Is Happiness All that Matters?

Marc Fleurbaey
Didier Blanchet

DOI:10.1093/acprof:oso/9780199767199.003.0006

Abstract and Keywords

This chapter analyzes the ethical underpinnings of subjective well-being indicators, their connections to psychological and philosophical background theories, their relation to economic concepts, and also the empirical strategies for possible improvements. It argues that happiness scores cannot be expected to reflect people’s values and goals in life directly, although they can provide useful information if properly handled. The chapter is organized as follows. Section 5.1 reviews the main arguments from the happiness literature in favor of shifting the focus from material to subjective aspects of well-being and relates them to the classical debate on welfarism. Section 5.2 proposes a theory of subjective well-being as it appears in questionnaires, in relation to the standard concepts of utility and preference orderings that economists are familiar with, but also taking account of perspectives and insights from psychology. Section 5.3 examines how the collection of data could be improved and how such data can be used for the empirical application of different conceptions of well-being.
Is Happiness All that Matters?

**Keywords:** subjective well-being indicators, happiness scores, welfarism, data collection


First, it has shifted priorities. The idea that social welfare can be equated with the growth of GDP has been much more powerfully shaken by the Easterlin paradox that derives from happiness surveys, than by theoretical objections against its underlying assumptions. Easterlin, (1974, 1995) observed that happiness scores in various countries seem quite flat in the long run in spite of a twofold or threefold increase in GDP over the last decades.\(^1\) Figures 5.1 and 5.2, borrowed from Clark, Fritjers, and Shields, (2008), illustrate this striking phenomenon for the United States and for five European countries. In addition, some happiness data suggest that employment, family, leisure, and pain are important aspects of well-being. Growth, therefore, can no longer be the only policy priority for economic advisors.

![Figure 5.1 Happiness and Real Income Per Capita in the United States, 1973–2004](source: Clark, Fritjers, and Shields, (2008, fig. 1))
The other change to the economic perspective on well-being is methodological. A strong positivist tradition initiated by Robbins, (1932) convinced economists for a very long time that choice behavior is the only source of relevant information about individual preferences and well-being. The concept of utility, which was central in the utilitarian tradition and was then considered almost as tangible a notion as money, became relegated to the status of a mathematical convenience making it easier to manipulate preference orderings. Interpersonal comparisons were viewed as a normative exercise not fit for a scientific discipline. What happiness studies have changed is that the practice of listening to people’s verbal utterances is now becoming respectable. It has also greatly contributed to reviving interest in the measurement of well-being.

Beyond the methodological shift, the resurrection of the concept of utility is impressive. There are now papers that estimate the curvature of the utility function directly from happiness data (Layard, Mayraz, and Nickell, 2008), and many authors argue that the similar behavior of various happiness and satisfaction scores suggests that they track the same underlying magnitude (e.g., Di Tella, McCulloch, and Oswald, 2003), while others argue that clever observation techniques make it possible to measure an “objective” form of subjective well-being (Kahneman, 1999). Bentham’s views make an unexpected comeback (Kahneman, Wakker, and Sarin, 1997; Layard, 2005).

These developments make it possible to imagine a new and simple solution to the quest for a measure of social welfare. Happiness scores are not very costly to elicit and are now available for most countries of the world. Why not use such

---

Figure 5.2 Life Satisfaction in Five European Countries, 1973–2004 Source: Clark, Frijters, and Shields, (2008, fig. 2).
data to compute Gross National Happiness, and take this statistic as the new compass of nations and policymakers? While this idea is cautiously defended by some authors (e.g., Diener, 2000; Dolan and White, 2007; Kahneman et al., 2004; Oswald, 1997), many refrain from doing so, and some firmly oppose it. The ongoing debate is partly reminiscent of a previous debate about subjective welfarism that took place in political philosophy about 30 years ago.

In this chapter we shall connect the two debates in order to better elucidate the normative underpinnings of a measure of social welfare centered on the notion of happiness. This chapter argues that happiness scores cannot be expected to reflect people’s values and goals in life directly, although they can provide useful information if properly handled. Whether or not one endorses the view that a good measure of social welfare should reflect people’s preferences, the implication is that happiness data should not be used without a careful processing that eliminates undesired biases and noises.

The first section reviews the main arguments from the happiness literature in favor of shifting the focus from material to subjective aspects of well-being and relates them to the classical debate on welfarism. The second section proposes a theory of subjective well-being as it appears in questionnaires, in relation to the standard concepts of utility and preference orderings that economists are familiar with, but also taking account of perspectives and insights from psychology. The third section examines how the collection of data could be improved and how such data can be used for the empirical application of different conceptions of well-being.

**5.1 The Easterlin paradox: Have we been wrong for 70,000 years?**

Our species has considerably evolved since it left its African cradle some 70,000 years ago. It has progressively conquered the whole planet, assimilating or eliminating other human species, and has now gained an amazing and even self-endangering power over its surroundings. It has proliferated considerably, and about two-thirds of it has reached an unprecedented level of longevity, consumption, and sophistication. If happiness surveys had been conducted over this whole period, one would perhaps find a remarkable stability in happiness scores over the millennia. The Easterlin
paradox is, one may think, not just a post-World War II phenomenon, but a more profound phenomenon of stability of subjective evaluations and feelings in our species.

The underpinnings and plausibility of such stability will be discussed in more detail in this chapter. The question to be raised in this paragraph is the following. If this conjecture about the stability of happiness scores over the whole human history were correct, would it imply that all that has been achieved is a loss of energy, a staggering waste of labor and research, a sequence of wrong-headed, narrowly materialistic endeavors?

This formulation is meant to highlight that the happiness literature advocates not just broadening the concept of well-being, but radically shifting the focus in the formulation of human goals. The shift from material achievements to subjective well-being potentially implies repudiating most of the golden calves that have been adored throughout history. Whether subjective well-being scores accurately record what one would like to promote at the individual and social level is the foundation-shaking question that policymakers should answer before using them as their new compass.

It must be noted here that the Easterlin paradox is itself being disputed. First, another claim by Easterlin was that happiness was more or less identical among the nonpoor countries, but the data from the Gallup World Poll have shown a rather neat logarithmic relation between the average satisfaction with life in countries and GDP per capita (Deaton, 2008; Stevenson and Wolfers, 2008). What is called the “Easterlin paradox” here only refers to the long-run stability of satisfaction scores. But even such stability is questioned in Sacks, Stevenson, and Wolfers, (2010). These works will be examined in more detail in the next section. The purpose of this section is simply to argue that, even if it were confirmed over the very long run, the Easterlin paradox would not have the sweeping implications against growth and progress that are often derived by commentators.

(p.164) 5.1.1 Bentham is back

The literature on happiness is not just a cold-blooded observation of facts of the human psyche. For many authors, the “hope is that hedonic psychology will be relevant to policy,” and subjective well-being statistics are meant “to
Is Happiness All that Matters?

provide a more direct assessment of the final outcome about which people are most concerned” (Kahneman, Diener, and Schwarz, 1999, pp. xi–xii). “Economic things matter only in so far as they make people happier” (Oswald, 1997, p. 1815).

Putting happiness at the forefront of social goals raises a number of important questions. First, what is happiness at the individual level? Second, what is it at the social level? Third, why should it be the goal rather than something else, such as wisdom or control over one’s life? Obviously, answers to the last question will depend on how happiness is defined in the previous two questions.

At the individual level, the psychology literature distinguishes many different aspects of subjective well-being (Diener et al., 1999). A key divide opposes cognitive evaluations (what people think of their life) to affective, emotional states (how they feel in their life). The affects themselves come in many shapes and colors, with a surprising independence between the positive and the negative affects, which seem to be connected to different mechanisms in the brain. Attempts to construct classifications of affects along two or more dimensions (e.g., positive/negative valence and high/low activation, as proposed by Russell, 1980) or in terms of discrete categories (Ekman, 1992) have triggered many debates, but it seems that no simple classification is able to distinguish all emotions (for instance, anger and fear are both intense negative emotions). These issues are well summarized by Frijda, (1999).

In practice, it is also not trivial to separate affects from judgments. Ordinary language, which has to be used in questionnaires, is itself ambiguous, as it is not so obvious how to distinguish “to be happy in one’s life” from “to be happy with one’s life,” for instance. When the General Social Survey in the United States asks “Taken all together, how would you say things are these days? Would you say that you are very happy/pretty happy/not too happy?” it is actually hard to tell for sure whether this refers to feelings or to an evaluation. As it turns out, moods are shown to have a strong influence on evaluative judgments (Schwarz and Strack, 1999), and conversely, reports on personal emotions cannot avoid judgmental thoughts (Kahneman, 1999). The happiness literature often appears to assume that there is a single object, like “utility,” that operates behind the scene to generate the answers of respondents. The fact that econometric regressions
of happiness scores and satisfaction scores are generally similar is often taken as an argument in favor of this assumption (Layard, Mayraz, and Nickell, 2008). \( \text{(p.165)} \)

There is, however, a growing evidence of a different pattern for affective scores and evaluative scores. Affective scores seem to be less sensitive to objective conditions of life, more prone to adaptation, than evaluative scores (Kahneman and Krueger, 2006; Krueger and Schkade, 2008; Kahneman and Deaton, 2010).

An important finding of the psychology literature is that people are not very good at predicting their future feelings (Kahneman, 1994; Kahneman and Thaler, 2006). They typically underestimate how easily they can adapt to a change in their situation and exaggerate the effect of changes over future feelings. They also give too much weight to their current mood. This may lead them to badly organize the sequence of activities. Another phenomenon is that at a time of decision about a particular aspect of their life, people typically overestimate the importance of this aspect and do not anticipate that when this aspect drops from the center of attention, it will have much less impact on their well-being. Such phenomena create a wedge between “decision utility” and “experienced utility.” From this standpoint, the Easterlin paradox, for instance, may be imagined to be due to the fact that people underestimate that their efforts to improve their standard of living will not have a long-lasting effect on their subjective well-being.

People also appear to have a peculiar memory about past episodes, giving more attention to the peak and the end of episodes than the overall content (Kahneman, Wakker, and Sarin, 1997). This may be so strong that adding a mildly unpleasant end to a more severely unpleasant episode improves individuals’ memories about it. This phenomenon makes it delicate to measure “experienced utility,” because people’s memories differ from a moment-to-moment recording of the episode.

There are also collective mistakes. Comparisons to others seem an important element of subjective well-being as reported in questionnaires, which suggests that there are large externalities, and therefore room for large inefficiencies. It may in particular happen that people make great efforts to improve their relative position in society. Most people doing
the same, their relative position does not change much, but growth has ensued, without improving subjective well-being substantially. Economic growth could then be, at least in part, the residual waste of the fight for positions. The Easterlin paradox is explained by Clark, Fritjers, and Shields, (2008) as a combination of this phenomenon and of the fact that people adapt to their improved standard of living.

The happiness literature has shown hesitations about whether to focus on affects or on satisfaction evaluations. The latter are cognitively demanding and appear fragile at the individual level. Schwarz and Strack, (1999) emphasize that respondents are influenced by the order and wording of questions and by transient information, or simply by their current mood, (p. 166) and conclude that affects may be a more reliable source of information about subjective well-being. This is also the general tone of Kahneman’s chapter in the same volume (Kahneman, 1999), with the hope that a good measure of affects approaches an “objective” notion of happiness. On the other hand, the fact that adaptation seems stronger for affects makes indicators of affects less attractive if one wants a subjective indicator that has some good link with objective conditions of life. Affects may be easier to measure objectively, but may also give a less relevant perspective on people’s lives.

Let us now turn to the definition of happiness at the social level. At the social level, the distribution of happiness across individuals has to be considered. Most studies focus on average scores and therefore neglect the distribution, but this is seldom defended as a definite neutral attitude about the distribution of happiness. Layard, (2005), for instance, argues in favor of giving some priority to the most miserable. The average score has the advantage of eliminating noises and disturbances that are randomly distributed over the population. Estimating the distribution of happiness is statistically more demanding and requires more confidence about the possibility of comparing score levels across individuals. Truly enough, the average score also requires great confidence about the possibility of comparing score variations across individuals. The risk is that individuals with greater variations will sway the general movement of the indicator. At any rate, the important point here is that it is not
obvious that “happiness” at the social level should be understood as the average score.

Then comes the key normative question: Why take happiness as the goal of society? The happiness literature is not always explicit about an answer; but when it is, the main argument is that this is the goal of the population, “the final outcome about which people are most concerned,” as quoted in the first paragraph of this section. More specifically, many authors argue that happiness is the ultimate goal, and that everything else that is desired is only a means to this end (Diener, 1994; Frey and Stutzer, 2002; Layard, 2005; Oswald, 1997). Bentham’s principle of the Greatest Happiness is then brought back to the fore. As one prominent author in this field writes: “I believe that Bentham’s idea was right and that we should fearlessly adopt it and apply it to our lives” (Layard, 2005, p. 112).

The general outlook emerging from this approach is the following, as well summarized in the preface of Kahneman, Ritov, and Schkade, (1999):

A scientific understanding of hedonic experience will allow for the development of valid hedonic indicators that reflect the pleasantness of life in the everyday experiences of people. At present, economic indicators hold the most sway in (p. 167) policy circles. Yet, the economic approach is limited in several ways. First, it focuses on those aspects of life that can be traded in the marketplace. Thus, desirable goods such as love, mental challenge, and stress are given little consideration. As people reach what Ronald Inglehart has labeled a “post-materialist” level in which basic physical needs are met, they become increasingly concerned with fulfillment in less materialistic realms. Second, the economic view presupposes that individuals will choose the greatest amount of utility for themselves; yet a great deal of evidence now contradicts this proposition. Third, economics assesses variables that are only indirect indicators of something else—of subjective fulfillment. (pp. xi-xii)

In other words, according to this literature the focus on material performance should give way to a focus on subjective well-being. When individuals and societies appear to strive for
something else, such as territorial expansions, technical inventions, greater possessions, or more consumption, they ultimately seek subjective well-being and do so awkwardly. Their strategies actually fail to maximize experienced utility. The new focus on hedonic fulfillment will eventually give people what they have wrongly sought in material performance.

5.1.2 The debate about subjective welfarism

There are many other reasons than the Benthamite principle one may like the political conclusions derived by the happiness literature. The main political conclusions are indeed that one should have a greater concern for poverty alleviation, mental health, pain relief, social relations, leisure. But attractive policy implications do not necessarily make a good normative theory. It is important to assess the soundness of social objectives.

Moreover, real policy challenges may be lurking. Although this is not broadcast by the happiness literature, for instance, the impact of education and technical innovation on happiness is dubious and, from the point of view of happiness, one should wonder whether the high priority they currently receive is deserved. The same could be said about health care and medical research that deals with nonfatal conditions, because subjective well-being tends to adapt to bad health, except mental health and pain. Similarly, while poverty alleviation is important in the hedonist perspective because poverty is associated with low subjective well-being in various direct and indirect ways, one may wonder if inequalities in subjective well-being track the kind of inequalities that one should seek to eliminate. As Graham, (2009) emphasizes in the title of her book, there are “happy (p.168) peasants and miserable millionnaires.” Therefore it is important to examine the arguments for and against the subjective approach to well-being measurement.

There was a philosophical debate about subjective “welfarism” in the 1980s, from which welfarism has not emerged unscathed, and it is interesting to rehearse the main ingredients of this debate.

Three objections have been raised against subjective welfarism. Dworkin, (1981) has argued on grounds of fairness that people who are more difficult to satisfy because they have
more expensive tastes cannot claim that they are worse off and should not request more resources than the others. A more ambitious goal in life cannot be presented as a handicap in the achievement of satisfaction and a reason for an extra amount of help from the other members of society.

Rawls, (1982) has raised a similar point and moreover objected to subjective measures on the ground that they cannot provide a sensible method of interpersonal comparisons. Individuals have different goals in life that are incommensurable. One can compare the resources available to individuals but not their success in achieving their different goals. A measure of success that would operate across a diversity of goals would require an overarching notion of success, and such a notion does not exist and would not make sense.

Sen, (1985) has objected that subjective well-being is too malleable and adaptable to serve as a good proxy for the conditions of life that people care about. One shouldn’t count against a poor and oppressed individual that his subjective well-being is well adapted.

Importantly, these objections are not based on the rejection of individual preferences as a guide for the evaluation of personal situations. Sen’s view probably introduces a distinction between people’s actual preferences, which may be distorted by oppression and social conditioning, and the more authentic form of valuation they would have in better circumstances. Such a distinction is, however, rather standard. As we have seen, the happiness literature also introduces a (different) wedge between a misguided “decision utility” and an authentic “experienced utility” and wants to make people happy against their will if necessary. As Layard puts it, he is “advocating the promotion of good feelings, not the satisfaction of desires” (2005, p. 260). All in all, however, there is a general liberal consensus that in some form, people’s perspective should be the ultimate guideline in the evaluation of individual situations, even if immediate preferences may not be trustworthy.

The above objections against subjective welfarism are therefore not based on the idea that people may be deeply mistaken about their own (p.169) good, and that one should promote their good against their view. Rather, the objections
have to do with the fact that subjective well-being does not provide a good metric of advantage for interpersonal comparisons. One may have a high level of subjective well-being and be disadvantaged, and conversely. The objections all turn around the idea of “happy peasants and miserable millionaires.”

5.1.3 Is happiness the ultimate goal?

In the next section we will examine in more detail how it may happen that subjective scores fail to record a sensible notion of advantage. Let us here focus on a possible response from the happiness perspective. One may think of biting the bullet and arguing that a happy peasant is truly better off than a miserable millionaire, the reason being that happiness is really the ultimate goal of people in life.

If happiness is really the ultimate goal of people in life, the objections listed in the previous subsection are not compelling. Insofar as happiness is comparable across individuals, Rawls’s objection misses the target because happiness provides the overarching measure of success that encompasses all the diversity of possible individual preferences. Sen’s objection points to the real phenomenon of adaptation, but the correct attitude about this phenomenon is to worry about the ills to which people do not adapt, such as pain and noise, rather than insisting that the ills to which they do adapt are still priorities. Finally, if the person with expensive tastes could benefit from psychological counseling that would raise his well-being by adjusting his goals, this would invalidate Dworkin’s view that this person is not disadvantaged.

The argument that happiness is the ultimate goal in life is so popular in the happiness literature that it deserves some scrutiny. The fact that happiness is multifaceted makes it difficult to assess. It may be true for some dimensions of subjective well-being and wrong for others. There may also be differences between individuals. And the picture is complicated by the fact that most people are mistaken, in their daily decisions, about what drives happiness.

There is an ambitious and a modest version of the argument. The ambitious form seeks to prove that happiness is the ultimate goal. One finds such an attempt in Layard, (2005, p. 113): ‘Happiness is that ultimate goal because, unlike all other
goals, it is self-evidently good. If we are asked why happiness matters, we can give no further, external reason. It just obviously does matter.’

(p.170) This argument has two weaknesses. First, it is not true that happiness cannot be sought for external reasons. People want to be happy, among other things, in order to carry out their tasks and duties unencumbered by depression and bad feelings. There are other good effects of happiness: it makes one more pleasant to one’s family and relatives, it is good for health and longevity, and so on.

Second, and more importantly, even if happiness were the only aspect of life that is desirable only for itself, and all other aspects had some impact on happiness, this would not show that the other aspects are not also partly desirable for themselves. In economics, this is a familiar pattern. It is common, for instance, to consider the possibility that health is an argument of the utility function (it is desirable in itself) but is also an argument of the wage function (good health increases earnings) and is therefore indirectly desirable. The fact that health is indirectly desirable does not prove that it is not also directly desired. Likewise, the fact that many good things in life enhance happiness does not imply that people want them only for the purpose of a greater happiness. Therefore, the strong form of the argument fails.

The more modest form of the argument is that happiness is actually people’s ultimate goal. This is a factual statement, which may be true or false. Presumably, it cannot be true simultaneously for all aspects of subjective well-being, and we will examine its application to hedonic scores and satisfaction scores successively.

Let us therefore first contemplate the thesis that people, in absence of psychological mistakes, would be most concerned, for themselves, with experiencing certain feelings (i.e., experiencing certain positive feelings, avoiding certain negative feelings, with a certain temporal pattern). This thesis is, however, factually wrong. There may be people who are, ultimately, exclusively interested in experiencing certain feelings but they most probably are a minority of the population. Most people are interested in other things than their own mental states and care about a variety of achievements. Taking happiness as the ultimate goal in life is
far from normal and popular. It corresponds to an outlier in the range of life goals that ordinary people adopt, and an outlier that does not appear to deserve much praise in ordinary morality.\footnote{Psychologists have studied the link \textit{(p. 171)} between goals and affects and noted the complexity of human goals and their multiple associations with affects and cognitions \cite{Pervin,1983}.}

At this point, an advocate of hedonism may turn normative and say that people should care only about happiness (as far as they are concerned—they may of course also care about the others’ happiness) and are mistaken if they do not. But it is hard to see what their mistake is, given that, as explained previously, there is no logical proof that happiness is the only ultimate value.

Let us therefore abandon the idea that people are most concerned with feelings, although this may be true for some, and turn to the alternative idea that people are most concerned with their life evaluation, that is, with the cognitive judgment they are requested to make in a satisfaction questionnaire. There is a sense in which the thesis then appears tautologically true. Life evaluation is supposed to synthesize what people care most about in their life, and therefore it is definitely \textit{about} what they are most concerned with. There is, however, a difference. People are tautologically concerned ultimately about the \textit{object} of their satisfaction. This is not the same as saying that they are concerned about their satisfaction. In a sense, their satisfaction is what they care the least about! It is not satisfaction that makes a good life, but a good life that gives satisfaction. Satisfaction is the consequence of having a life full of valuable things, but is not a source of value by itself.

Another way of explaining this point is that one must distinguish “obtaining what one wants” from “being satisfied,” and realize that people are interested in the former, not in the latter. There are three ways to be satisfied, and obtaining what one wants is only one of them. One can also become satisfied by adapting one’s aspiration level or by adapting one’s preferences. Everyone can check for oneself if it is more attractive to obtain satisfaction by getting what one wants or by manipulating one’s aspirations and preferences.\footnote{Another way of explaining this point is that one must distinguish “obtaining what one wants” from “being satisfied,” and realize that people are interested in the former, not in the latter. There are three ways to be satisfied, and obtaining what one wants is only one of them. One can also become satisfied by adapting one’s aspiration level or by adapting one’s preferences. Everyone can check for oneself if it is more attractive to obtain satisfaction by getting what one wants or by manipulating one’s aspirations and preferences.}
There is, truly enough, a positive effect of feeling satisfied on emotions, self-confidence, and the like. This is why people do try also to manage their level of satisfaction by adopting modest aspirations and reasonable preferences. But even then satisfaction has an instrumental value, not an intrinsic value. The true source of value remains in feelings and objective achievements.

**Economists are used to describing individuals that “maximize $u(x)$,” and this gives the profession a natural belief in the fact that utility is the ultimate goal of individuals. But if one thinks of $u(.)$ as a representation of the preference ordering, and of $x$ as a comprehensive description of all things that the individual cares about, the maximizing model is compatible with an individual who does not care at all about $u(x)$ and only about $x$ itself. In ordinary economic models in which $x$ is a limited array of economic resources that have merely instrumental value in the achievement of higher goals, the situation is different because the true object of preference is not $x$. Then $u(x)$ may indeed capture other dimensions of life that are not covered by $x$ and that matter a lot to the individual. Once $x$ covers everything that matters, however, utility loses all value except being a representation of an ordering.

Similarly, when one says that an investor is ultimately concerned about the balance sheet, this is a misleading way of saying that he is really interested in the profit—an accountant that would cook a nice balance sheet that does not represent actual profit would not please the investor.

This line of reasoning shows that the argument that happiness is the ultimate value fails, no matter how one defines happiness. Feelings need not be the only source of value, and life satisfaction is not a source of intrinsic value at all, but rather the valueless summary of all other things of value.

**5.1.4 The key objection to subjective scores**

The above discussion of happiness as the ultimate goal leads us to formulate a key objection to subjective welfarism. This objection has not been explicitly articulated in the classical debate on welfarism, but it seems more damaging than the
classical objections because it undermines the core ambition of subjective welfarism to reflect people’s own perspectives.

The objection is the following. Not only does subjective welfarism fail to capture real advantage and to provide a sensible metric for interpersonal comparisons, but by doing so, it betrays people’s own concerns. When subjective welfarism focuses on affects, it betrays people’s concerns when people are interested in achievements other than feelings. When subjective welfarism focuses on satisfaction, it also betrays people’s concerns because it focuses on the outcome of the evaluation instead of the object of the evaluation (p.173). For instance, if it were possible to move to a situation in which people would have less of what they want but would have lower aspiration levels and end up being more satisfied, the welfarist approach would approve this change, but the people themselves would oppose it.

Layard acknowledges that “the thought that you can be happy in that way has led many people to reject altogether the idea of happiness as a goal” (2005, p. 73). His reply is that this way of raising happiness does not work in practice because people want to set themselves aspirations that are sufficiently challenging to avoid boredom. This reply is insufficient. First, the objection was a conditional. If one could make people happier by lowering their standards of satisfaction, that would be recorded as an improvement by the subjective well-being indicators. This reveals a flaw in such indicators even if this phenomenon never occurs. Second, a similar phenomenon does occur, on a very large scale, when satisfaction remains stable due to a shift in aspirations while conditions of life have changed tremendously.

The Easterlin paradox is, precisely, the best illustration of this problem. It appears most likely that people have a strong preference for the greater standard of living and longevity they enjoy now than 50 (or 70,000) years ago. The object of their satisfaction has improved and in this sense they are definitely more satisfied. But the level of satisfaction expressed in questionnaires may remain stable because the standards of answers shift with people’s actual conditions of life. There is no paradox in the flatness of satisfaction curves because such curves do not reveal anything substantial about how much of people’s goals is achieved.
An important caveat must be made here. These observations are compatible with the possibility that the direction of human development has been ill-chosen, that hubris and materialism have made the success of our species a catastrophe for the rest of the planet. Perhaps human goals and values are deeply questionable and should be radically revised. These are important moral issues that will not be discussed here at all. The point of this section is not to say that development has been good in spite of the stability of satisfaction curves. The point made here is that such stability, if confirmed, is not a proof that human development has been pointless or against human values and goals. Most probably, development has reflected actual human objectives. Such objectives may be wrong for moral reasons, but happiness data do not show it.

The “paradox” in Easterlin’s paradox is that it has served to popularize the happiness challenge to the economic approach, whereas it is the best proof that the happiness approach is problematic. Subjective well-being indicators fail to accurately record the object of people’s values and goals because they either focus on a narrow aspect of people’s lives (their feelings) or focus on happiness and satisfaction scores that rely on standards of comparisons that are heterogenous across individuals (happy peasants, miserable millionaires) and drift along history (Easterlin’s paradox).

This does not imply that happiness data are useless. Imagining a good use for them requires a little more exploration of how they are generated and can be processed. This is the topic of the next sections.

5.2 A theory of subjective well-being
In this section we will try to construct a model of subjective well-being as it can be retrieved with typical questionnaires. This will put more flesh on some of the ideas put forth in the previous section and make it possible to think about how to make use of happiness data, which is the topic of the next section.

We will build on Schwarz and Strack, (1999), who propose the following informal model—what is drawn in figure 5.3 is a simplified version, as their model also covers the evaluations of specific domains of life such as income or family life.
Their model reads as follows. When assigned the task of evaluating life, the individual considers if her current mood is informative. If not, she has to retrieve other relevant information, define what the target (“my life”) consists of, and choose a standard of evaluation (e.g., “my neighbors”). A key element of the evaluation is that some elements may either be assimilated into the object of evaluation or act as contrasting background. After evaluation, she still has to see if the answer is appropriate in the context of the interview. If not, she can edit the initial judgment (e.g., make it more rosy).

5.2.1 Affects and judgments

Let us now develop our own model. First, let us think of the life of an individual \(i\) as a vector \(l_i\) of many dimensions that cover the diversity of states, activities, possibilities enjoyed or endured by the individual over the course of her life. At the time of answering a questionnaire, the individual is alive and not sure about her future life; therefore the dimensions of life include the various scenarios for the future that may unfold. Let us take the convention that beliefs about the probabilities of these scenarios are not part of life. This is convenient because it is not obvious that one should adopt the individual’s own beliefs for the evaluation of life. For instance, if the individual is much too optimistic about her life expectancy, it may be advisable to use better expertise in order to evaluate her situation. This convention permits some flexibility about the choice of probabilities. Let \(l_{ts}\) denote the subvector of life events that happen for \(i\) at time \(t\) in state (i.e., scenario) \(s\). The individual’s belief about the probability of state \(s\) is denoted \(\pi_{ts}\).
Where does subjective well-being find its place in the model? First, the feelings experienced at a particular date \( t \) form a subvector \( f_{\text{fts}} \) of vector \( l_{\text{fts}} \). Importantly, feelings are multidimensional, and different people may put different weights on different feelings depending on their psychology or their values, and also depending on the circumstances. As explained in the previous section, for most people the desired feelings are not a constant sequence of intense exhilaration, or even of quiet contentment, but rather a sequence that fits the circumstances of life and enable the individual to cope with them in the best way according to her own values. It is interesting that the ranking of feelings cannot be “objective” because the feelings one considers fitting with the circumstances depend on how one wants to react to the circumstances. For instance, some people consider it weak to mourn for a long time, while others consider it the best way to pay tribute to the beloved deceased. Even something as simple as the distinction between negative and positive feelings is, in part, value laden and culture dependent. Some people like feeling proud, while others find it an (p.176) uncharitable and egoistic emotion. Hatred is generally considered a negative feeling but is praised by some. Love and attachment for certain objects (animals, gold) is good for some and despicable for others.

In summary, we have \( l_t = (f_{\text{fts}})_{ts} \) and \( l_{\text{fts}} = (f_{\text{fts}} \cdots) \). For present and past periods, there is only one state, and one simply writes \( l_t = (f_{\text{fts}} \cdots) \).

It remains to find a place for life evaluation (as opposed to feelings). Here there are two possibilities. The simplest one assumes that the evaluation of life is not part of life, even though it is an act of the individual. When the individual \( i \) answers a happiness or satisfaction questionnaire, she looks at \( l_t \) and there is a function \( \xi_i \) that maps lives into the set of possible answers to the questionnaire. Therefore, \( i \)'s answer is \( \xi_i(l_t) \).

The more complex alternative takes the act of evaluating life as part of life. When the individual formulates the evaluation \( e_i \) at time \( t \), this is a component of vector \( l_t \). However, one may safely assume that the evaluation is not a function of itself but only of all other components of \( l_t \), which means that there is a
mapping $\xi^*_{i}$ such that $e_{it} = \xi^*_{i}(l \setminus e_{it})$, where the notation $l \setminus e_{it}$ means that the component $e_{it}$ is removed from vector $l$. What may happen, nevertheless, is that $e_{it}$ has an impact on other components of $l$. For instance, a positive judgment may produce good feelings, or may give self-confidence that enhances future prospects. It would strange, however, if such effects were to make the computation of $e_{it} = \xi^*_{i}(l \setminus e_{it})$ problematic because of divergence.

We will work here with the first, simpler approach $\xi(l_i)$, which does not appear to miss much of the relevant phenomena. The question is now to understand how $\xi(l_i)$ is constructed by the individual.

The value of $\xi(l_i)$ must lie in a given scale, which can either be a verbal scale (e.g., very happy/pretty happy/not too happy, or very satisfied/fairly satisfied/not very satisfied/not at all satisfied), or a numerical scale (e.g., from 0 to 10). The cognitive problem for the individual is to put the many dimensions of $l_i$ into one of a few ordered categories.

5.2.2 The three problems of the respondent

This problem can be decomposed into three problems: (1) the scope problem: what part of $l_i$ is relevant? (this is what Schwarz and Strack call the target); (2) the ranking problem: how does $l_i$ stand in the set of relevant possible lives? (3) the calibration problem: how does a position in the ranking translate into a category of the questionnaire? (this is what Schwarz and Strack call the standard). Schwarz and Strack do not distinguish the last two problems, but this distinction, familiar to the economist, appears (p.177) crucial. Giving a grade from 0 to 10 to one’s life involves first identifying where one’s life stands in an ordinal ranking and then associating positions in the ordinal ranking to grades on the numerical scale. The two tasks are cognitively different.

5.2.2.1 The scope problem

The scope problem is shaped by the question itself and especially its ambiguities. Table 5.1, reproduced from Layard, Mayraz, and Nickell, (2008), lists the various questions featuring in the main questionnaires. 
## 5.1 Happiness and satisfaction questions

<table>
<thead>
<tr>
<th>Survey</th>
<th>Variable</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Social Survey</td>
<td>Happiness</td>
<td>Taken all together, how would you say things are these days? Would you say that you are very happy, pretty happy, or not too happy?</td>
</tr>
<tr>
<td>World Values Survey</td>
<td>Life sat.</td>
<td>All things considered, how satisfied are you with your life as a whole these days? Please use this card to help with your answer. [Range of 1–10 with 1 labeled. “Very Dissatisfied” and 10 labeled. “Very Satisfied”]</td>
</tr>
<tr>
<td>European Social Survey</td>
<td>Happiness</td>
<td>Taking all things together, how happy would you say you are? Please use this card. [Range of 0–10 with 0 labeled. “Extremely unhappy” and 10 labeled. “Extremely happy”]</td>
</tr>
<tr>
<td>European Social Survey</td>
<td>Life sat.</td>
<td>All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card, where 0 means extremely dissatisfied and 10 means extremely satisfied. [Range of 0–10 with 0 labeled. “Extremely dissatisfied” and 10 labeled. “Extremely satisfied”]</td>
</tr>
<tr>
<td>European Quality of Life Survey</td>
<td>Happiness</td>
<td>Taking all things together on a scale of 1 to 10, how happy would you say you are? Here 1 means you are very unhappy and 10 means you are very happy.</td>
</tr>
<tr>
<td>European Quality of Life Survey</td>
<td>Life sat.</td>
<td>All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied.</td>
</tr>
</tbody>
</table>
Survey | Variable | Question
---|---|---
German Socio-Economic Panel | Life sat. | In conclusion, we would like to ask you about your satisfaction with your life in general. Please answer according to the following scale: 0 means “completely dissatisfied,” 10 means “completely satisfied.” How satisfied are you with your life, all things considered?

British Household Panel Survey | Life sat. | How dissatisfied or satisfied are you with your life overall? [Range of 1–7 with 1 labeled. “Not satisfied at all” and 7 labeled. “Completely satisfied”]


Most questions indicate that everything must be taken into account, but this does not dispel difficult boundary questions. A first and obvious difficulty for all such questions relates to the time frame. Does the expression “these days” or “nowadays” that appears in many of the questions refer to the time of evaluation (what is your judgment these days) or to the phase of life to be considered (your life in the current period)? Even when the expression is absent it is not obvious to decide whether one should think of one’s whole life, or only of one’s past life, or only of the current period. If the latter, the length of the period can go from a few days to several years.

Another difficulty has to do with relatives. Is the question purely personal, or should one incorporate one’s family in the evaluation? Even if personal life partly depends on how one’s relatives are doing, the weight attributed to their situation may greatly vary with how strongly the survey suggests that only the purely personal aspects are relevant for the question.

A third related ambiguity has to do with the general social context. It is part of a good life to be the member of a good polity, to be part of a world that goes well. For many people, what goes on in the world around them, or at least in their community, is like a public good or bad that directly affects their satisfaction. Whether this is part of their life or not is not easy to tell in the context of such a satisfaction question.
Deaton, (2012) observes that when the question is preceded by a question on politics, the level of satisfaction goes down substantially. This may due to the fact that the presence of the political question induces people to broaden the scope of their thoughts when answering the satisfaction question. As it is a popular thought to imagine that one’s strictly personal situation is better than the situation of one’s country and that the government’s record is bad, the decrease in satisfaction level in this context is not surprising.

(p.178) More generally, one may suspect that the ambiguity about scope is most likely to create a strong dependency of answers on the order of questions and various framing aspects of the questionnaire. When a question is ambiguous, people seek clues in the other aspects of the interview. The ambiguity of scope is likely to create uncontrolled diversity among respondents and to generate substantial noise.

(p.179) Formally, let $\vec{f}$ denote the subvector of $\vec{t}$ that the individual considers relevant in the context of the questionnaire. For simplicity’s sake, we can retain the same notation for the evaluation function and let $\xi(\vec{f})$ describe how the answer is determined.
5.2.2.2 The ranking problem

A happiness or satisfaction question does not expressly request the respondent to form a ranking. And one may suspect that very few people have well-formed preferences about their life in general. But when asked such a question, one is forced to shrink the complexity of a multidimensional life into an ordinal pattern, and this requires the prior construction of a ranking, no matter how coarse.

The cognitive complexity of this exercise may induce respondents to focus on some salient aspects of their situation and forget many relevant dimensions of their life. This is linked to the “focusing illusion,” by which people tend to exaggerate the importance of aspects under consideration at the moment (Kahneman et al., 2006). Here again, the general structure of the questionnaire plays an important role. When the question is asked in the beginning, or after a series of questions that are not directly relevant, the construction of the ranking is particularly complex. When the question comes after a comprehensive examination of the various domains of life, the task is made somewhat less complex and the respondent should be less vulnerable to a focusing illusion—although the respondent may then be more vulnerable to the framing impact of the earlier questions.

Given that the scope problem has led the respondent to specify a subvector $l^*_i$ of $l_i$ to which attention is restricted, the respondent only needs an ordering of the possible vectors $l^*_i$. But the respondent is likely to have some preferences about the other components of $l_i$ and also about other aspects of the universe than $l_i$. Deriving a ranking of subvectors from the ranking of full vectors is easy if the excluded components are assumed to have a fixed value. One would naturally expect the respondents of a questionnaire to take the current values of these excluded components as the fixed value.

But fixed values for the excluded components may not be easy to assume in the context at hand, and this creates specific difficulties. Consider the case of a respondent from a rich country who has to imagine where to put her life on a 0–10 scale. Presumably, 0 is the grade of the worst life she can think of and 10 is the grade of the best life. But she may think that the worst life would consist of living in a destitute country with a lot of insecurity. Some aspects of this different situation
would change the components of \( l_i \cap l_i^* \), that is, the excluded components. Now, these different values for \( l_i \cap l_i^* \) may alter the ranking of \( l_i^* \). For instance, in her quiet and affluent country she likes devoting her life to art, but if she were in the poor country she would prefer devoting her life to politics and would consider it scandalous to be an artist. It may therefore be that in her current situation she prefers \( l_i^* \) (artist) to \( l_i^* \) (activist), but if she were in the poor country she would hold the opposite preference. It may, however, be the case that in the two contexts she would agree on what the worst life \( l_i^* \) is, but this cannot be guaranteed. In summary, while defining an ordering of \( l_i^* \) is in principle easy when one has an ordering of \( l_i \) and the excluded components are fixed, it may be unrealistic to assume that the excluded components are fixed when \( l_i^* \) takes all possible values. Then there may not be a well-defined ordering over \( l_i^* \).

Expectations about the future are an interesting element to discuss at this point. If \( l_i^* \) includes the future, then the respondent’s expectations over \( l_i^* \) and in particular her beliefs \( \pi_{lw} \) will directly play a role. If \( l_i^* \) does not include the future, the respondent’s preferences over her whole life \( l_i \) will make it tricky to define preferences over \( l_i^* \). Very different values for \( l_i^* \) should, realistically, go with different expectations about the rest of life. Such different expectations may alter the preferences over \( l_i^* \). For instance, if one expects to have a comfortable retirement one may want to adopt a safe lifestyle. If one’s situation is so bad that one does not expect to live long, a less healthy lifestyle may become more attractive. It is then difficult to know what the preferences over \( l_i^* \) are, if this depends on how one revises expectations in the light of realizations in \( l_i^* \).

This issue reveals that satisfaction indicators are, in a sense, much more informationally demanding than measures of well-being that are local in the preference space. The computation of an equivalent income, for instance, only requires information about the current indifference curve of the individual, and there is no need for the individual, is she were asked questions about this indifference curve, to imagine situations that are far away in the ranking, especially the
extreme situations of the “best possible” and “worst possible” lives.6

Let us finally assume that the individual manages to construct an incomplete ranking of vectors $\mathcal{I}$, and let $R^*$ denote this reflexive and transitive but not necessarily complete relation (with associated strict preference (p.181) relation $P^*$ and indifference relation $I^*$). Incompleteness must be limited in some way because it would be problematic if a subset of lives was completely incomparable to its complement subset. One may reasonably assume that for each life it is easy to find a similar life that differs only in a few dimensions and is clearly better or clearly worse. For a given vector $\mathcal{I}$, let $L(\mathcal{I}; R^*)$ denote the subset of vectors that are comparable to $\mathcal{I}$ and no better than $\mathcal{I}$, that is, the set of $\mathcal{I}$ such that $\mathcal{I} R^* \mathcal{I}$, and let $H(\mathcal{I}; R^*)$ denote the subset of vectors that are comparable to $\mathcal{I}$ and no worse than $\mathcal{I}$, that is, the set of $\mathcal{I}$ such that $\mathcal{I} R^* \mathcal{I}$.

5.2.2.3 The calibration problem
The calibration problem is the most interesting in the context of questionnaires. Equipped with an incomplete ordering $R^*$ of life vectors, how can the individual transform the position of his actual life $\mathcal{I}$ into a specific category $\xi(\mathcal{I})$ of the questionnaire?

Here there is a strong framing effect due to the fact that the scale offered in questionnaires is closed. While many aspects of life are open (there is no maximal possible income, and probably no worst possible pain) or have very fuzzy and remote physical limits (bodily and intellectual performance, as well as longevity, have no known limits), giving a closed scale to respondents forces them to move from a reasoning in terms of life content to a reasoning in terms of statistical distribution. They therefore have to determine where their situation lies in a particular distribution. The problem is then to choose which distribution.

Of course, this reference to a statistical distribution should not be taken too literally. Very few respondents will exactly identify the problem with finding a quantile in a distribution of possible lives. But it is clear that, given a fixed number of ordered categories, a respondent must choose a relative position for the grading of his life. It would be strange to hear a respondent saying that the human condition is appalling
anyway, so that he chooses the worst category, or that life is the greatest gift, so that whatever happens he should pick the top of the categories. The ladder-of-life questions that explicitly refer to the “best possible” and the “worst possible” life most effectively force the respondent to think in relative terms. In the Gallup World Poll, the formulation is as follows:

Please imagine a ladder with steps numbered from zero at the bottom to 10 at the top.

The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you.

(p.182) On which step of the ladder would you say you personally feel you stand at this time? (ladder-present)
On which step do you think you will stand about five years from now? (ladder-future)

This approach was proposed by Cantril, (1965), whose purpose was to devise a “self-anchoring scale” in which people would think in relative rather than absolute terms, with standards adjusted to their own perspectives and possibilities. It is interesting to note the ambiguities of the anchoring. The expression “best possible life for you” can mean “the best possible life for a human being, according to your preference ranking of lives,” with a remaining ambiguity about whether “possible” refers to logical conceivability or to practical attainability in the current world; but it can also mean “the best possible life that you could achieve,” here again with ambiguity about how the possible-impossible distinction is to be understood. These ambiguities introduce considerable noise in the data. How can one compare the answer of someone who defines the best possible life as the dreamed life of a human being in two millennia with the answer of someone who defines the best possible life as a moderately optimistic interpretation of his own possibilities? The same kind of ambiguity occurs for the “worst possible life.”

In Cantril’s study, there was no such ambiguity because the ladder question was preceded by the following questions:
1. All of us want certain things out of life. When you think about what really matters in your life, what are your wishes and hopes for the future? In other words, if you imagine your future in the best possible light, what would your life look like then, if you are to be happy? Take your time in answering; such things aren’t easy to put into words.

2. Now, taking the other side of the picture, what are your fears and worries about the future? In other words, if you imagine your future in the worst possible light, what would your life look like then? Again, take your time in answering.

What is remarkable about Cantril’s study is that it gives us an idea of what the best and worst life looks like for different people. Here are two contrasting examples about the best life, one from India and the other from the United States:

I want a son and a piece of land since I am not working on land owned by other people. I would like to construct a house of my own and have a cow for milk and (p.183) ghee. I would also like to buy some better clothing for my wife. If I could do this then I would be happy.

I would like a reasonable income to maintain a house, have a new car, have a boat, and send my four children to private schools. (Cantril, 1965, pp. 206 and 222)

It is a popular rebuttal of figures describing the Easterlin paradox that GDP is on an open scale whereas satisfaction scores are bounded by construction. But the real difficulty with satisfaction questions is that the true scale of life is more like GDP and is not naturally bounded. Respondents are therefore induced to reason in relative terms when they must describe an open-ended object, their lives, in a closed scale.7

Sometimes this does not look so artificial. There are dimensions of life that are bounded. One’s social status, one’s position in the distribution of income, and all similar relative positions are already cast in the relativistic frame. Health is also conceived in comparison to perfect health. Obviously, for individuals who care only about their relative position in their community, their task is made easier because they do not have
to make a conversion to a relative scale as they already think in these terms.

In summary, the fact that life is made of a combination of open-ended and bounded dimensions makes it necessary to answer the satisfaction questionnaires in relative terms. This does not necessarily mean that a comparison to other members of society has to be made. But a comparison to some benchmark seems unavoidable, and the reference to the best and worst possible lives in the ladder questions makes it explicit.

Formally, thinking about life in relative terms amounts to defining a distribution over the set of $l_i^*$. Let $\mu_i$ denote the probability measure that the individual chooses. Then a natural way to evaluate the relative position of $l_i^*$ is to compute $\mu_i(L(l_i^*; R_i^*))$, or $1 - \mu_i(H(l_i^*; R_i^*))$.

In the literature it is common to introduce relative terms that take the form of a ratio in which the denominator is an average value. This is a special case of what is described here. For a real random variable $y$ that is uniformly distributed over an interval, the ratio $y/\bar{y}$, where $\bar{y}$ is the mean value, is an affine transform of the cumulative distribution function. For a distribution with unbounded support (such as the normal or the lognormal distribution), however, $y/\bar{y}$ is not a bounded variable and is not a convenient way of squeezing $y$ into a bounded scale.

(p.184) From $\mu_i(L(l_i^*; R_i^*))$, or $1 - \mu_i(H(l_i^*; R_i^*))$, to $\xi_i(l_i^*)$, it only remains to partition the $[0, 1]$ interval into subintervals. This creates a secondary calibration problem, but we can assume that a uniform partition is adopted and that the main calibration operation does all the hard work.

Let us now examine the key task for the respondent, which is to define a standard, that is, the distribution $\mu_i$ that will do the calibration. Three natural possibilities arise. A first one is the distribution of possible lives as expected, some time earlier, by the respondent. The calibration is then driven by expectations or aspirations. A second possibility is to look at the distribution of other people’s lives, in some reference group. The calibration is then driven by comparisons to others. A third possibility, which makes sense only when $l_i^*$ only refers to the
current live vector $l_{t_0}$, where $t_0$ is the time of interview, is to look at the past experience of the respondent. The calibration is then said to rely on the past. The past experience of the respondent is typically a degenerate distribution with which it is not possible to directly compute $\mu(l_0; R^*)$ because $l_0$ does not belong to its support. One must then define a distribution around the past experience in order to define what is a “small” or a “big” improvement or deterioration.

Note that for each of these three possibilities, there are many possible choices. Aspirations can be more or less ambitious, comparisons to others may refer to various reference groups, and reference to the past may look at different periods. The consequence for analysts using happiness data is that the calibration problem is a source of heterogeneity across individuals. Moreover, respondents can construct $\mu_1$ by mixing various sources of inspiration. Different measure functions can be combined easily to form a new measure function, simply by computing a weighted average. Some individuals may give a greater weight to comparison to others, while other respondents may primarily refer to their own aspirations, but all of them can give some positive weight to these different considerations.

Steffel and Oppenheimer, (2009) study the comparisons standards used by respondents in the context of simple happiness questions (“How happy are you?”). First, their study confirms that, even for such a direct question that could be interpreted as referring to affects rather than life evaluation, people do make comparisons to calibrate their answers. Second, there is a variety of standards used, with a majority of intrapersonal comparisons (the reference distribution being “what is typical for me”) over interpersonal comparisons in their study. Third, the respondents comparing themselves to others tend to give higher happiness scores than those making intrapersonal comparisons. One of the conclusions of their paper is that the surveys should try to eliminate the heterogeneity of standards that renders the answers hard to compare across respondents.

Finally, it must be emphasized that the model that is proposed here, and that can be summarized by the formula

$$\xi_1(l) = \mu_1(l; R^*)$$
has variants. For instance, instead of going through the scope-ranking-calibration stages listed above, the respondent could do the three exercises at the level of life domains first, obtaining satisfaction scores for each life domain. In this case, global satisfaction can then be constructed as a weighted average of the scores of the various life domains. This has perhaps the advantage of making the ranking exercise easier because life domains contain less heterogeneous subdimensions. But the hard work of making life domains commensurable is then postponed to the last stage of aggregating scores and cannot be escaped completely. This variant is particularly plausible for questionnaires in which the respondent is indeed asked to give scores for life domains before being asked the overall satisfaction question.

Schwarz and Strack’s model gives great importance to the role of mood as an informational shortcut. Given that current mood at the time of interview has been shown to influence satisfaction scores, it is indeed interesting to understand how that can happen. The hypothesis that mood makes the respondent bypass the whole reasoning and delivers an immediate score is a rather extreme possibility. Other hypotheses that might be worth exploring are that a good mood distorts the ranking phase by making the respondent focus on positive aspects of her situation (selective attention), or distorts the calibration phase by making the respondent take a less favorable reference distribution \( \mu \), or by inducing the computation of \( \mu(L(t; R')) \) to be biased upward. All three mechanisms can be partly driven by the unconscious desire of the respondent to generate feelings that confirm the current mood. A good mood is reinforced by a good evaluation.

5.2.3 Heterogeneous and shifting standards

The model proposed in the previous subsection suggests that respondents give answers that are hard to compare when they use heterogeneous strategies or different sorts of information to cope with the scope, ranking, and calibration problems.

Although this is an empirical statement that has to be ultimately decided by the data, one may safely conjecture that the main source of (p.186) heterogeneity lies in the calibration problem. A happy peasant and a miserable millionaire cannot diverge so much in the evaluation of lives because they focus on different aspects of their lives when
they define \( I_i \) and construct \( R_i^* \). They must take very different reference distributions \( \mu_i \) in order to transform their presumably similar rankings into bounded scores. Their personal expectations, their community, their own past, all three typical sources of inspiration for the choice of the distribution \( \mu_i \) can indeed differ exactly in the way that generate their different evaluation standards. One is more easily satisfied if one had low expectations, if one compares oneself to a less well-off community, and if one comes from a disadvantaged background. Cantril, (1965) showed this very early on by eliciting free descriptions of the “best possible life.” Loewenstein and Ubel, (2008) and Graham, (2009) emphasize such diversity, as induced in particular by the adaptation phenomenon (Brickman and Campbell, 1971).

In particular, the stability of satisfaction scores as epitomized in the Easterlin paradox can easily be explained by the fact that all three sources of calibration tend to follow the evolution of people’s actual situations. Along the development path, expectations move up, the rest of society moves up, and the past experience, if defined by a fixed lag to the present, also moves up. Therefore, if, due to these shifting references, the distribution \( \mu_i \) follows the change in \( I_i^* \), the stability of \( \mu_i(L(I_i^*; R_i^*)) \) can be observed in spite of a constant improvement of the position of \( I_i^* \) in the ranking defined by \( R_i^* \).

Di Tella and McCulloch, (2006, p. 35) observe that “for both adaptation and relative income effects to be relevant explanations of the Easterlin, (1974) paradox we would need a very specific pattern: it would have to be the case that individuals adapt to income, but do not adapt to their relative position.” They go on to note that this pattern seems to be confirmed by the data. The model proposed here offers a simple explanation of this phenomenon. The relative position does not need a calibration operation, whereas an unbounded variable such as income does. Someone who is relatively well-off in her reference group can steadily give a higher satisfaction score, but the long-run increase in income shifts the distribution \( \mu_i \) and therefore disappears in the satisfaction score.
Adaptation to relative position is, truly enough, logically compatible with the model. It could happen if the individual cares about her relative position as such and uses her recent past experience in the relative position to calibrate the score. But that requires a very specific pattern of ranking (caring about the relative position) and calibration (using the recent past about the relative position). In particular, it appears rather unnatural to recalibrate a variable that already comes in a relative form.

The model also suggests why it is possible for satisfaction scores to appear more sensitive to objective conditions of life than affective scores that record emotion flows, although emotions appear negatively affected by low income (Kahneman and Krueger, 2006; Kahneman and Deaton, 2010). Suppose that, outside the stressful conditions of deprivation, emotions are managed by individuals in such a way that they always come back to their set level, and that this set level depends on personality more than on the contents of life. Then affects are not very sensitive to conditions of life, except in seriously disadvantaged situations. In contrast, the shifting standards that reduce the sensitiveness of satisfaction scores to life do not attenuate the relative outcomes (except in the very special configuration described in the previous paragraph). Insofar as individuals care about their relative position or use it to calibrate their answers, their scores will therefore be stable at various unequal levels corresponding to their position in the distribution of objective outcomes.

A challenge to the idea that people use heterogeneous standards across space and time has been raised by Deaton, (2008), Kahneman and Deaton, (2010), Stevenson and Wolfers, (2008) and Sacks, Stevenson, and Wolfers, (2010). Consider figure 5.4, showing the average level of satisfaction in various countries against GDP per capita. This figure reveals a rather nice log-linear relation, although it also shows great variations of satisfaction at every level of income. Moreover, Sacks, Stevenson, and Wolfers, (2010) show that the slope of this log-linear cross-section relation between countries is generally similar to the slope of the relation within countries. As far as change over time is concerned, they note that the long-run evolution of satisfaction in various countries has been quite diverse, with some countries exhibiting a negative relation in the long run between income and satisfaction. However, looking at the whole set of countries in their sample, they
show that the data are, on average, compatible with a positive relation with income, with a similar slope as in the cross-section picture, although the degree of confidence is low because the data are still sparse for the long run.

The log-linear relation between income and well-being can be interpreted in at least two ways. Kahneman and Deaton, (2010) and Deaton, (2012) refer to Weber’s law. This law says that the smallest noticeable variation of a stimulus is proportional to the initial magnitude of the stimulus. The associated Fechner law implies that the relation between stimulus and response is log-linear (which derives from Weber’s law if the response is linear with respect to perceptions). Although income is measured in monetary terms (p.188) and has objective smallest units, it is plausible that the stimulus variation that people feel is the percentage increase in income rather than the absolute increase. This is how a log-linear relation can be obtained.

This interpretation suggests a stable relation between income and satisfaction over time, but is compatible with substantial heterogeneity across individuals and countries.

Note that a logarithmic curve goes from $-\infty$ to $+\infty$, so that the Weber-Fechner laws cannot be the whole story for a bounded scale. Prolonging the curve drawn by Deaton, (2008) would cross the 10 ceiling at about $300,000.

There is, moreover, an alternative and not less natural interpretation of the logarithmic curve suggested by the figure. Such a curve is also the shape of a cumulative distribution function. In particular, the CDF of a distribution of
income may easily display this shape. Figure 5.5, from Sala-i-Martin, (2006), shows an estimation of the CDF of income in the world, with income measured in logarithmic scale.

The curve is not straight, but if one removes the levels of income below $500 and above $40,000, the 2000 curve is relatively straight. So the shape of satisfaction answers can also be explained by the fact that people take inspiration from their rank in the world distribution of income in order to (p.189) answer the satisfaction questionnaire. Of course, people’s actual position in the distribution need not correspond to their perception, and it is the latter that would be the basis of their satisfaction.

We therefore have two conflicting interpretations of the shape of the satisfaction curve. They induce different predictions about long-term movements. With the Weber-Fechner interpretation, all countries move along the fixed curve. With the CDF interpretation, the curve moves horizontally across time as all countries grow and changes shape as the distribution changes.

The two interpretations are not exclusive and can be combined in the following way. One can imagine that the area of satisfaction below 5 is one in which subsistence concerns matter, giving a role to the Weber law, whereas the area of satisfaction levels above 5 is more influenced by comparisons with the current distribution. Therefore one could imagine that along history, there is an increase in the average satisfaction of countries that move out of dire poverty, whereas satisfaction in well-off countries remains stable. If the world evolved toward a less unequal distribution of income, the curve would move up above the 0–5 area, but would retain its CDF shape above that.
Imagine a scenario in which, over the next century, the situation of European countries stagnates, whereas Northern America benefits from a multiplication of its GDP per capita by 10, and the life expectancy of its population doubles (this would mimic the current gap between the rich countries and the poor countries). The Weber-Fechner interpretation would have it that European satisfaction would remain as high as it is (p.190) nowadays. It appears more realistic to conjecture that, contemplating the much greater standard of living and longevity in Northern America, the Europeans would revise their view of what the “best possible life” is and would have a substantially lower satisfaction as a result.

Coming back to the question of whether calibrations are homogeneous and stable or heterogeneous and shifting, we make three concluding remarks. First, figure 5.4 reveals not only a general trend, but also a great variety. The countries are not on the same line. In particular, one sees that the intermediate levels of satisfaction are experienced by countries of vastly different levels of income. It is really hard to believe that Brazilians are on average better off than the Japanese, for instance. Therefore these data show not only a general pattern, but also a lot of noise that reveal the different calibrations used by respondents of different circumstances. Far from defeating the thesis that there are heterogeneous standards in the surveys, these data confirm it.

Second, the analysis of the long-run relationship between income and satisfaction in Sacks, Stevenson, and Wolfers, (2010) reveals a great variety of situations over the world. Whereas the data are compatible with an average positive relationship, Sacks, Stevenson, and Wolfers, (2010) emphasize that this relation is not robust to the composition of the sample and show that the situation is not a uniform move of all countries along the same slope. There is a great variety of slopes, and a sizable group of countries even experienced a negative relation between income and well-being. The circles in figure 5.4 are, therefore, not slowly moving along the average log-linear line, they are following complex patterns. Di Tella and McCulloch, (2010) separate the sample into two subsets according to the rank of countries in 1960 in terms of GDP per capita and show that the satisfaction level in the poorer half did benefit from growth between 1960 and 2005, whereas the richer half did not show any significant impact of GDP growth on satisfaction. These authors interpret these
results in a way that is similar to the combination, proposed earlier, of the Weber law for low levels of satisfaction with the CDF interpretation for high levels. Another possible explanation is that the CDF theory applies everywhere, with a sizable group of respondents in the poorer group of countries feeling that their rank in the distribution has improved.

Third, as explained in the previous section, these debates about whether satisfaction is stable over space and time or increases with growth are really about the quality of such surveys, not about the true evolution of well-being. If the data revealed no link between satisfaction and income, this should create suspicion about satisfaction surveys, primarily. If the data reveal some link between satisfaction and income, this is better but still (p. 191) does not prove that satisfaction surveys are reliable in tracking what people care about, especially if the link appears very noisy.

5.2.4 What do people care about?

The model proposed here also helps understanding how subjective well-being data relate to “the final outcome about which people are most concerned” (Kahneman, Diener, and Schwarz, 1999, p. xii).

As we have said, individuals may care about many things, of which the components of $I_1$ are just a part and those of $I_2$ an even smaller part. Let us make abstraction of the excluded dimensions and focus on how individuals care about $I_2$. The question is whether $\xi(I_2)$ is a good representation of “the final outcome about which people are most concerned,” and can be used to compare situations across time and across individuals and populations.

The fact that different people may focus on different parts of $I_1$ when they address the scope problem, and on different parts of $I_2$ when they try to make a synthetic ranking of lives, diminishes the comparability of answers, as we have seen. But the key problem is, once again, the heterogeneity of calibration strategies. Forgetting the excluded components, a good life is a good $I_2$, as assessed by $R^*_1$. If one takes two individuals, Abel and Bael, who share the same $R^*_2$, and they both agree that Abel has a better life, it seems reasonable to
conclude that Abel is better off. Let us call this the “same-preference” principle (Fleurbaey et al., 2009):

**Same-Preference Principle:** If $R_i^a = R_j^a$ and $P_i^a P_j^a$, then $i$ is better off than $j$.

But it may happen that these individuals use different calibration strategies, so that Abel actually has a lower $\xi_i(t_i)$. Trusting the satisfaction score would betray the concerned individuals’ views about the comparison. The violation of the same-preference principle is the formal way to capture this problem.

This is just a more precise formulation of the problem raised in the earlier discussion of the Easterlin paradox. The object of people’s concern is what happens in their life, not how they calibrate the answer to fit their evaluation in a bounded scale. To be fully precise, people do care about the calibration and may engineer it so as to enhance their mood, as explained in subsection 5.1.3, or to appear honorable in front of an interviewer—this is the social consideration appearing at the end of Schwarz and Strack’s model. But they do not care about it so much that a greater $\xi_i(t_i)$ is generally more important than a better $t_i$, when they have to choose between the two.

Are there approaches that satisfy the same-preference principle? The equivalence approach, epitomized in the equivalent income indicators for instance, does satisfy it. When two individuals have the same preferences, necessarily the one with a better life has a better equivalent life in the reference set, for example, a greater equivalent income. Intuitively, this is because, as they have the same preferences, the one with a better life has an indifference curve that is everywhere higher. This property carries over to the multireference equivalent income that is computed on the basis of several reference situations by taking the average (or generalized average, or lowest) value of the equivalent incomes obtained for each of them.\(^8\)

Another example, closer to what has been studied in this section, is a measure that would rely on a fixed calibration, that is, on a given measure $\mu_d$ can also satisfy the same-
preference property. When two individuals have the same $R^*$ and $f^i R^*_j$, then $\mu^0(L(f_j; R^*)) \mu^0(L(f_j; R^*))$ provided that $\mu^0$ is fine enough to be able to distinguish the two sets $L(f_j; R^*) \supset L(f_j; R^*)$.

Interestingly, a measure $\mu^0(L(f_j; R^*))$ is not very different, in its formal structure, from a multireference equivalent income. Indeed, the value of $\mu^0(L(f_j; R^*))$ is a measure of the lower contour set $L(f_j; R^*)$. Now, assume that the model of this chapter can be linked to that of the previous chapter by setting $f^* = (x, y)$ and letting $u_1$ be a utility representation of $R^*$. For instance, the average equivalent income

$$\int_{(p, y) \in \mathcal{P}(p, y, u(f_j))} f(p, y) d(p, y)$$

is a measure of the lower contour set of situations $(m, p, y)$ such that

$$v(i, m, p, y) \leq u(f_j)$$

(This set is the lower contour set in figure 4.8, for instance.)

There is one technical difference. The probability measure $\mu^0$ is supposed to be bounded, whereas the equivalent income is more like a measure of volume of the set and is therefore unbounded. The former is a relative magnitude; the latter is expressed in monetary units. This difference is not essential because one could easily normalize equivalent incomes to make (p.193) them relative, or multiply $\mu^0(L(f_j; R^*))$ by some conventional magnitude in order to give an absolute meaning. However, the fact that life is intrinsically open-ended may be a reason to prefer an open-ended measure like the equivalent income to a bounded measure. In particular, it may be easier with a measure that is expressed in the natural units of a life dimension to think about inequality aversion than with a measure that has no units and already contains a reference to a distribution. This is an interesting issue that may deserve further analysis.
5.2.5 Comparisons across preferences

It is more important, with respect to the example of \( \mu_0(L(t^*_i; R^*_i)) \), to notice that there are not just one, but two important issues with ordinary subjective scores \( \mu_i(L(t^*_i; R^*_i)) \).

We have insisted in this chapter on one of these issues, namely, tracking what people care about, which is hampered by the heterogeneity in the calibrations \( \mu_i \). Such heterogeneity generates violations of the same-preference principle and makes it problematic to respect the views of individuals sharing the same preferences and values in life. This problem is indeed alleviated when one takes a uniform calibration measure \( \mu_0 \).

The other important issue, which was highlighted in the classical debate on welfarism (subsection 5.1.2), is that comparisons across individuals with different preferences would not necessarily be well performed even if all individuals relied, in their expressions of subjective well-being, on the same calibration measure \( \mu_0 \). Such a measure would correctly compare individuals with identical preferences \( R^*_i \), but it might fail to suitably compare individuals with different preferences. As advocated by Rawls, (1982), Dworkin, (2000), and even Bergson, (1954) earlier on, such comparisons require fairness principles, like the restricted dominance principle or similar notions as found in the theory of fair social orderings.

This additional issue reveals how difficult it may be for subjective data to provide good proxies of well-being for social evaluation. While the problem of heterogeneous calibrations may perhaps be roughly eliminated by the law of large numbers when one looks at average statistics, the estimation of the distribution of well-being for a population with heterogeneous preferences about life is much more demanding. It appears heroic to hope that spontaneous answers to subjective well-being questionnaires could directly deliver data for ethically reasonable interpersonal comparisons.

For instance, Fleurbaey, Schokkaert, and Decancq, (2009) have compared the populations of the worst off (lowest quintile) in terms of satisfaction and the worst off in terms of equivalent income in a Russian sample and (p.194) found that these populations differ widely—much more than when
different reference parameters are adopted for the computation of equivalent incomes. Specifically, the worst off in subjective terms include people who are quite well off on all objective counts, even if they are not all millionaires.

5.3 Making use of happiness data
Equipped with the conceptual apparatus of the previous section, we can examine the various uses of subjective well-being data that have been proposed and explore various possibilities.

5.3.1 Proposed indicators
There is first the simple average score (of satisfaction or happiness), which is the most common statistic in the literature:

\[ \frac{1}{n} \sum_{i} \mu_i (L(\gamma_i; R^*_i)) \].

Such an indicator assumes that the individual scores \( \mu_i (L(\gamma_i; R^*_i)) \) are cardinally measurable and interpersonally comparable, and that social welfare increases when this average number increases. These two assumptions are distinct. The former is justified if \( \mu_i (L(\gamma_i; R^*_i)) = \mu_0 (L(\gamma_i; R^*_i)) + \epsilon_i \), where \( \mu_0 (L(\gamma_i; R^*_i)) \) is a cardinally meaningful measure and \( \frac{1}{n} \sum_{i} \epsilon_i \) does not fluctuate across measurement operations.

None of these assumptions is plausible. Variations in scores are not directly comparable across individuals when heterogeneity in calibrations involves more than a fixed individual shift; comparisons across populations (e.g., countries) are problematic when \( \frac{1}{n} \sum \epsilon_i \) varies with culture; and summing scores ignores inequality.

Introducing inequality aversion would require computing a generalized mean

\[ \frac{1}{n} \sum_{i} \varphi(\mu_i (L(\gamma_i; R^*_i))) \],

with a concave transform \( \varphi \). Unfortunately, the informational demands on individuals scores \( \mu_i (L(\gamma_i; R^*_i)) \) are then even greater, as they must be fully measurable and comparable.

(p.195) There have been apparently less demanding uses of these data. The Gallup-Healthways Well-Being Index focuses on the percentage of respondents in various levels, considering in particular that on the 0–10 ladder, the
respondents at 4 or below are “suffering” while those at 5-6 are “struggling.” Such an approach is also advocated by Hammond, Liberini, and Proto, (2011). This requires that individual score levels, rather than differences, be interpersonally comparable, and, for applications to social welfare, it also requires the additional ethical assumption that individuals with lower scores should be considered worse off than individuals with higher scores. Moreover, seeking to minimize the proportion of the population below a certain score would involve a form of maximin criterion with some aggregation at the bottom.

This approach is not compatible with the assumption

\[ \mu_i(L(t_i; R_i^f)) = \mu_0(L(t_i; R_i^f)) + \epsilon_i \]

and it essentially requires that for all \( i \),

\[ \mu_i(L(t_i; R_i^f)) = \mu_0(L(t_i; R_i^f)) \]

but without the assumption that

\[ \mu_0(L(t_i^f; R_i^f)) \]

is a cardinally meaningful measure. As suggested by the study of the worst-off population in Fleurbaey, Schokkaert, and Decancq, (2009), it is not very likely that this approach suitably identifies the subgroup of the population that deserves the most priority in social policy.

Kahneman and Krueger, (2006) have proposed a more parsimonious indicator, the U-index, which bears on emotions and records the average proportion of the day that people spend in a mood in which negative emotions predominate. The distinction between positive and negative emotions is perhaps the least controversial aspect of subjective well-being for comparisons across individuals. However, some interpersonal variations may occur in how people identify the dominant emotions (some may be more sensitive to negative emotions than others in their evaluation, in a way that makes comparisons difficult). Nevertheless, one can probably be more hopeful about coming up with interpersonally comparable indicators for emotions than for satisfaction judgments.

In conclusion, it appears that happiness and satisfaction scores cannot be used directly as representations of “the final outcome about which people are most concerned,” because they involve calibrations that are heterogeneous in space and time and may not reflect ethical priorities. Measures of emotions may be more promising on the measurement side, but they cannot be taken to measure “the final outcome about
which people are most concerned,” because they represent only a part of what matters in life for the individuals who are not pure hedonists.

The fact that subjective well-being data do not deliver the desired output directly does not mean that they do not provide anything useful. If one would like to measure well-being on the basis of individual preferences, for instance with a measure like the equivalent income or like $\mu_{e}(L(R_i^t, R_i^t'))$, such data are useful if it is possible to filter the calibration variations and retrieve information about individual preferences.

In the remainder of this section, we briefly examine how to devise a strategy for the extraction of useful information from data on subjective well-being.

5.3.2 Putting affects in their place

Regarding affects, it is easy to conclude from what precedes that they should be treated as a life domain, along other domains like income or health. In fact, the boundary between emotions and mental health is fuzzy, which reinforces the idea that they belong to a domain of this sort.

The interesting research question that emerges from this perspective is to find out how much weight people give to this dimension of life in their global evaluation. This is not rendered easy by the correlation between emotions and other dimensions. Success in other dimensions produces positive emotions.

In a different vein, Kimball and Willis, (2006) propose to use emotions as an indicator of preferences in a clever way. It is argued in psychology (Kahneman and Thaler, 2006) that emotions are carried by events more than by states, that is, by changes rather than by stable situations. Suppose that an emotion score then reflects the change in satisfaction recorded in the most recent period. If one can regress an emotion score on the change in the vector of life dimensions that has occurred recently, one can interpret the relative weight of the various dimensions in such a regression as the relative weights they have in people’s preferences.
The difficulty of this approach is to find data that have the right temporal pattern, and to be confident that emotions reflect variations in satisfaction. The risk is that emotions reflect changes in life domains that occur separately at different times, without a clear link between the intensity of the emotion and the weight of the life domain in the respondent’s preferences. One may be happy after finding a job, then happy after marriage, then happy after having a child, but it may be very hard to retrieve weights from such a sequence.

5.3.3 Identification problems

With satisfaction and happiness scores, the estimation of preferences over life dimensions is rendered particularly arduous by identification problems. An important characteristic influencing responses is the respondent’s personality. Few data sets have information about personality, and it is often thought that panel data are preferable because they make it possible to isolate individual fixed effects. This, however, works well only if personality influences the answers only by moving them up or down by a constant (or by moving the probability of answering in a high category by a constant). If personality exacerbates or attenuates the variations of the score over time, this is not eliminated by fixed effects, and the typical regression will give a greater importance, in the estimation, to the individuals with a “sensitive” personality over the others.

Panel data have the drawback of eliminating not only fixed characteristics that appear irrelevant such as personality, but also the fixed characteristics that are objects of preference. A disability that is fixed over the waves of the survey and reduces well-being will be assimilated to a particularly gloomy personality. Only the people acquiring this disability over the survey will have a role in the estimation of its importance. And the individual characteristics that are fixed over one’s life but may nevertheless have an impact on satisfaction, as objects of preferences or as influences on objects of preferences that are not recorded in the data, necessarily go unnoticed. One may think of social or ethnic background as examples of such characteristics.

The calibration problem also makes it very hard to determine whether relative variables have an impact as objects of
preferences or are only due to the calibration. Consider Abel, whose preferences are not affected by comparisons with others, and are represented by the utility function

\[ a_0 + a_1 \ln x_1 + a_2 x_2 \]

It may nevertheless happen that Abel, on a questionnaire with a bounded scale, gives a satisfaction score that reflects the quantile of this expression in the uniform distribution over some interval \([\bar{u} - A, \bar{u} + A]\), where \(\bar{u}\) is the average value of \(a_0 + a_1 \ln x_1 + a_2 x_2\) in a reference group and \(A\) is a constant. His satisfaction score is then equal to

\[ \frac{a_0 + a_1 \ln x_1 + a_2 x_2 - \bar{u} + A}{2A} \]

Compare Bael, who cares primarily about his relative position in the distribution of each good, and whose preferences are representable by the utility function

\[ b_0 + b_1 \ln \frac{x_1}{\bar{x}_1} + b_2 \left( x_2 - \bar{x}_2 \right) \]

(p.198) where \(\bar{x}_1\) is the geometric mean and \(\bar{x}_2\) the arithmetic mean of these variables in a reference group. Because his preferences are already about relative magnitudes, Bael uses a simple affine transformation in order to calibrate his satisfaction grades for the questionnaire.

With satisfaction data, there is no way in which Abel, who does not care about comparisons at all in his preferences about life, can be distinguished from Bael, whose sole concern is his relative position with respect to the reference group. This example shows that it is very hard to derive from standard happiness and satisfaction questionnaires reliable information about how much people really care about comparisons to others. The same problem affects any estimation of adaptation over time. People who care mostly about the evolution of their situation are hard to distinguish from people who care about their current situation but look at the past to calibrate their answers.

In other words, the data may tell us more about the calibration strategies of the respondents than about their true preferences and values in life. Importantly, this does not mean that the distinction between preferences and calibration is metaphysical and cannot be observed in any way. It is rather easy to track preferences by making the respondents face trade-offs between improvements in their absolute situation and improvements in their relative position. This requires
different questionnaires that ask directly about preferences over alternative options rather than about grading the current situation.

Tversky and Griffin, (1991) and Benjamin et al., (2012) have explored such trade-offs in questionnaires about stated preferences. Moreover, they have asked respondents to estimate their satisfaction or happiness level with the various options. In a majority of cases respondents prefer the improvement in their absolute situation, even if some of them forecast that their happiness would be higher with an improvement in their relative position. The latter fact may be attributable to various possible factors. Respondents may not care that much about their relative position but forecast that their satisfaction, which is relative, will evolve otherwise. Or they may understand the question about future happiness as referring to feelings and know that their feelings are strongly influenced by their relative position. Or they may choose the absolute improvement because they care about the good externalities it has on their relatives, but their personal concern is about their relative position.

Coming back to happiness data, it may be reassuring to see that, whether people are more like Abel or more like Bael, if the average values for the reference groups do not vary much, the estimation of the coefficients is not necessarily biased. What is difficult is to know whether respondents care about absolute or relative values of the variables. But for personal trade-offs about the absolute values of these variables, the estimation is indicative of the relative weight that respondents assign to them and of the choices they would make if they had to choose.

In other words, ordinary estimations with happiness data do not tell us much about problems like the Easterlin paradox, because shifting calibrations are not distinguishable from the presence of relative variables in preferences, but they do tell enough about personal indifference curves to know if people give a great weight to health or to employment, as opposed to income. They may therefore be useful to estimate the indifference curves that are needed to compute equivalent incomes or similar measures of well-being. They must, however, be supplemented with additional data in order to know whether people care about comparisons and how much.
Another consideration that must be borne in mind is that many variables in happiness or satisfaction regressions are not truly exogenous. For instance, income may be influenced by personality, which also influences happiness. The influence of income on happiness may then be overestimated. Marital status may be influenced by happiness as much as happiness is influenced by it. Panel data and natural experiments are generally thought to provide partial solutions to this difficulty. Ideally, one would need a structural model of the choices made by the individuals and of the various influences of such choices before one starts deriving equations to be estimated. So far, the structural approach has not been tried much. It is obviously quite hard to develop a realistic structural model in a model with multiple life dimensions.
5.3.4 Can happiness data be improved?

There are two doctrines about the design of happiness questionnaires. One doctrine says that the question must be as little prepared as possible and as vague as possible, so that respondents are left free to flesh out the notion of happiness with whatever they care about and are submitted to little framing by preparation questions. The opposite doctrine is that the question should be as precise as possible, leaving no ambiguity about the target concept, and come after a series of questions helping the respondent to recap the various relevant aspects of her life.

The former doctrine is convincing if happiness for herself is something that each respondent has a clear idea of and can give an answer right away. The latter doctrine fits better with the observation that the evaluation of one’s life is a cognitively demanding exercise and is not something that people have in mind, ready to be uttered. As argued by Schwarz and Strack, (1999, p.200) the answer to a satisfaction question is actually constructed on the spot, with the various elements at the respondent’s disposal.

Answering a satisfaction question is not a completely novel exercise for most respondents. People generally have a rough idea about whether their life is going well. They also have rough preferences about the contents of their life, as they are used to making decisions repeatedly, which forces them to think about what they want. Even in life domains over which little choice is possible, the human mind is accustomed to fantasize and imagine how nice or bad a different situation would be. Finally, people make comparisons to others, or reference to their aspirations, sufficiently often so that the exercise of calibrating answers by reference to a distribution is not a totally new exercise. Nevertheless, it is hard to deny that helping the respondents with sufficient time and occasions to think about the various life domains before giving a global evaluation seems the most reasonable way to proceed.

It would be particularly interesting to see if the quality of answers varies with the kind of questionnaire that is administered. The test-retest robustness of happiness data is rather weak compared to more objective data (Krueger and Schkade, 2008), and it would be natural to seek to determine if
giving more time and preparation to respondents improves the test-retest quality.

Another direction of potential improvement has to do with calibration. As in Steffel and Oppenheimer, (2009), it is possible to ask respondents how they construct their answer and what elements of comparison, if any, they use. A more radical move is to guide the calibration by asking the respondents to refer to a particular standard. One can also control the calibration by the use of vignettes. That is, one can describe a list of lives and indicate what grades they should have, or in what category they fall. This may be the most promising route if one hopes to obtain happiness data that have an absolute value in spite of the bounded scale. It is then sufficient to keep the vignettes unchanged across waves of the survey. This may be untenable over a long period of time because changes in techniques and vocabulary may render the descriptions outdated. Imagine vignettes conceived in the seventeenth century and used nowadays. A good life would probably involve a noble status, many servants, and feats of arms. But for one century static vignettes may be possible. They would make it possible to test the true meaning of the Easterlin paradox.

(p.201) Similarly, the scope problem can be alleviated if the question makes it precise what part of life is relevant. The most interesting question is probably about the whole life rather than the current period, but this may depend on the purpose of the survey. Asking two questions, one about the whole life and one about the current period, might be the best strategy. It would also be useful to somehow specify the boundaries of “personal” life, as opposed to the rest of society and the whole universe.

Finally, happiness data can be supplemented with direct questions on preferences. Asking such questions would make it possible to check if the relative weights of life domains obtained through satisfaction scores are similar to what is obtained directly from stated preferences.

5.4 Conclusion
Happiness data provide an invaluable source of information on the population’s preferences and values in life. While there remains a great potential for improving the quality and meaningfulness of the data, the great achievement of this field
is the methodological recognition of verbal expressions as relevant for the evaluation of welfare. Such data make it possible to go beyond the very limited domain of preferences revealed by choices, epitomized by the classical economic analysis of market demand. Looking back at chapter 3, one can appreciate the progress made in seeking to widen the scope of well-being considerations beyond the narrow economic aspects of life.

However, happiness data have often been offered as the promise of a direct measure of well-being. “These kinds of statistics are probably the only ones available to us if we wish to measure well-being,” writes Oswald, (1997, p. 1827). “Self-reported happiness has turned out to be the best indicator of happiness. ... It exhibits considerable intrapersonal stability and interpersonal comparability and therefore can be used without major problems for many purposes,” write Frey and Stutzer, (2002, pp. 26 and 28).

As we have seen in this chapter, the calibration problem makes it impossible to use such data directly. Doing so is likely to betray people’s actual concerns with the contents of their life, as the Easterlin paradox—if confirmed—suggests. It is very implausible that people are indifferent to the growth in living standards and longevity that has been witnessed in the last decades. A stability in the satisfaction curves would reveal more about the shifting calibrations of the answers than about how people really feel about the issue.

(p.202) Some authors in the literature have seen that there is an apparent contradiction between preferences and subjective well-being scores and that this cannot be simply attributed to mistakes in “decision utility.” “Obviously, everyone would choose to be healthy rather than paraplegic, and rich rather than poor. But it is not obvious how to demonstrate that the rich are actually happier than the poor if both groups report the same level of well-being. At the same time, it is clear that an adequate measure of well-being must distinguish between rich and poor, and between paraplegic and quadriplegic” (Tversky and Griffin, 1991, p. 724).

“Assuming that future research provides a deeper understanding of hedonic adaptation, is it likely that such information would cause people to conduct their lives differently? Would they stop wearing seatbelts with the assurance that they would get used to being paralyzed? Would
they exploit an embezzlement opportunity knowing that prison wouldn’t be all that bad in the long run? We suspect not” (Frederick and Loewenstein, 1999, p. 320).

As we have seen, the problem appears simple once one understands that subjective well-being indicators give information either about a calibrated \( f' \), or about feelings \( f'' \) (a part of \( i \)), whereas people actually care about \( i \).

The fascination of the happiness literature for the magnitude of subjective well-being indicators may come from the meeting of two disciplines by coincidence. Psychology seeks to understand mental states and behaviors, and emotions as well as life evaluations are mental states that are certainly worth studying and explaining. Economics has been constructed around the concept of utility, and it is apparently easy for economists to forget that when the economic model makes the individual maximize \( u(x) \), this means that the individual cares about \( x \), not \( u(x) \). If individuals cared about \( u(x) \), they would spend their time working on their mind-set rather than changing the world around them.

The simple model that has been proposed in this chapter to describe the mental operations that respondents have to perform when they face a happiness question has shown that the calibration problem is pervasive, making it hard to compare individuals who use different calibration strategies, and making it impossible, for instance, to identify from such data how much the individuals actually care about their relative position versus their absolute situation.

The calibration problem comes from the gap between the open-ended nature of many dimensions of life and the bounded scale imposed by questionnaires. The violation of the “same-preference” principle shows how, precisely, this problem may make the measure go against the concerned people’s views about their own situations.

(p.203) We have seen that taking a fixed calibration would eliminate the problem. But, as explained in this chapter, it would not eliminate the difficulty of making ethically sound comparisons between individuals with different preferences about the dimensions of life. Interpersonal comparisons involve not only respectfully tracking what people care about, but also adjudicating conflicts of interests in order to
determine who deserves greater priority in social evaluation. The capability approach, studied in the next chapter, is the prominent example of an approach that is primarily based on such considerations.

Notes:

(1) This claim is disputed by Stevenson and Wolfers (2008) and Sacks, Stevenson, and Wolfers (2010). More on this in this chapter.

(2) Hedonism is often viewed as morally questionable. Assimilating a great life with a life of pleasant feelings appears rather mediocre. Very few people want to have good feelings all the time. What people typically want, as far as feelings are concerned, is to have the appropriate feelings for the circumstances they encounter. A good life is a life with strong social ties, which imply much pleasure and shared laughter, but also their full lot of mourning and bereavement. A good life is a life of daring enterprises, which imply excitement but also fear, anxiety, and stress, all negative feelings without which courage could not be displayed (Nussbaum, 2008).

(3) Barry, (2007) compares an individual who would seek to be satisfied per se—instead of getting what he wants—to a football fan who would support whatever team is most likely to win. What kind of football fan would that be?

(4) As explained, it may have instrumental value via its effect on feelings and other achievements.

(5) This list omits the ladder-of-life question used in the Gallup World Poll. This one will be discussed later on.

(6) The equivalent income may require going far away “along” the indifference curve, and this may be cognitively demanding, too (e.g., imagining being healthy when one is chronically ill). But this is not the same problem as locating the current situation on a scale containing the best and the worst possible situations.

(7) Note that the problem does not come from the limited number of rungs on the ladder. Nor would it be solved if the ladder had no upper bound, because people would still have to generate their own understanding of the gap between two rungs.
Recall, however, that the equivalence approach has been defined for complete preference orderings. Extending it to incomplete orderings is not very complex, but it requires considering imprecise values for equivalent incomes.

Another possibility, as exemplified by Kapteyn, Smith, and Van Soest, (2009), is to ask respondents to freely grade lives described in vignettes and then use the answers to recalculate satisfaction scores afterward. These authors show that American and Dutch respondents have different grading patterns (but the difference is not a simple shift).