achievements against them. But such quantitative characterisations have not always come easily.

We noted in Section 1.2 that all measurement can be thought of as a mix of two aspects, the representational and the pragmatic. Representational measurement is concerned with the mapping from the system of objects and the relationships between them to a corresponding numerical representation. A simple example would be placing two rocks on the two pans of a weighing scale and assigning a larger number to the rock which forced its pan down. By this means we could assign a set of numbers to the rocks which represented their weight, in the sense that larger numbers were associated with heavier rocks. Then, going further, we could place two rocks together on one pan, and find a third rock which just balanced this pair of rocks. It is then possible to assign numbers to the rocks such that the sum of the two numbers assigned to the pair on the same pan equalled the number assigned to the other rock. These numbers have the properties of our usual measure of weight. Indeed, we could take a particular rock and let the number assigned to it be unity, and then all the other rocks would have weight defined in terms of it. In standard usage, such a procedure determines the unit of measurement, with 'unity' corresponding to 1 oz or 1 kg and so on.

Representational measurement, as illustrated by our rocks example, is all very well, but clearly something more is needed when we attempt to measure wellbeing. This 'something else' is the pragmatic aspect of measurement. The representational aspect of measurement constrains the numbers we can use to represent the magnitudes of attributes of objects by requiring the relationships between these numbers to be the same as the relationships between the objects (e.g. the number assigned to represent the weight of one rock is greater than that assigned to represent the weight of another if the first rock 'is heavier than' than the second, i.e. the first rock tips the scale). The pragmatic aspect of measurement also places constraints on the numbers we can use, but these constraints are chosen for reasons of convenience, practicality or to encapsulate our intended meaning of the attribute in question. In measuring wellbeing, for example, as the preceding sections will have made abundantly clear, different researchers regard different properties as being part of the overall concept of wellbeing. To assign a number to the wellbeing of an individual, we first have to decide which of these properties we regard as relevant, and then we have to decide how to combine these properties. As Hand (2004, p. 13) puts it: '[Pragmatic constraints] *crystallize* the numerical assignment, so that we know exactly what we are talking about.'

There are a number of consequences of this. Different researchers may well have different views on what is relevant or important. The various sets of relevant characteristics listed in the previous section illustrate some different perspectives. It is reassuring, however, that the various sets have much in common. The main difference appears to be the scope of the lists.

Different researchers will also have different views on how to combine the various characteristics which are parts of wellbeing. Should one use a weighted sum – a sum of the values of the characteristics, with each one weighted according to its perceived importance? If so, what weights are appropriate? Is some more elaborate combination method appropriate? Since different choices can be made, different researchers will have different definitions, as well as different ways of measuring wellbeing. This is fine: unlike 'weight', 'wellbeing' is a complex concept, and we should expect different definitions, meanings and measurement procedures to emerge as we dig down into exactly what we mean by it and as we tap into and emphasise different aspects of it. But we should not lose sight of the concomitant obligation that when one talks of wellbeing one must be explicit about precisely what definition one is using.

Domains such as wellbeing, which have a very substantial pragmatic component, present tougher measurement challenges than largely representational domains, such as weight. But the differences are perhaps not so great as some believe. As Quetelet put it in the mid-nineteenth century: 'Although we are here in a new field, where facts cannot be estimated mechanically, as in the physical sciences, the difference, nevertheless, is not to be held so great as it may appear at first sight. Even the physical sciences sometimes rest on facts which are not identically the same, as deaths and births should be, and which may lead to appreciations and conclusions more or less great. With the use even of an instrument, when one wishes to discover a temperature, a magnetic declination or the force and direction of a wind, does one really find the quantities which are sought? When one measures an individual, is the real height positively discovered? Errors, greater or lesser, may be committed, and observation alone can recognise the limits within which they range. Has the consideration of the average life of man been rejected, because that average rests upon numbers which vary, without doubt, within limits as extended as can be conceived?' (Quetelet, 1842, p. ix).

And, one might add, measurement error aside, individual heights are not well defined: they change over the course of a day, under the inexorable force of gravity squashing the cartilages and organs of the human body.

In all measurement, whether representational or pragmatic, one must be sure one is measuring the right thing. Now, 'right', out of context, is meaningless. This means that, before we can decide how to measure an attribute, we need to decide what is the purpose of our measurement. And then we need to match our measurement procedure to that purpose. Someone on a slimming diet basing their success on their height or the length of their hair would not make much progress.

Measurement procedures which are heavily pragmatic are typically found in very complex areas, and usually the social or behavioural sciences, but not solely: Baggott (2004, p. 320) gives an example from quantum mechanics. He says 'Remember that we have no way of knowing the "actual" signs of the phase factors because this is information that is not revealed in experiments. However, we can adopt a phase convention which, if we stick to it rigorously, will always give results that are both internally consistent and consistent with experiment'. But in the social and behavioural areas in particular, one needs to tread carefully to ensure that one really is measuring what one wants to measure, and also that the measuring procedures have desirable characteristics. The next two subsections look at just some of these issues.

To conclude this subsection, we should comment on gold standards in measurement, and the role they might play in measuring wellbeing. A gold standard is the true value of the attribute being studied, and sometimes situations arise where it is possible or perhaps possible in principle, to measure this true value directly. For example, in studies of osteoporosis, we might regard radiography as leading to the true extent of bone deterioration (the gold standard). But for reasons of cost we might prefer to adopt a questionnaire based on items concerning age, diet, exercise, etc., which we know to give scores correlated with the radiographic result, and which are much quicker and cheaper to determine, as well as being non-invasive. Or, as another example, we might wish to know something about the true state of a disease so that we can treat it, in situations where the true state can be discovered only at a post mortem examination. Here we might again try to develop a screening instrument which is highly correlated with the true disease status.

Unfortunately wellbeing admits of no such gold standard. Indeed, as we have seen, the very complexity of the concept means that there are multiple interpretations of what is meant by 'wellbeing'. Even such straightforward approaches as simply asking people questions like 'are you satisfied with your life?' arguably touch on only certain aspects of wellbeing.

However, even though no gold standard exists, there are indicators which one might expect to be highly correlated with wellbeing. When this is the case, advantage can be taken of this expected relationship to test and refine potential definitions. For example, at the aggregate level, where we are trying to determine the wellbeing of a population, we might reasonably expect suicide rate to be negatively correlated with wellbeing. If it is not, it suggests either that our measure of wellbeing is missing something or that we have misunderstood something about the potential relationship between these two variables. Either way it will lead to exploration and improvement of understanding and the measures.

1.4.2 What is measured matters

In Section 1.2, we noted that one informal definition of wellbeing was 'what matters'. But this can be inverted, so that what is measured may be taken to be what matters – purely because the solid numbers assigned to it give it a spurious sense of accuracy, validity and reality. Many authors have cautioned against this mistake.

For example, there is the McNamara Fallacy. Charles Handy (Handy, 1995) describes it thus: 'The first step is to measure what can be easily measured. This is OK as far as it goes. The second step is to disregard that which cannot be easily measured or to give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what cannot be measured easily really is not important. This is blindness. The fourth step is to say that what cannot easily be measured really does not exist. This is suicide'. Handy summarises this as 'What does not get counted does not count'.

Chambers (1997) put a similar message more poetically:

Economists have come to feel
What can't be measured isn't real
The truth is always an amount
Count numbers, only numbers count.

Enrico Giovannini, then the OECD Chief Statistician, has put a more positive spin on this: 'by measuring progress, we foster progress'. While sympathetic to the intent, we feel that may be a claim too far, at least if it is interpreted as requiring measurement and only acting on the results of that measurement.

Goodhart's law is a related phenomenon – see Hand (2007). This says that 'any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes' (Goodhart, 1984). It basically means that an indicator adopted as a measure of effectiveness of policy gradually loses its usefulness as people and organisations work towards the target. For example, on the subject of health indicators, McDowell and Newell (1996, p. 11) say 'Social reforms are based on the information that is available to us, so the selection and publication of indicators of health both reflect and guide social and political goals. Hence the very choice of indicators tends to affect the health of the population; publication of an indicator focuses attention on that problem, such as infant mortality, and the resulting interventions (if successful) will tend to reduce the prevalence of the problem, in turn reducing the value of the indicator as a marker of current health problems'. The measures themselves contain the seeds of their own irrelevance.

Goodhart's law can be taken even further in the variant sometimes known as Campbell's law: 'The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor'. In some cases this can even take the form of explicit manipulation of the data. For example, *The Telegraph* of 28th February 2003 reported that 'the Commission for Health Improvement rebuked bosses at West Yorkshire ambulance service for reporting better than average performance … According to the watchdog, West Yorkshire Metropolitan Ambulance Service NHS Trust (WYMAS) downgraded some category A calls, if the ambulance crew arrived at the scene and decided the call was not, in hindsight, serious enough to warrant a category A response … The watchdog also found a substantial time lag between the time a call was received and the time the trust started the clock to time its response' (Telegraph, 2003).

This example also illustrates one of the intrinsic difficulties with social and behavioural measurement systems used for social policy development, namely that such measurements may have (at least!) two rather distinct but typically intertwined purposes. The first purpose is simply to gauge performance. Areas of strength and weakness can be identified, so that effort and resources can be focused on those domains where improvement is necessary. This

purpose is essentially internal. But the second purpose is to *report* performance – perhaps to compare with a stated target, or to provide input to a league table. This usage is essentially external.

A straightforward example of this dichotomy is that of school tests. For the first purpose, a test will show how individual students are doing, where difficulties are arising with the taught material, and on whom the teacher needs to focus attention. For the second purpose, a test will show what level the students have reached, so that, for example, employers know that the graduates have reached a requisite level of knowledge. From the perspective of a student, it would be advantageous if they admitted to doing poorly on those aspects of the test where they were weak when the test was being used for the first purpose, but concealed this when it was being used for the second.

One way to reduce the potential for 'gaming', or working towards the target, is to use multiple indicators, in what is sometimes called a 'dashboard'. The different components of wellbeing discussed in Section 1.3 could all be reported separately, so that a performance profile results. This clearly gives a better picture of what is going on, but at the cost of increased complexity. Furthermore, in many situations a single performance measure will be needed (e.g. the media, especially with their insatiable appetite for league tables, will doubtless require one). We shall return to this point below.

A second strategy for tackling such problems is to change the target indicator in unpredictable ways periodically. This may be a good strategy for the first purpose above, but it creates difficulties for the second purpose, because consistent time series cannot be constructed. It has been suggested that deliberately leaving some flexibility or ambiguity in definitions – of wellbeing, for example – can serve the same purpose of making gaming more difficult. But it could equally have the opposite effect as different users refine the measures to their particular advantage.

There are many further complicating issues. As we have seen, most researchers regard health as an intrinsic part of wellbeing. But health care provision is not static and, in particular, health care technology progresses over time. This could have various consequences which could work in either direction: the overall health of a population might improve over time or the expectations within a population might exceed practical results, inducing dissatisfaction. Certainly, more medical investigations and procedures become possible as things progress, but at an increased cost. This is one of the dilemmas with which health care systems, worldwide, are currently battling. Likewise, medical screening can serve to identify potentially fatal diseases at an earlier stage (so increasing the apparent survival rate merely because the disease has been identified earlier), but this can come at a very substantial cost in terms of large numbers of medical procedures carried out on the false positives or those with benign conditions (and all medical procedures carry some risk).

It is well known that blood pressure can become elevated by the tension of having a nurse or doctor take one's blood pressure. This illustrates a basic problem: How can we measure something if the very act of measurement distorts the thing we want to measure? The same effect is manifest elsewhere. In particular, the subjects of a study can respond positively to the mere fact of being studied (the attention they receive, for example), so that their scores are better than they should be. This phenomenon, called the Hawthorne Effect, is described in detail in Hand (2007). Wellbeing studies are potentially particularly vulnerable to this sort of effect: Paul Dolan has reminded us that Schkade and Kahneman's (1998) observation, that 'Nothing in life is quite as important as you think it is while you are thinking about it', clearly applies in the measurement of subjective wellbeing (Dolan and Powdthavee, 2012).

1.5 Properties of measurements

We referred to 'measurement error' in Section 1.4. This represents the discrepancy between the observed measurement and the underlying 'true' value. Clearly this is a difficult concept when, as with measuring wellbeing, the very concept of an underlying 'true' value is troublesome. More generally, we will be interested in a range of aspects of a measurement, including such things as the following.

- *Data availability*: Can the elements which are combined to yield a wellbeing score themselves be measured? We saw, for example, that sustainability was regarded as an important pillar of wellbeing by the OECD in its *How's Life* study, but that problems of data availability meant that it was not discussed in detail in that report.
- *Timeliness*: If a wellbeing measure is to be used to monitor change over time, for example, in an exploration of the effect of policy changes, then clearly it is important to be able to collect the data and derive the wellbeing score on a timescale which makes it possible to detect the relevant changes.
- *Sensitivity*: It is critical that the wellbeing score should be sufficiently sensitive to changes of a size which matter in its component factors. In contrast, the fluctuations arising from changes which do not matter should have a limited impact in terms of changes to the wellbeing score. At the two opposite extremes we have a score which fluctuates dramatically from day to day and a score which barely changes over time. Something in between is needed.
- *Data quality*: There are many aspects to this. Examples include floor and ceiling effects (in which minimum and maximum values to the measurement scale lead to bunching near those values), incomplete data (if a wellbeing questionnaire includes questions which are likely to elicit many missing values, such as 'what is your spouse's income?', then it will be problematic) and distortions arising from poorly worded questions. At an aggregate level, an important aspect of data quality is sample selection bias.

Such issues are related to the characteristics of measurement scales known as *validity* and *reliability*.

1.5.1 Validity

Validity indicates the extent to which the test measures what you want it to measure: How close the measured values are to the underlying 'truth'. So validity is about 'meaning'.

The concept of validity has been the subject of a very considerable amount of attention, especially in the behavioural and social sciences. This has led to the exploration of different kinds of validity, including the following.

- *Content validity*: Does the measuring instrument properly capture the entirety of the concept in question. For example, if the instrument consists of a number of questions, the responses to which are combined to yield an overall score, does the collection of questions cover the full range of aspects of the concept?
- *Criterion validity*: This is an indication of how well a measure correlates with some gold standard. It will be used when a proposed new measure is quicker, cheaper, or more timely (e.g. a diagnostic test vs. a posthumous dissection) than the existing gold

standard measure, and is only relevant when there *is* a gold standard, so that it is seldom relevant in measuring wellbeing. However, even if a gold standard is not available, there may be other external characteristics with which one might expect it to be correlated, such as suicide rate, in the context of national wellbeing.

– *Construct validity*: When there is no gold standard, one strategy is to relate multiple measures of the same phenomenon, to see if they are correlated and indeed if these correlations fit in with theoretical expectations.
– *Face validity*: If a test looks as if it should measure the attribute one is aiming to measure, then it is said to possess face validity; one might say that 'at face value' it measures what one wants to measure. Looking as if it measures what one wants to measure is not the same as being *proved* to measure what one wants to measure, of course, so in a sense face validity is a complement to content validity.

1.5.2 Reliability

Reliability measures how close the measured values are to each other – how consistent are the results. If one repeatedly measures the same thing when there has been no opportunity for the underlying characteristic to change, how much variability is there in the scores? This is hypothetical: If one actually measured the same thing in quick succession, people are likely to remember their previous response, so that an overestimate of the reliability of the measurement procedure would result. Some more sophisticated way to measure reliability needs to be devised. Various strategies have been explored to tackle this, including such things as parallel forms approaches (in which different questions aim at the same concept).

More details of the concepts of validity and reliability can be found in Hand (2004).

1.6 Objective or subjective?

We mentioned several objective aspects of wellbeing in Section 1.3, including education, ability to perform daily activities, political voice, security, extent of social network and environmental conditions. But one might argue that the *subjective* aspects are the key to measuring wellbeing. After all, we want to know how satisfied and content people are, and who better to judge someone's state than the person oneself? This raises rather subtle questions. For example, as we mentioned earlier, in the United Kingdom (and indeed elsewhere) the crime rate has dropped in recent years, but fear about it has gone up. It is a subtle question whether objective or subjective measures, as a component of wellbeing, are more appropriate in this context.

At their most basic level, subjective measures can be based on a single question (e.g. 'Taking all things together, would you say you are: very happy, quite happy, not very happy or not at all happy?', van Hoorn, 2007). However, both experience and psychological measurement theory shows that better estimates can be found by identifying and measuring different component aspects, and also by combining multiple measures of the same thing.

Regarding the first of these points, we have already noted that subjective wellbeing may be regarded as being composed of a cognitive component of life evaluation, and an affective component, including both positive and negative aspects, capturing feelings at the time. Some researchers also include an explicitly eudemonic component describing sense of engagement or purpose. Given these various aspects, we are then back to the issue of how to combine them to yield an overall measure, discussed further below.

Subjective wellbeing, even more than objective measures, is susceptible to difficulties in making cross-national and cross-cultural comparisons. The term 'cultural bias' has sometimes been used to describe the differences between measured values which are the consequence of measurement instruments rather than underlying reality. Inadequacy of measuring instrument is one thing, but another is simply the way in which people *report* their personal experiences, regardless of what those experiences are. Despite these difficulties, there are various surveys which include subjective wellbeing measures and attempt to make them comparable across different countries, such as the *Gallup World Poll* and the *World Values Survey*.

Perhaps the defining feature of subjective measures of wellbeing is that they are, well, *subjective*. There is no ground truth against which we can compare the subjective rating. Nonetheless, the fact that subjective measures do reflect some sort of reality is indicated by correlations between such measures and objective measures or outcomes: for example, one might expect that people who rate themselves as dissatisfied with their jobs tend to leave them at a greater rate than others who claim to be satisfied. There are also various feedback mechanisms. An optimist, feeling positive about their chances of winning a lottery, might buy a ticket. A pessimist, 'knowing' they will not win, will not buy a ticket. Hardly surprising, then, that a study might show a correlation between an optimism/pessimism scale and the probability of winning a lottery. Michaelson *et al.* (2012, p. 6) have something to say about this: 'Whilst well-being covers more than happy feelings, recent research suggests that positive feelings like happiness can actually lead to better well-being overall. This is because positive feelings broaden people's potential responses to challenging situations and build their personal resources and capabilities. We should think of feeling happy not only as a goal in itself, but also as a way of increasing people's potential for doing well'.

Broadly, speaking, measurement of subjective wellbeing has been approached in two different ways. One way seeks immediate reports as things happen in normal surroundings (and not, for example, in psychology laboratories); the advantage of this is that it has pragmatic validity – it is how people evaluate their own experience as it occurs (see, for example, Scollon *et al.*, 2003), and is not subject to retrospective perceptual and memory distortion. The other approach is a retrospective one, in which respondents keep a diary of the day before, evaluating how they felt at each hour or event during the day (e.g. Kahneman *et al.*, 2004). All retrospective methods are at risk of memory bias.

It is perhaps worth commenting that subjective measures of wellbeing are potentially intrusive, and as with all intrusive survey questions, responses may be susceptible to various kinds of distortion. This is tied in with ethical issues. It is incumbent on statistical agencies to have a clear purpose for collecting data, not only because of the public cost of collecting the data, but also because of the intrusiveness involved.

1.7 Combining multiple aspects

As we have stressed, wellbeing is not a simple representational measurement concept, like length or weight, which readily admits to being mapped to a unidimensional numeric variable, but is a confluence of various aspects. This means that, rather than trying to represent wellbeing by a single indicator or index, it might be better to use a multivariate profile – a 'dashboard', as we said above. Such profiles cannot have too many component indicators or they are useless: one would not know where to start if presented with a profile of some hundreds of measures. (An obvious trap to avoid, one might think, but its obviousness has not prevented various public bodies falling into this trap when devising profiles of key performance indicators.)

One strategy for easing things is to use a hierarchy of indicators, with a few (or even just one) headline indicators, and a range of secondary indicators associated with each of the headline indicators.

All this is very well, but it is probably inevitable that at some stage the various components of the profile will be combined to yield a single measure. With single scalar measures it is much more straightforward to make comparisons – over time, culture, policies, etc. It is important that this summarisation be done in a principled and informed manner by the designers of the measurement scale than in a simple *ad hoc* approach by people not versed in the subtleties of measurement technology.

There are, broadly speaking, two philosophically different perspectives on how to combine multiple indicators to give a single overall score. These two perspectives have gone under various names. For example, Feinstein (1987) calls them *psychometric* and *clinimetric*.

Fayers and Hand (2000, p. 241) explain the difference between the two perspectives: 'The objective of the psychometric approaches … might be characterised as *attempting to measure a single attribute using multiple items*. In contrast, clinimetric methods *attempt to summarise multiple attributes with a single index'*. (our italics)

A prime illustration of the psychometric perspective is factor analysis. This is perhaps hardly surprising, since factor analysis was initially developed by and for the psychological and psychometric communities. Factor analysis postulates the existence of an underlying 'factor' or characteristic, which cannot be measured directly, hence also called a 'latent' or hidden factor. 'Intelligence' is a classic example. In fact, factor analysis typically generalises this structure to postulate the existence of several underlying factors which cannot be measured directly, but for the purposes of this exposition it is sufficient to assume just one.

Observable variables (in the case of intelligence these might be scores on different tests – numerical reasoning, verbal reasoning, etc.) are correlated with the unobservable latent factor. In fact, in the standard factor analysis model, a subject's score on any one of these observable variables is the sum of that subject's value of the underlying latent factor and a 'specific' part due to the particular variable being observed. Covariances between the observable variables are explained by virtue of the fact that they have the latent variable in common. Statistical methods, applied to the covariance matrix between the observed variables, allow one to separate out the contributions from the latent variable and the specific variables. In particular, by such means one can estimate a subject's score on the latent variable, even though it cannot be observed directly.

A key thing about factor analysis is that it is based on a postulated *model*. It says 'the observed covariance matrix arises because there is this underlying structure'. And then seeks to estimate the details of the structure from the observed matrix. This is clearly very much in the psychometric mould, as described above; it seeks to measure a *single* unobservable attribute using the observed values of *multiple* items which are related to the unobservable attribute by the assumed model structure. (Of course, if the assumed model structure is wrong, then things can go awry, but a discussion of that is beyond the scope of this book.)

This situation appears rather different from the case of wellbeing and quality of life. Not everyone agrees that there is some underlying characteristic – wellbeing – with observed attributes (such as subjective contentment, adequate income, strong social connections, low fear of crime, a warm place to live, etc.) arising largely because of this underlying factor. Such an assumption seems rather odd – the observable attributes clearly lack any intrinsic homogeneity. Rather it is thought that there are a number of *completely distinct* characteristics

which together can be taken as an indicator of wellbeing. This is the clinimetric perspective, in which the aim is 'to summarise multiple attributes with a single index'.

Of this aim of combining completely distinct attributes, Fayers and Hand (2002) say '… a single index loses the intrinsic differences between the attributes and it also sacrifices any possibility of allowing for an interaction … Instead the researchers aim their strategies at choosing and suitably emphasizing the most important attributes to be included in the index … In fact, of course, we are choosing how to *define* the concept being measured by our choice of variables and the way of combining them … Clinimetric scales are different from psychometric scales because, in the former, the items to be included are chosen according to what we want the scale to do, whereas in the latter they are chosen because they are thought to be related to an underlying concept which defies explicit measurement'.

The choice of characteristics to be included in a clinimetric index, along with the method by which they are to be combined, is very much a pragmatic one – in the technical sense defined in Section 1.4. It is a deliberate choice, to reflect one's aim, not a choice in any way determined by empirical relationships (such as an observed covariance matrix). A classic example of a clinimetric scale is the Apgar score for the state of health of newborn infants. This combines the very distinct items of body colour, heart rate, respiration, reflex response and muscle tone.

The word 'clinimetric' sits a little uncomfortably with the topic of this book, since wellbeing is not intended to have any clinical implications. We have therefore decided to stick to the word 'pragmatic' to describe the approach to constructing a measurement scale based on an explicit combination of the components regarded as relevant. This term is also being adopted by others (e.g. Paruolo *et al.*, 2013).

This abstract (clinimetric and pragmatic) structure is the approach adopted in almost all strategies for constructing wellbeing measures. But one can go further in various ways.

One extension is to combine psychometric and pragmatic approaches. This has been done in a rather elegant model described in the New Economics Foundation report *Measuring Our Progress* (NEF, 2011, p. 13). At its base is a fundamentally psychometric model, linking two aspects of wellbeing (good feelings, such as happiness, joy, contentment and satisfaction; and good functioning, such as autonomy, competence, safety and social contacts) to external conditions and personal resources in a causal model. But these two aspects then have to be combined to yield an overall measure of wellbeing, and the relative balance taken between good feelings and good functioning must be a pragmatic decision.

Another extension is to adopt a form of *conjoint* measurement. Conjoint measurement derives additive relationships between different components purely from ordinal scales. One can regard it as essentially seeking to balance the various contributory components so that a unit change in one component has an equal effect to a unit change in another. In essence it finds contours, in the space spanned by the constituent components, so that all profiles lying on the same contour have the same degree of the characteristic to be measured (wellbeing, in our case). This approach has occasionally been applied in the wellbeing context (e.g. EPICURUS, 2006) and in related areas (e.g. Bridges *et al.*, 2008).

The characteristics thought relevant to wellbeing can be thought of as of three types. First, there are characteristics which are regarded as fundamental aspects of wellbeing. Subjective states might mostly be regarded as of this type. Second there are those which are thought to affect or influence wellbeing. For example, one might regard income, education and skills as things which influence wellbeing: we might expect low income and lack of education and skills to lead to lower wellbeing. And third, there are those which are regarded as being affected by or influenced by wellbeing. A sense of competence, and hence satisfaction

with life, might be regarded as something which is a consequence of or an indicator of wellbeing.

It is not always clear whether a particular item is best regarded as an influence on, a consequence of or simply an aspect of wellbeing. Sometimes it is more than one, since feedback loops can occur. For example, a simple explanation of the fact that marital status is positively related to wellbeing amongst men might be that the support it provides is beneficial, so that marital status affects wellbeing. In the other hand, it is entirely possible that positive wellbeing induces a more optimistic and outgoing outlook which in turn is more likely to lead to strong personal relationships. In contrast, something like gender (women tend to have higher subjective wellbeing scores than men) can clearly work in only one direction: enhanced wellbeing cannot cause a gender change (we realise, as we write this that there are occasional exceptions, in which lack of wellbeing arising from gender dysphoria leads to a gender change operation; but these are very rare, and certainly occur in insufficient numbers to have a material effect on measures of national wellbeing!). And, likewise, the U-shaped relationship between age and wellbeing already mentioned can work in only one direction: enhanced wellbeing does not lead to one's age decreasing or increasing.

Quality of life models of this kind, involving an explicit recognition that there is more than one kind of characteristic (in fact two kinds – those which influence wellbeing and those which are consequences of wellbeing) are explored in detail in Fayers and Hand (2002). Characteristics which are influenced by wellbeing might well be summarised usefully by a factor form of model, but those which influence it can, at least in principle, be completely unrelated. The overall result can be thought of as a combination of, or a generalisation of, both factor models and regression models. The same sort of structure, applied in a different context, but one which has similar characteristics, is described in Hand and Crowder (2005). Sometimes such models are called *multiple indicator, multiple cause* (MIMIC) models.

1.8 What is national wellbeing?

Interpreting the wellbeing of a population in a very broad sense, one can discern a continuum of progress in developing measures, from simple mortality rates, via crude economic measures, to the current interpretation in terms of quality of life and beyond; again recall the World Health Organisation's definition of health as being 'a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity' (WHO, 1946). It is interesting to recall that when this definition was introduced it was criticised by some as immeasurable. It is a mark of how far measurement technology has come that this criticism no longer holds any force.

So far we have been discussing what is meant by individual wellbeing, and how to measure it. We have seen that this is difficult enough. But measuring, and even defining, national wellbeing opens up further layers of difficulty. National wellbeing is, at least certainly in part, some kind of aggregate or average of individual wellbeing, which raises the question of how that aggregation or averaging should be effected. This is discussed in the next section. However, as we noted earlier, national wellbeing is more than merely a summary of the individual wellbeing of the members of a population. There are also additional higher level issues which should be taken into account when measuring things at the national level. For example, use of non-renewable resources, while having little impact on individual short-term day-to-day experience, may be accumulating difficulties for the nation's future wellbeing. Perhaps this suggests that a concept of national wellbeing should avoid the short-term nature

which is an inevitable part of individual wellbeing by integrating or averaging over time. This explains why sustainability plays a part in many discussions.

Chapter 2 gives a history of measuring national wellbeing. The history is a long one, inevitably, if one regards a government's prime aim as being to promote national wellbeing. Familiar examples we give there include the United States Declaration of Independence of 1776, with its reference to 'the pursuit of happiness', and Sir John Sinclair's Statistical Account of Scotland of 1791, an 'inquiry into the state of a country for the purpose of ascertaining the quantum of happiness enjoyed by its inhabitants and the means of its future improvement'. That last is precisely the concern of the current interest in national wellbeing – it took only two centuries to get there.

Two more recent examples of measures which combine characteristics at different levels are the *Human Development Index* (see Chapter 2) and the *Happy Planet Index*. The definition of the *Human Development Index* has changed over time, but broadly speaking it is the geometric mean of three basic dimensions: (i) life expectancy at birth; (ii) education; (iii) and an income index. The *Happy Planet Index* (see link in Appendix), introduced by the New Economics Foundation in 2006, is defined as a combination of a subjective measure of life satisfaction, life expectancy at birth, and a measure of individual demand on the Earth's ecosystems (so-called 'ecological footprint').

Chapter 6 discusses in detail the different levels which need to be combined, covering national economic accounts (for which aggregate measures already exist), extension of the national accounts to cover such things as time spent on voluntary work and the distribution of income within and across households, specific social conditions such as environmental conditions and aggregates of individual wellbeing and life satisfaction. Note that since one use of a national wellbeing measure will be to compare different countries, measures such as GDP that depend on the size of the population will need to be standardised – giving, for example, GDP per capita.

1.9 And how to measure *national* wellbeing?

We see from the above that, to produce a measure of *national* wellbeing, we have to *aggregate* the measures of individual people into an overall measure for the population, and then combine the resulting aggregate(s) with other higher level populations aspects.

Methods of aggregating or averaging individual values, to yield an overall population summary, have long been used in other areas. One domain is that of health statistics, evaluated at the population level by aggregating from the individual level. Perhaps the best known example is that of mortality statistics (dating back at least as far as John Graunt's *Natural and Political Observations Made upon the Bills of Mortality*, in 1662 (Graunt, 1662)), but there are also population measures of morbidity. Indeed the discipline of *epidemiology* is defined as the study of patterns of diseases in populations, and there are many classic examples, including William Farr's collection of data relating to cholera in the nineteenth century. The strong link between public health and population wellbeing is illustrated by the following extract from *The Spirit Level* (Wilkinson and Pickett, 2010, p. 277).

> It is now almost universally accepted amongst scholars and practitioners of public health that the most important determinants of health are social and economic circumstances. Geoffrey Rose, who was one of the most highly influential and respected epidemiologists of the second half of the twentieth century, said,

> 'medicine and politics cannot and should not be kept apart'. Our growing understanding of human health and wellbeing are so deeply affected by social structure inevitably pushes science into politics.

In principle, aggregation is straightforward: individual level scores are determined using a survey, and then one simply calculates the proportion who respond positively to some question (e.g. 'overall, are you satisfied with your life nowadays?') or derives an appropriate average, such as an arithmetic mean (e.g. from the question 'overall, how happy did you feel yesterday, on a scale from 1 to 10?') or quantile values (e.g. what percentage of people had a score of three or lower on the previous question?).

Adolphe Quetelet, working in the nineteenth century, was a founder of the notion of extracting an 'average' or aggregate view of the individuals making up a population. He wrote: 'It is of primary importance to keep out of view man as he exists in an insulated, separate or in an individual state, and to regard him only as a fraction of the species. In thus setting aside his individual nature, we get quit of all which is accidental, and the individual peculiarities, which exercise scarcely any influence over the mass, become effaced of their own accord, allowing the observed to seize the general results' (Quetelet, 1842, p. 5). Quetelet thus stressed the need to strip away the individual-specific variation, to reveal the larger scale underlying property. It is this, in essence, which the measurement of *national* wellbeing seeks to do.

Quetelet was very much aware of the advantages which derived from aggregation. For example, (Quetelet, 1842, p. 6): 'It would appear, then, that moral phenomena, when observed on a great scale, are found to resemble physical phenomena; and we thus arrive, in inquiries of this kind, at the fundamental principle, that *the greater the number of individuals observed, the more do individual peculiarities, whether physical or moral, become effaced, and leave in a prominent point of view the general facts, by virtue of which society exists and is preserved'*. (His italics)

The first part of Quetelet's *A Treatise on Man and the Development of his Faculties* (Quetelet, 1842) is concerned with population growth, birth rates and death rates, and the second part is concerned with physical characteristics such as height, weight and strength. The third part is entitled 'Development of the moral and intellectual qualities of man', and is predominantly concerned with attributes such as memory, imagination, judgement, insanity, temperance, courage, genius, prudence and propensity for evil, including crime. He could so easily have extended this to wellbeing! Again he stresses the advantages of working with a mass of people (p. 73): 'It appears to me that it will always be impossible to estimate the absolute degree of courage, etc, of any one particular individual: for what must be adopted as unity? – shall we be able to observe this individual long enough, and with sufficient closeness, to have a record of all his actions, whereby to estimate the value of the courageous ones; and will these actions be numerous enough to deduce any satisfactory conclusion from them? Who will guarantee that the dispositions of this individual may not be altered during the course of the observations? When we operate on a great number of individuals, these difficulties almost entirely disappear, especially if we only want to determine the ratios, and not the absolute values'. If anywhere, it would be this part of Quetelet's book which dealt with wellbeing and happiness. But such a discussion does not figure there. The nearest Quetelet gets is a discussion of suicide rates.

There are subtleties in what we have called the 'aggregation' from the individual level to population level. One, in particular, has already been mentioned. This is the fact that there

seems to be a relationship between inequality and national wellbeing (Wilkinson and Pickett, 2010). Inequality, of wealth for example, is not a property of an individual, and a simple average of the wealth of the individuals in a population will not reveal such inequality. At the least one needs to look at measures of dispersion, and for properties such as wealth, which cannot take negative values and which tend to be very positively skewed; measures of skewness are also needed. For a skew distribution with a long positive tail, the majority of values will be less than the average, when 'average' is taken to be the arithmetic mean (so leading to the old joke about most people earning less than the average), with a very long tail of exceptionally large values being balanced by a large number of very small values. This would seem to be a natural driver of envy. Perceptions are probably aggravated by the constant media comment on large salaries and wealth, even to the extent that newspapers regularly report the value of the house owned by a story's protagonist. The fact is that wellbeing is influenced by relative perceptions. The saying 'It is not enough to succeed; it is also necessary that others must fail' or variants of it, has been variously attributed to a wide range of authors.

Inequality can manifest at various levels. One is the distributional level of wealth or income, but others appear in the form of gender or ethnic disparities leading to exclusion from education (and the enhancement of wellbeing that brings), jobs and even health care. The question of inequality and the various ways in which it can arise mean that measures of national wellbeing have to be sufficiently fine grained that subgroup analysis can be undertaken. A small simple random survey of an entire population may be insufficient to tease out important sources of poor wellbeing.

When surveys are being used to collect wellbeing data, and when the intention is to aggregate the wellbeing of the individual members of a population, it is perhaps unnecessary to note that great care must be taken over the data collection. The measured individuals must be representative of the population in some sense. Perhaps more than most areas, wellbeing is susceptible to distortions of various kinds arising from poor data collection strategies.

Our discussion on aggregating low-level values to yield a (component of) a national measure of wellbeing has focused on individuals as the elements. But some aspects aggregate other elements. Air quality, noise levels and even global temperature are examples. They can be measured at specific points, but precisely how to determine an overall 'average' requires some careful thought, not least because if one chose different points one could get very different values.

Examples of higher level population characteristics, not generally associated with values describing individuals, are high-level economic indicators, public value and sustainability. High-level economic indicators, of the kind currently measured, such as GDP and GNP, are based on the system of national accounts. Despite their imperfections as measures of wellbeing, they are important aspects of it. Compiling the national accounts involves much data collection including from individual businesses and households. The resulting aggregates, though, refer only to the economy as a whole (or to whole sectors of the economy). We do not talk about the GDP of an individual business. While we can measure household or individual debt, we also need to look at the national debt in terms of the amount that the government owes, from the securities and bonds it has issued or the amounts it has borrowed from international organisations for example. Paradoxically, much of a government's income has to be drawn from individuals and businesses through taxation, so that government debt is indirectly the debt of individuals and businesses.

When looking at public value, the essence is that something has value to the public at large. A public park, for example, enhances wellbeing, but it is part of the national fabric

and cannot be attributed to individuals. Public value is deemed to be intrinsic to things like parks: they are valued by the public whether or not they use them. However, we will see in Chapter 6 that cost-benefit analyses of public policy decisions might draw on how they affect individual wellbeing.

Sustainability, in a number of arenas, tells us whether our present way of life can be maintained. It is therefore necessarily a high-level assessment, but it is also something that can only be fully assessed in retrospect. We will look more closely at sustainable development in Chapter 2, but to summarise the point here, sustainable development is often defined as development that 'meets the needs of the present without compromising the ability of future generations to meet their needs' (World Commission on Environment and Development, 1987). Such foresight is asking a lot of current measurement systems.

There have been many attempts to quantify sustainability: Chapter 3, of Stiglitz *et al*. (2010) includes a review, and as they say 'This abundance of measures is a serious drawback insofar as different synthetic indicators convey widely divergent messages. This leads to a great deal of confusion among statisticians and policymakers'. They argue that sustainability is complementary to current wellbeing or economic performance and 'must be examined separately', going so far as to say that attempting to combine these two aspects into a single indicator leads to confusion. As a result, they recommend a dashboard of indicators.

1.10 Structure of the book

We began this chapter by presenting three fundamental questions: what is national wellbeing, why should it be measured, and how should it be measured? Chapter 1 has thus set the scene for what follows. Since national wellbeing is clearly built on individual wellbeing, we described what we mean by individual wellbeing, we looked briefly at its component aspects and then outlined notions of how to measure it, including some basic measurement theory. This introductory discussion included examination of objective and subjective aspects, and also how the various different aspects should be combined to yield a single overall measure (given the inevitable need for a single measure even where a dashboard containing several measures is presented). We also argue that the wellbeing of individuals is but one aspect of *national* wellbeing. A nation, as a whole, must address issues at a population level, consider the future, and generally take note of aspects which may not impact individuals, at least at the time of measurement.

Chapter 2 sets our development in context by giving a history of the concept of national wellbeing and its measurement. The concept certainly has a long history, with the economic aspects of national wellbeing going back at least as far as the 1600s. But, as we repeatedly stress in this book, echoing one of the key drivers behind the global initiatives to measure national wellbeing, economics is but one aspect of wellbeing. The ideas covered in this chapter span philosophies of government, utilitarianism, the United Nation's System of National Accounts, quality of life, the rise of psychological measures of wellbeing, the capabilities approach, issues of sustainability and others.

From the history, we move in Chapter 3 to examine more recent developments. The report by Stiglitz, Sen and Fitoussi, produced by a commission established by the then president of France, Nicholas Sarkozy, has been seminal, as have the detailed developments by the OECD. We also discuss concurrent initiatives, especially to replace the Millennium Development Goals, in which measures of wellbeing and progress are also core.

Because individual wellbeing is so central to national wellbeing (some would say individual wellbeing is the entirety of national wellbeing). Chapter 4 is devoted to how individual wellbeing can be measured. The very notion of measuring wellbeing sometimes encounters lay scepticism. But this is in ignorance of profound developments which have been made in recent decades, both in terms of the theory of measurement, and in terms of its practical implementation. The fact is that degrees of subjective feeling can be quantified, and indeed, can be quantified accurately. These measures build on a huge amount of work on measuring psychophysical responses, opinions and attitudes, as well as subjective phenomena such as pain and depression.

Chapter 5 steps up a level, from the individual to the nation, examining what measures of national wellbeing and progress might be used for: it is in their very nature that different measures, and different measurement procedures, will be needed if the results are to be used for different purposes, just as a price index for macroeconomic purposes may not be suitable for use as a measure of inflation pressure on households.

Chapter 6 then looks at the details of how to construct a suitable measure of national wellbeing. The economic aspects (with the ubiquitous GDP leading the charge) are but one aspect of national wellbeing, and they need extending – in order to capture economic aspects properly, but, more importantly, more widely to capture non-economic social, cultural and environmental aspects. The chapter also examines ideas, methods and tools from around the world.

Measuring national wellbeing is of global interest, with a large number of initiatives being undertaken. To complement the broad brush approach elsewhere in the book, and to give the flavour of the depth of investigations being carried out, Chapter 7 provides a case study: that of the United Kingdom. The chapter describes the historical development of the measuring national wellbeing initiative in the United Kingdom, and outlines where it has got to.

Finally, in Chapter 8, we look back at where we have got to, and look ahead to the future. We note the potential that new data capture and analytic technologies have for the measurement of wellbeing. And we note that, to go beyond economic measures of wellbeing and progress will not be straightforward; we describe it as requiring a paradigm shift. Although the signs are promising, it is clear that, if all this effort is to do any good, the measures must start to be used. Words are all very well, but the measures must be integrated into government policy decisions, with businesses and households also making decisions that similarly go beyond the purely economic.

We conclude the book with an appendix of sources. Perhaps one of the most striking things about the entire area is the amount of work being undertaken. It really is a global movement. And it is a dynamic movement: we were aware that new developments were being made even as we were writing. Despite that, it is clear that ideas are converging: we expect the material covered in this volume to continue to hold true, as national wellbeing measures are crystallised.

References

Baggott J. (2004). *Beyond Measure: Modern Physics, Philosophy and the Meaning of Quantum Theory*. Oxford University Press, Oxford.

BBC. (2012). Kabul Bank: Afghan Politicians 'interfered with enquiry'. http://www.bbc.co.uk/news/world-asia-20523217 (accessed 15 March 2013).

Belgian Federal Planning Bureau. (2005). *Transdisciplinarity and the Governance of Sustainable Development*. Belgian Federal Planning Bureau, Task Force Sustainable Development.

Bentham J. (2009). *An Introduction to the Principles of Morals and Legislation*. Dover Publications Inc.

Brickman P. and Campbell D.T. (1971). Hedonic relativism and planning the good society. In: M.H. Apley (ed.) *Adaptation 0-Level Theory: A Symposium*. Academic Press, New York. pp. 287–302.

Bridges J., Haber B., Marshall D. *et al.* (2008). A Checklist for Conjoint Analysis Applications in Health: Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. http://www.ispor.org/taskforces/documents/A_CHECKLIST_FOR_CONJOINT_ANALYSIS_APPLICATIONS_IN_HEALTH.pdf (accessed 25 April 2014).

Chambers R. (1997). *Whose Reality Counts? Putting the First Last*. Intermediate Technology, London.

Christmann K. and Rogerson M. (2004). *Crime, Fear of Crime and Quality of Life Identifying and Responding to Problems*. Research Report 25, New Deal for Communities: The National Evaluation, Sheffield Hallam University.

Christopher J.C. (1999). Situating psychological wellbeing: exploring the cultural roots of its theory and research. *Journal of Counseling and Development*, **77**, 141–152.

CSLS. (2011). Happiness as a Goal for Public Policy: Ready for Primetime? A Synthesis of the CSLS-ICP Conference on Happiness and Public Policy, CSLS Research Note 2011-1. http://www.csls.ca/notes/Note2011-1.pdf (accessed 4 July 2013).

DeNeve K.M. and Cooper H. (1998). The happy personality: a meta-analysis of personality traits and subjective well-being. *Psychological Bulletin*, **124**, 197–229.

Dickens C. (1850). The Personal History, Adventures, Experience and Observation of David Copperfield the Younger of Blunderstone Rookery (Which He Never Meant to Publish on Any Account).

Dolan P. and Powdthavee N. (2012). Thinking about it: a note on attention and well-being losses from unemployment. *Applied Economics Letters*, **19**(4), 325–332. doi:10.1080/13504851.2011.5770008

Dunn G., Sham P.C. and Hand D.J. (1993). Statistics and the nature of depression. *Journal of the Royal Statistical Society, Series A*, **156**, 63–87. (Reprinted in Psychological Medicine, **23**, 871–889).

Easterlin R.A. (2003). Building a better theory of well-being. Discussion Paper No. 742, Institute for the Study of Labour, Bonn.

Edgeworth F.Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. C. Kegan Paul and Co, London.

EPICURUS. (2006). Societal and Economic Effects on Quality of Life and Well-being: Preference Identification and Priority Setting in Response to Changes in Labour Market Status: Final Report, HPSE-CT-2002-00143. European Commission, Directorate General for Research, Brussels.

Fayers P.M. and Hand D.J. (2002). Causal variables, indicator variables, and measurement scales, with discussion. *Journal of the Royal Statistical Society, Series A*, **165**, 233–261.

Fayers P.M. and Machin D. (2000). *Quality of Life: Assessment, Analysis and Interpretation*. Wiley, Chichester.

Feinstein A.R. (1987). *Clinimetrics*. Yale University Press, New Haven, Connecticut.

Felce D. and Perry J. (1995). Quality of life: its definition and measurement. *Research in Developmental Disabilities*, **16**, 51–74.

Flynn J.R. (2012). *Are We Getting Smarter: Rising IQ in The Twenty-First Century*. Cambridge University Press, Cambridge.

Goodhart, C.A.E. (1984). *Monetary Theory and Practice*. Macmillan, Basingstoke.

Graunt. (1662). Natural and political OBSERVATIONS mentioned in a following INDEX, and made upon the Bills of Mortality. http://www.edstephan.org/Graunt/bills.html (accessed 21 December 2013).

Hand D.J. (2004). *Measurement Theory and Practice: The World Through Quantification*. Wiley, Chichester.

Hand D.J. (2007). *Information Generation: How Data Rule Our World*. One World Publications.

Hand D.J. and Crowder M.J. (2005). Measuring customer quality in retail banking. *Statistical Modelling*, **5**, 145–158.

Handy C. (1995). *The Empty Raincoat: Making Sense of the Future*. Arrow, London.

Harrington R. (1804). *The Death-Warrant of the French Theory of Chemistry*. Longman, London.

Kahneman D., Diener W. and Schwarz N. (eds) (2003) *Well-Being: The Foundations of Hedonic Psychology*. Russell-Sage Foundation, New York.

Kahneman D., Krueger A.B., Schkade D., Schwarz N. and Stone A. (2004). Toward national well-being accounts. *American Economic Review*, Papers and Proceedings, **94**, 429–434.

McAllister F. (2005). *Wellbeing Concepts and Challenges*. Sustainable Development Research Network.

McDowell I. and Newell C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires* (2nd ed.). Oxford University Press, Oxford.

Michaelson J., Mahony S. and Schifferes J. (2012). *Measuring Well-being: A Guide for Practitioners*. New Economics Foundation, London.

NEF. (2011). Measuring Our Progress: The Power of Well-Being. http://dnwssx4l7gl7s.cloudfront.net/nefoundation/default/page/-/files/Measuring_our_Progress.pdf (accessed 2 August 2013).

OECD. (2011a). *How's Life: Measuring Well-Being*. OECD Publishing.

OECD. (2011b). *Compendium of OECD Well-Being Indicators*. OECD Publishing.

OECD. (2013). *OECD Guidelines on Measuring Subjective Well-Being*. OECD Publishing.

Paruolo P., Saisana M. and Saltelli A. (2013). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society, Series A*, **176**, 609–634.

Quetelet M.A. (1842). *A Treatise on Man and the Development of His Faculties*. Burt Franklin, New York.

Ryan R.M. and Deci E.L. (2001). On happiness and human potentials: a review of research on hedonic and eudemonic well-being. *Annual Review of Psychology*, **52**, 141–166.

Santor D.A., Gregus M. and Welch A. (2006). Eight decades of measurement in depression. *Measurement*, **4**, 135–155.

Schkade D. and Kahneman D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, **9**, 340–346.

Self A., Thomas J. and Randall C. (2012). *Measuring National Well-being: Life in the UK, 2012*. Office for National Statistics.

Stiglitz J.E., Sen S. and Fitoussi J-P. (2010). *Mismeasuring Our Lives: Why GDP doesn't Add Up*. The New Press, New York.

Strange T. and Bayley A. (2008). *Sustainable Development: Linking Economy, Society, Environment*. OECD.

Telegraph (2003). Ambulance service 'lied over response rates'. http://www.telegraph.co.uk/news/1423338/Ambulance-service-lied-over-response-rates.html (accessed 26 December 2012).

van Hoorn A. (2007). A short introduction to subjective well-being: its measurement, correlates and policy uses. International Conference on Is Happiness Measurable and What Do Those Measures Mean for Policy? University of Rome 'Tor Vergata', 2–3; April, 2007.

Verme P. (2011). Life satisfaction and income inequality. *Review of Income and Wealth*, **57**, 111–127.

Watson D., Clark L.A. and Tellegen A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS Scales. *Journal of Personality and Social Psychology*, **54**, 1063–1070.

WHO. (1946). Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22; June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

Wilkinson R. and Pickett K. (2010). *The Spirit Level: Why Equality is Better for Everyone*. Penguin Books, London.

World Commission on Environment and Development. (1987). *Our Common Future*, Oxford University Press, Oxford.

Scollon C.N., Kim-Prieto C., and Diener E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, **4**, 5–34.