MONASH University
Business and Economics

Department of Economics

Issn 1441-5429

Discussion paper 50/10

A Framework of Happiness Survey and Evaluation of Gross National Happiness*

Haiou Zhou†

Abstract

Happiness surveys based on self-report can generate unreliable data due to respondents' imperfect retrospection, vulnerability to context, and arbitrariness in measuring happiness. To overcome these problems, this article proposes to incorporate a method of measuring happiness, which is developed by Ng (1996) based on Edgeworth's notion of "Just Perceivable Increment" of happiness, with the Day Reconstruction Method developed by Kahneman et al (2004a) to form a new happiness survey procedure. Distinguished from many surveys that simply ask respondents to rate their subjective wellbeing on a given scale, this happiness measuring method provide detailed instructions to help respondents determine and use their personal happiness units, which are interpersonally comparable, in measuring happiness. While the Day Reconstruction Method helps avoid the effects of imperfect retrospection and external disturbances, the proposed method of measuring happiness can help reduce the arbitrariness in self-measurement and derive accurate, coherent and interpersonally comparable happiness metrics. Therefore, data collected from such a survey can be used as a more reliable informational foundation for the evaluation of gross national happiness.

JEL Classifications: J00, J17, J18

Key words: Happiness survey, Day Reconstruction Method, Just Perceivable Increment, Wellbeing, Gross National Happiness

* This paper is supported by an Australian Research Council grant.
† Department of Economics, Monash University

1. Introduction

Developments in behavioral and experimental economics have shown that a person's choices can, at times, fail to reveal his or her subjective utility or satisfaction, which has led economists and sociologists to place more emphasis on happiness measurement based on self-reports, trying to collect happiness data that can be used as the information foundation of measuring gross national wellbeing or for policy making (e.g., Kahneman et al. 2004b; Loh & Wackernagel 2004; Ng 1996; Veenhoven 1996). However, as shown by many authors, such as Conti & Pudeney (2008), Kahneman & Krueger (2006), Kristoffersen (2010) and Ng (2008) among others, although existing studies have produced many useful results, plenty of work remains to be done to improve the accuracy, coherence and interpersonal comparability of the collected information.

The existing literature on happiness surveys has identified at least three sources of the unreliability of self-reported happiness. The first source is individuals' imperfect retrospection of their experienced happiness, which implies self-reported information about happiness over a past period can be inaccurate or biased. The second source is respondents' vulnerability to external disturbances, which suggests self-reported information may be unreliable because it is susceptible to be manipulated or affected by unexpected factors. The third source is individuals' arbitrariness of measuring and describing their happiness, which means self-reported information can be inaccurate, incoherent and low interpersonally comparable as respondents do not consistently follow a standard approach to measuring their happiness with interpersonally comparable units.

A desirable measurement of happiness must address all these problems. Kahneman et al. (2004a) proposed a happiness survey procedure – Day Reconstruction Method (DRM) – that can largely reduce the effects of imperfect retrospection and help prevent respondents from exposing to manipulation. However, DRM does not sufficiently address the problem related to individuals' arbitrariness in measuring happiness. On the other hand, Ng's (1996) conducted a happiness survey in which respondents were asked to define their own interpersonally comparable happiness units and use the units to measure their happiness. This approach reduces the respondents' arbitrariness of happiness

measurement but ignores the issues of imperfect retrospection and framing effect. A standard happiness survey procedure that can provide reliable self-reported information for further use is thus available by combining Ng's happiness measuring method with DRM.

The next two sections analyze the two methods respectively, including their advantages and disadvantages, with a main purpose to show why and how they can be combined to form a standard survey procedure to improve the quality of the data. Section 4 focuses on how to apply the new survey procedure to establish national wellbeing accounts and evaluate the gross national happiness (GNH). Section 5 provides brief concluding remarks and suggests avenues for further research.

2. DRM: advantages and disadvantages

Many happiness surveys ask respondents to give information of their subjective wellbeing of past episodes. However, as shown in Fredrickson & Kahneman (1993) and Kahneman et al. (1993), when respondents are asked to report their subjective satisfaction or other affect experience over a certain past period of time, they tend to overweight experiences that are either salient or recent, and assign little or no weight to the duration of an experience. Therefore, while the information collection method used in the survey largely relies on retrospective self-evaluation, the reported data may be biased (Kahneman & Krueger 2006).

Respondents are also frequently asked to provide information of happiness that is not in direct reference to specific daily experiences, such as the information of *global* life satisfaction. In additional to the inaccuracy caused by imperfect retrospection, responses to such questions are susceptible to circumstance and context in which problems are presented to the respondent due to various psychological mechanisms (Kahneman & Krueger 2006). The context is partly controlled by the researcher while the extent to which the researcher can control the context relies on the specific survey procedure. For example, the wording of a question can set up the frame for respondents' decision making (Tversky & Kahneman 1981); the respondents' emotion or reference can be influenced by what happened to them before answering the question or what are asked in

previous questions (Schawtz 1987), or who are the interviewers and audience in an interview (Conti & Pudney 2008), etc. While individuals' emotion and evaluation may be influenced by the context and the researcher can to some extent "design" the context, the data collected from the survey to some extent can be manipulated by the researcher.

These findings suggest that self-reported happiness or wellbeing can be reliable only if "they are reported closer to the time of, and in direct reference to, the actual experience" (Kahneman & Krueger 2006). It is thus important to avoid asking people to evaluate happiness or life satisfaction based on retrospection or related to nonspecific life experience.

To avoid the bias caused by imperfect retrospection and manipulation or unconscious disturbances by the researcher, information collection methods that aim to reflect individuals' actual daily experiences in real time and natural environments have been proposed. Examples of such methods include the Experience Sampling Method (ESM) proposed by Csikszemtmihlyi (1990) and Stone & Shiffman (1994), and the Daily Reconstruction Method (DRM) proposed by Kahneman et al. (2004a). While ESM tries to avoid potential errors caused by imperfect retrospection and the misdirection of the context by asking individuals to report their happiness at real times in daily life, DRM in contrast asks subjects to fill out a diary at the end of each day by reconstructing the events of the day and recall the beginning time, end time and their experiences during each episode on selected affect dimensions. DRM has the major advantage that it is less costly than ESM for both the researcher and the respondent. Kahneman et al. (2004a) shows empirically that data collected from ESM can be well approximated by the more practical and less costly DRM method.

The DRM proposed by Kahneman et al. (2004a) can avoid the bias caused by the imperfect retrospection and framing effects; however, the quality of the data still remains to be improved. There are mainly two reasons. First, in the DRM survey, respondents are asked to report their affect experience in different dimensions, as illustrated in Table 1, and the overall happiness or the "net affect" of an episode is simply defined as "the average of three positive adjectives (Enjoyment, warm, happy) less the average of five negative adjectives (Frustrated, depressed, angry, hassled, criticized)" (Kahneman et al

2004a). This approach, however, can decrease the reliability of the reported data. On the one hand, while an individual cannot clearly distinguish two dimensions, say, “worried” and “depressed”, from each other, the emotion can be repeatedly counted. This may lead to systematic bias in the data. On the other hand, it is arbitrary to assume that the overall happiness is “the average of three positive adjectives (Enjoyment, warm, happy) less the average of five negative adjectives”. The overall happiness may be a complicated function of these dimensions and the function may vary across individuals. Given that we cannot observe these functions, it might be better to let each respondent herself directly estimate and report the overall happiness.

Table 1. Fixed scale used in DRM

How did you *feel* during this episode?

Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.

	Not at all						Very much
Happy	0	1	2	3	4	5	6
Frustrated/annoyed	0	1	2	3	4	5	6
Depressed/blue	0	1	2	3	4	5	6
Hassled/pushed around	0	1	2	3	4	5	6
Warm/friendly	0	1	2	3	4	5	6
Angry/hostile	0	1	2	3	4	5	6
Worried/anxious	0	1	2	3	4	5	6
Enjoying myself	0	1	2	3	4	5	6
Tired	0	1	2	3	4	5	6

Source: Kahnmeman et al (2004a)

Secondly, DRM, similar to many other surveys, does not give any explicit instruction to respondents to help them define a happiness unit and use that unit consistently to measure their happiness; instead, it simply asks respondents to choose from a given set of numbers {0, 1, 2, 3, 4, 5, 6} to represent their subjective measures, by stating that 0 means “not at all” and 6 “very much” (see Table 1). Such an approach, called Fixed

Scale Method (FSM) in this paper, is too rough to reduce the arbitrariness of happiness self-evaluation and therefore it can be the source of a number of potential problems[‡].

(1) FSM can cause arbitrariness in determining the maximum happiness experiences. A fixed scale is usually a finite set of numbers or a close interval. Therefore, there is a maximum number in the set that must be used to represent the maximum happiness level, such as number "6" in Table 1. Defining an experience that can be rated at "6" and consistently using this experience or any equivalent as an example of "6" in all measurements is thus a necessary condition to ensure the accuracy and consistency of the measures. However, under FSM, the person may not have a concept that she needs to define such a happiness experiences as the reference.

Let $A=\{a_1,a_2,\cdots\}$ be a set of activities or events that an individual experienced. If an individual has an explicit happiness unit and uses this unit to measure all activities, she will be able to assign to each activity a number that represents the actual happiness level of the activity. In this process, activities that have the maximum happiness level will be represented by the largest number. However, under FSM, the respondent does not have a happiness unit in mind and which activities have the maximum happiness level is ambiguous to her. In that case, even she notes that she needs to define a reference, she may use experiences that have different happiness levels as the reference in different measures. This leads to the inaccuracy and inconsistence of the reported measures.

(2) FSM can cause informational distortion. This results from the fact that while an individual does not have a clearly defined happiness unit in mind, it is possible that she uses different units in measuring different levels of happiness. I.e., the unit that a respondent uses can rely on the happiness level that she needs to measure.

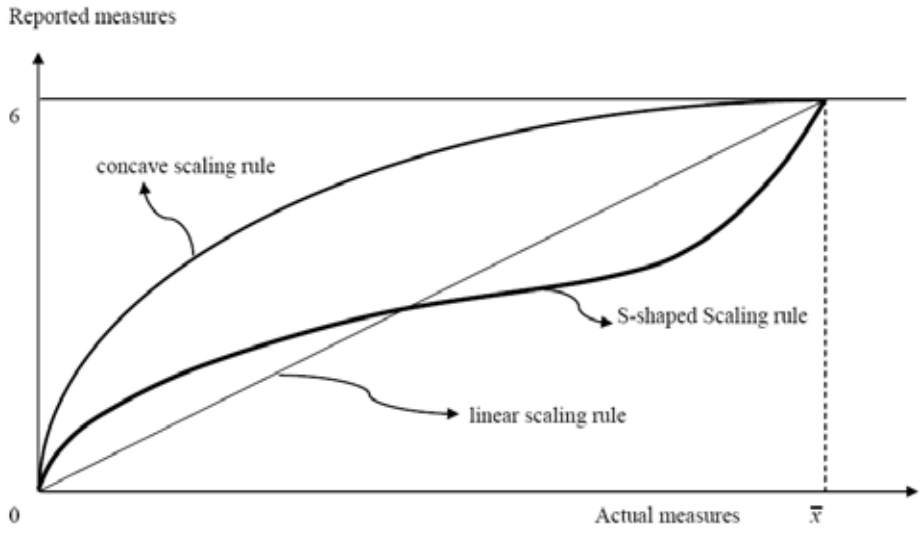
Given a set of activities $A=\{a_1,a_2,\cdots\}$ and a happiness unit, the actual happiness measure of these activities can be described by a function $\varphi: A\to S$ that assigns each $a\in A$ a real number $\varphi(a)=x$. We distinguish happiness from unhappiness; therefore the

[‡] What can be first noted is that while the respondent is forced to use discrete scales to represent her continuous subjective measures of happiness, there is a significant loss of information (Kristoffersen 2010). However, this problem might not bother researcher too much as it can be simply overcome by introducing an interval scale, such as [0,6].

minimum level of happiness will be zero. We further assume that the maximum happiness level under the given unit is $\overline{x}$. The function φ thus maps A into $X = [0, \overline{x}]$.

Assume now that the respondent is not given a unit but a fixed scale, say $S = [0, \overline{s}]$, and denote the reported happiness measure by $\psi : A \to S$, where $\psi(a)$ represents the reported happiness level of $a \in A$.The difference between ψ and φ can be characterized by a function $F : X \to S$, where $\psi(a) = F[\varphi(a)]$ for all $a \in A$. We call this function the *scaling rule*. A scaling rule can be of different forms as shown in Figure 1.

Figure 1. Three types of scaling rule

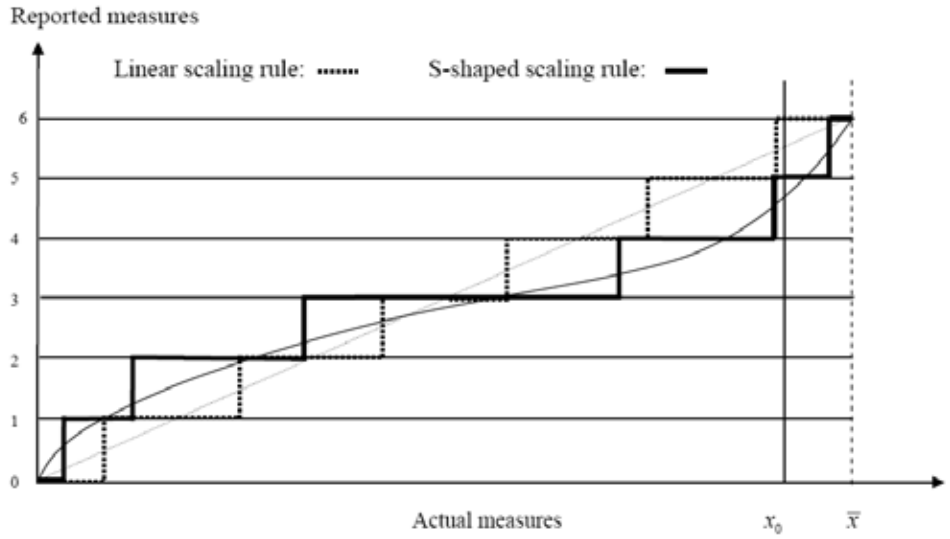


Given that the individual's happiness is a cardinal variable[§], if the reported measures are perfectly accurate, the person must correctly map the actual happiness level 0 and $\overline{x}$ into the zero point and the maximum number in the scale S respectively, i.e. $F(0) = 0$ and $F(\overline{x}) = \overline{s}$, and F must be a linear function, i.e., $\exists a > 0$, such that $F(x) = ax \in S$ for $\forall x$, where $a = \overline{s} / \overline{x}$. Such a linear scaling rule can be available only when the respondent can define a happiness unit under the given scale and use the new unit consistently in measuring all happiness levels.

[§] Cardinality of individual happiness is a necessary condition for the summation form of social happiness measurement. See Sen (1977) for a detailed discussion. In Kahneman et al (2004b), the social happiness is defined as a summation of the indiviudals' happiness obtained from DRM, which implicitly assumes that happiness is a cardinal variable.

Under the FSM, however, a respondent typically does not have a clear concept of a specified happiness unit in mind and she may fails to use the same unit consistently in measuring different happiness levels, the scaling rule will thus become nonlinear. The nonlinearity of the scaling rule implies that the reported information distorts the actually measure. For example, Ng (2008) suggests that in practice a significant number of people would use a concave scaling rule (such as arc-tangent function). A concave rule reflects the tendency that people incline to use big units to measure high happiness and implies that people tend ti overstate the actual happiness, except the zero and the maximum level.

Figure 2. Avoid using extremes in scaling



It is also possible that the scaling rule takes a form of the S-shaped function as people may use big units to measure high happiness when happiness level is low and then gradually turn to another direction by using small units to measure high happiness. The S-shaped scaling rule resembles the pi function in prospect theory (Teversky & Kahneman 1979, 1992) by implying that individuals would overstate the happiness when happiness is low and understate the happiness when it is high. It also reflects that people tend to avoid using extreme measures, such as 0 or 6 in Table 1 or 0 and 1 in probability estimation. As shown in Figure 2, the actual happiness x_0 will be rated at number 6 by a

linear scaling rule; however, when the respondent is shy away from extreme responses (Kristoffersen 2010), she will rate x_0 at 5. While various scaling rules are possible, what a type of scaling rules that most people will have is an issue that opens to empirical tests.

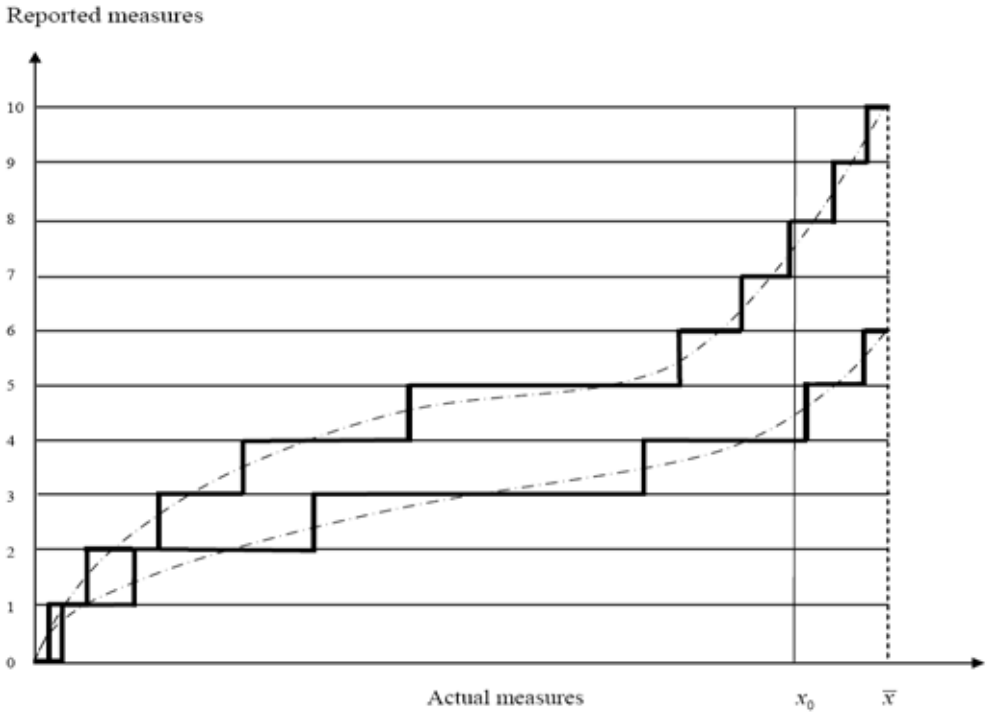
(3) FSM can cause informational inconsistence. While an individual does not have a clear concept of a specified happiness unit, it is possible that she use different scaling rules at different times. So the reported information may not have consistent welfare implications as well.

Many economists following Edgeworth (1881, p101) define an individual happiness over a given time period as the integral of the reported momentary happiness flow. However, the inconsistence of reported measures undermines the intertemporal additivity of happiness data. For example, assume that an individual's reported happiness at time t_1 and t_2 are h_1 and h_2 respectively and the number h_1 is greater than h_2. However, it is possible that the individual uses a greater unit of happiness at time t_2 and thus her happiness at t_2 is actually greater than that at t_1. It is also possible that the individual uses different scaling rules at the two points of time and the actual measures of happiness at time t_2 is greater than t_1. In each case, the sum of h_1 and h_2 is not a valid measure of the overall happiness of the two moments.

(4) FSM can be a new source of framing effect. Without a specified unit, when an individual has to use different scaling rules under different scales, the difference between the two scaling rules may not be a linear transformation. This implies that the reported information can be affected by the scale used in the survey.

For example, assume that an individual experienced a happiness level x_0 in some episode. As shown in Figure 3, when the scale is {0, 1, 2, 3, 4, 5, 6}, she to rates her happiness at 4 and with such a score her happiness may be interpreted as "quite happy"; however, when the scale is {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}, she rates it at 8 and her happiness can be interpreted as "very happy". Therefore, the scaling rule is affected by the scale. In this case, FSM constitutes a new source of framing effects and allows the survey designer to manipulate respondents' reports by choosing different scales.

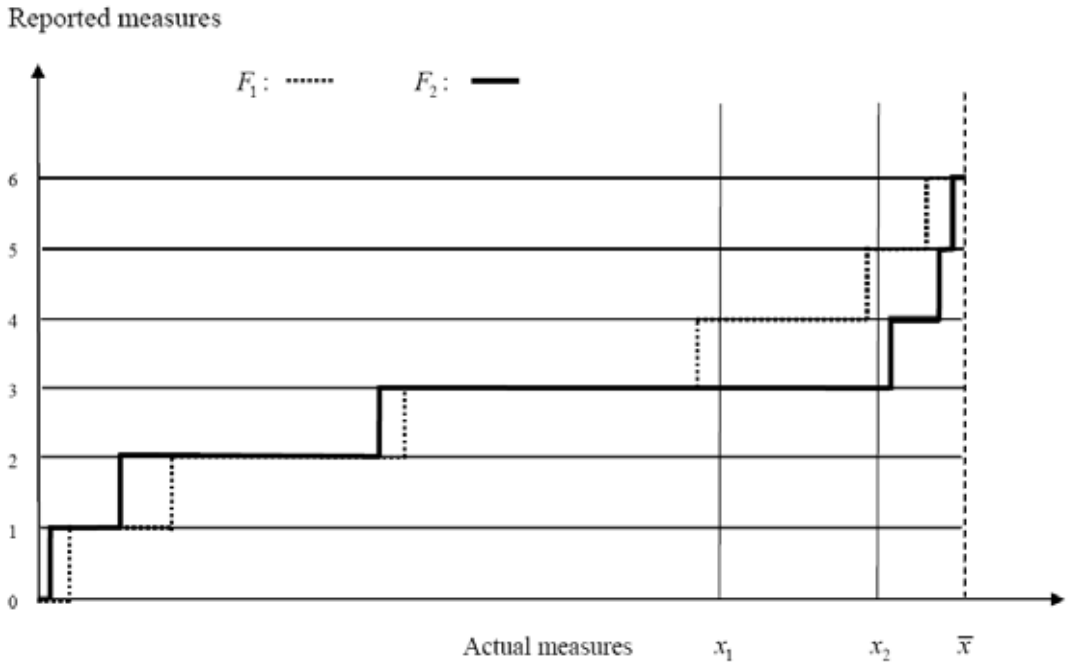
Figure 3. Framing effect of the scale



(5) FSM can reduce the accuracy of the interpersonal comparison information. While people fail to use happiness units that are independent of the happiness level, the given scale, time and context, the generated data may not correctly reflect the actual relations of interpersonal happiness comparison.

For example, assume individual 1 and individual 2 share the same maximum happiness level $\overline{x}$ and individual 2 experienced an actual happiness level x_2 that is greater than individual 1's happiness x_1. The relation $x_2 > x_1$ should be maintained when these measures are mapped into the given scale. However, as shown in Figure 4, because the two individuals use different scaling rules, the reported happiness level of individual 1 is 4 and that of individual 2 is 3. Therefore, the interpersonal happiness comparison is distorted.

Figure 4. Distortion information of interpersonal comparison caused by FSM



The analysis above shows that DRM still leaves some rooms to be desired. Especially, it does not well overcome the respondents' arbitrariness in measuring subjective wellbeing. As a result, it may derive inaccurate and inconsistent intrapersonal and interpersonal happiness information.

3. A happiness "meter" based on JPI

The weakness of FSM arises from a common source, that is, FSM does not provide respondents with a method to help respondents define their own happiness units and use these units consistently in measuring happiness. Therefore, people will not be able to ascertain the precise relative relations among different states in terms of the level of happiness, as well as to locate the numbers that reflect the relative relations.

The significance of such a method is not trivial even in measuring dimensions that

are much obvious than happiness. To see this, let us assume that you are invited to an experiment where we are asked to answer the following question:

Please take a visual observation on the room and rate "the width of the computer keyboard", "the height of the bookshelf" and "the width of the wall behind the computer" in this room with numbers between 0 and 10. You should use number 10 when you think "the item is very long, wide or high" and 0 if it has "no length at all".

In this question, a fixed scale [0, 10] is given and you are not allowed to use in the measurement a ruler or anything, such as your palm, of which you already know the length.

Although the keyboard, the bookshelf, and the wall are directly and simultaneously presented to you and the length (height/width) of an object is much more observable and perceivable than happiness, I believe most readers will still find it not easy to accurately rate the length of these items. After careful thought we can find that, first, without a ruler or some equivalent instruments it is hard to ascertain the precise relative relations among the three items in terms of length (or width or height), and second, under FSM we need to define for ourselves what is "very long, wide or high", which is typically not easy.

It may not be hard for us to just rank the three items in terms of their lengths, given that the sizes of these items differ saliently. However, to rate happiness with a number on the given scale, we need to do more precise comparisons such as "given A is longer than B and B is longer than C, is the difference in length between A and B greater or less than that between B and C" or "given A is longer than B and B is longer than C, how many times A is longer than B and how many times B is longer than C", etc. These subtle comparisons (cardinal information) are much less perceivable to respondents. To have such comparisons precisely done, we need a sufficient small unit of length that can be applied to help ascertain the subtle differences between objects.

Further, assume that after training, we are now very good at perceive the relative relations between items in terms of their lengths. As a result, we are now able to find out that height of the bookshelf is, for example, 6 times of the width of the keyboard and the width of the wall is 2.5 times of the height of the bookshelf. However, it may still be difficult for us to transfer the perceived relations into numbers on the given scale as we

need to figure out first what a length should be labelled by the number 10. The numbers that we use to represent the lengths of these objects depend on what a length is visualized by us as "very long". The determination of such a reference is arbitrary and can be affected by many irrelevant factors[**].

In the experiment above we can see that with FSM, people may confront as many problems in measuring length as in measuring happiness and the reported data reported can also be arbitrary, inaccurate and inconsistent, exactly as the self-reported happiness numbers (Kristoffersen 2010). However, in practice, we do not see too many complaints on the unreliability of length data simply because people do not use FSM but use well-developed instruments, such as a ruler, in measuring length. A ruler is a reference object that specifies a sufficiently small unit of length, such as centimeter or inch. Measuring length becomes a very easy job with a ruler. Although the arbitrariness, inaccuracy and subjectivity cannot be 100% eliminated by a ruler, however, they are reduced to an acceptable range.

Defining a unit and using an instrument to make sure that objects can be measured in terms of the same unit in various environments is the most important approach to reducing the arbitrariness, inaccuracy and inconsistence of metrics. To improve the quality of self-reported happiness data, we should also ask respondents to define their own happiness units and use these units to measure their happiness. The question facing us is how to realize this in practice.

Edgeworth (1881) first posited the development of a hedonimeter based on the notion of the "just perceivable increment (JPI)" of happiness[††]. That the "JPI" can be used as the subjective unit of happiness measurement is justified by the assumption that the welfare implication of a JPI to an individual is the same no matter where it takes places. Under this assumption, when an individual's happiness in different activities or events are represented by real numbers, these numbers must satisfy the condition that for any two activities or events x and y, the happiness difference between x and y must equate a constant real number that is independent of x and y as long as the happiness

[**] The difficulty in determining the absolute scale can also be illustrated by the fact that when a person can tell the notes, such as Do Re Mi, in hearing a song, she may feel difficulty to identify the absolute pitch of these notes.

[††] See Colander (2007) for a review on how Edgeworth's approach is related to the development of utility measurement.

difference is just perceivable. This ensures that the representation functions are cardinal, i.e., unique up to a linear transformation. [⌗⌗]

This approach is turned into practice by Ng (1996) in a questionnaire survey in which he developed a procedure to help each respondent define her personal unit of happiness and use the unit consistently in measuring subjective wellbeing. According to this procedure, the happiness unit, which is called *util*, is simply defined as the "just perceivable increment of subjective wellbeing (or happiness) over a just perceivable interval of time". Respondents then are asked to define a reference event or activity, such as drinking a cup of coffee, and specify the happiness derived from the reference event measured under the defined unit. By comparing other happiness experiences with the reference event, which now plays a role as a ruler, respondents can measure these happiness experiences of happiness using the same unit. The procedure developed in Ng (1996) well reflects Blanton and Jaccard's (2006) idea that non-arbitrary self-reported happiness data is obtained when respondents can tie specific scores on a metric to specific events that are meaningful in the life of the respondent.

All respondents in Ng's survey completed and returned the questionnaire. Most respondents answered all questions, having no difficulty in understanding the concept of JPI, define the reference event and use JPI as a unit to measure happiness related to various activities and experiences. Although the sample size is small in Ng's survey (41 respondents participated the survey), it thus provides good evidence that JPI can be understood by respondents and applied as a unit in measuring happiness.

Because Ng's (1996) survey is based on the traditional questionnaire design that involves happiness evaluation not closer to the time of, and in direct reference to, the actual experience, it may have problems of imperfect retrospection and the vulnerability to external disturbances. However, it develops a standard and applicable method that can largely reduce the arbitrariness in self-evaluation of happiness and increase the accuracy and consistency of reported data. This method can replace the FSM in DRM to improve the quality of data.

[⌗⌗] This assumption is captured by "Convention 1" in Ng (1975). According to this convention, while a lot of utility functions can be used to represent the finitely sensible preferences, only those satisfying this condition are regarded as valid representation functions. It is shown in Sichelstiel & Sollner (1996) that the existence of such utility functions can be justified by some relatively weak assumptions.

4. A standard procedure of happiness survey: from individual to society

The preceding analysis shows that a desirable happiness survey procedure to collect personal happiness information based on self-evaluation would be attained by incorporating JPI method with DRM. The combination of these two methods can at the same time ensure that self-reported happiness data will not be much influenced by respondents' imperfect retrospection, vulnerability to context, and arbitrariness in measuring happiness.

Some small modifications should be made before the two methods are combined. As mentioned in Section 2, the current DRM asks respondents to report their affect experience in different dimensions and use a simple formula to transfer these dimensions into happiness and unhappiness. If the aim of the survey is to reveal a complete picture of peoples affect experience, it may be necessary to cover all these dimensions and ask respondents to rate them separately. However, if the purpose of survey is to derive the happiness information for policy making or other uses, such as evaluation of the GNH (Kahneman et al 2004b), then a supposed formula can be arbitrary. How happiness and unhappiness as the ultimate feelings rely on these dimensions is a question that remains to be answered by further empirical studies. Before we have good knowledge about the relation between these dimensions and happiness (unhappiness), it is more effective and costless to ask respondents to directly report their happiness and unhappiness.

In Ng (1996), each respondent is asked to define her personal happiness unit as the product of two items – her just perceivable increment of happiness and just perceivable interval of time – to measure her happiness amount over a certain period of time. I think it is better to replace the second item by a fix time interval, such as 1 second or 1 minute because the just perceivable interval of time is actually redundant. When someone perceives an increment of happiness, the perceived happiness must have lasted a sufficient long time; or it will not be perceived. Therefore, when we say someone's instantaneous happiness level at time t is x, the number x already implies the happiness has lasted a perceivable time interval. There is no point to multiply x by this item again.

Now we are ready to apply the combined procedure to collect happiness information from individuals to set up a national wellbeing account. Instead of asking individuals question about global happiness in the whole year, the survey should be ideally conducted in every day among a set of randomly picked respondents, N_t, to collect the real-time wellbeing information. If the same person participated in the survey two twice, she should be regarded as two participants. Assume that an individual, i, participated the survey on the t-th day of the year, i.e., $i \in N_t$. Let h_{ij} be the length of time individual i spent in activity j, u_{ij} the average happiness amount i can have in the activity. Individual i's wellbeing or happiness in that day is $WB_t^i = \sum_j h_{ij} u_{ij}$. The daily national happiness can be defined as the "average happiness" of all individuals in that day and the annual national wellbeing or the GNH can be defined as the sum or the average of the daily national happiness of all days in a year.

One question must be addressed in calculating the GNH WB. That is, how to compare the happiness measures reported by different individuals. For example, when a person reports a happiness of 1000 utils and another reports 800 utils, the "util" used by the first person does not necessarily represent the same amount of happiness as that defined by the second. It resembles to the fact that each country has a different unit of money. Thus if one have two amounts of income, USD 100 and RMB 800, we cannot simply add the two numbers up to calculate his total income. Instead, we need an exchange rate to transform one currency into another. If we want to aggregate the happiness measures from different individuals and reported at different times into the GNH of the whole year, we need such "exchange rates" as well.

To operate, we must prescribe the util of a given respondent on a given day as the standard unit for social happiness measurement. We call this unit the standard happiness unit. Assume that we already determined the exchange rate of any respondent i's util against the standard happiness unit, where $i \in \bigcup_t N_t$, according to some approach. A measure of the daily national wellbeing is

$$WB = \frac{1}{365}\sum_{t=1}^{365}\left[\frac{1}{N_t}\sum_{i\in N_t} e_t^i WB_t^i\right].$$

where e^i be the exchange rate of i. The question is how to determine the e^i for all $i \in \bigcup_t N_t$.

Three approaches can be used in determine the exchange rates. The first is a Utilitarianism approach, which assumes that the JPIs of different individuals at different times represent the same amount of happiness; i.e., 1 util of Tom is equivalent to 1 util of Jerry[§§]. This approach implies that e^i=1 for all $i \in \bigcup_t N_t$. Although the assumption that the JPI of pleasure of all individuals are equatable was taken as granted by Edgeworth (1881, p7, p60) and accepted by some other economists (e.g., Amostrong 1951; Goodman & Markowitz 1952; Ng 1975, 1984; Sichelstiel & Sollner 1996), many readers and policy makers may find it difficult to accept and apply this assumption, especially in practice[***].

The second is an egalitarian approach that assumes the exchange rate between subjective happiness units used by any two individuals must be determined to make the amount of happiness over a given period of time that individuals can experience in the happiest states (bliss) of their life equatable. This approach is based on the idea that the value or ethical significance of the bliss is the same for every individual no matter it is measured by which happiness unit. All surveys that use FSM imply this approach as it is assumed that the maximum happiness level of every person should be represented by the same number. For example, in Table 1, Kahneman et al (2004a) actually assume that all persons' maximum happiness amounts are equal to 6.

To determine the exchange rates of each respondent's util against the standard happiness unit, the following question is enough:

[§§] See Goodman & Markwitz (1952) and "Convention 2" in Ng (1975) for a formulation of this assumption. Ng (2000, Chapter 5 and Appendix B) defended this approach by arguing that "equality in welfare weights is the right equality at the ultimate level."

[***] Sen (1970b, p92-94) criticized this approach, saying it is "an arbitrary" and "partial" assumption.

Imagine the happiest moment in your life, which could be an experienced event or what you haven't experienced but you wish to experience in the future. Fill a number here ____ to represent the amount of happiness you can have over 1 minute in that moment measured by the happiness unit (util) that you defined above.

Assume that the response of Tom, who is a respondent of the first day of the year, to this question is 1000 and that of Jerry, who can be a respondent from the same day or another day, is 500. If we use Tom's util as the standard happiness unit, then the exchange rate of Jerry's util against the standard happiness unit is 2.

The third approach, which can be called the moderate approach, is to specify a set of standardised events or activities that every person can experience in life and set the exchange rate of util between any two persons such that the total amount of happiness of these events or activities reported by one person equates to that reported by the other person when being measured by one unit. The logic behind this approach is similar to that behind Purchase Power Parity theory in determining the exchange rate between two currencies. To determine the exchange rate between currencies, we define a standard combination of products and assume that the value of this combination is the same in two countries no matter it is priced by which currency. Therefore, if the total price of the combination is 800 in terms of RMB and 100 in terms of USD, the exchange rate between RMB and USD is given by RMB 800 = USD 100. We apply this method to happiness measurement by replacing the standard combination of products by a *happiness basket* – a set of standardized activities or events – that characterizes the basic feature of life. How to construct such a happiness basket to determine the exchange rates between happiness units used by different people is an interesting and meaningful question that remains to be solved.

All these approaches assume that happiness measures are perfectly interpersonally comparable and thus allow for the summation form of the social happiness measure[†††]. It should also be noted that the GNH obtained from any of these approaches is

[†††] See Sen (1970a) for a detailed discussion on how the "degree" of interpersonal comparability influences the availability of social measurement.

internationally comparable. The international comparison is direct in the first approach. In the second or the third approach , to compare the GNH between two countries we need to further determine the exchange rate between the standard happiness units used in the two countries.

5. Concluding remarks

Happiness information derive from respondents' self-evaluation can be inaccurate and inconsistent because respondents are imperfect in retrospection, vulnerable to contexts, and arbitrary in measuring subjective wellbeing. The Day Reconstruction Method developed by Kahneman et al (2004a) provides a non-costly data collection method that can help avoid the effects of imperfect retrospection and external disturbances. Ng's (1996) provides a happiness measuring method based on the concept of "just perceivable increment" of happiness that can reduce the arbitrariness in self-measurement. These two methods complement each other and thus can be combined to form a new survey procedure. Data collected from such a survey is reliable and interpersonally comparable, and can be used as a foundation for the evaluation of GNH. Therefore, the paper provides a complete framework of establishing national wellbeing account.

Around this framework, several problems remain to be solved to improve our understanding and practice of happiness measurement. First, while individual happiness or unhappiness is a comprehensive measure of various affect dimensions, to what an extent and according to what rules human brains aggregate these dimensions into a unique metric is an important problem. Social choice theory that aims to provide some normative theory about how individuals' preferences are aggregated into social preferences (see Sen 1970, 1999 for more introductions) may be applied to this question as a normative models as well. However, the actual process of such aggregation can deviate from the normative models and descriptive models are to be clarified by future studies.

Second, while it is suggested that FSM can potentially generate inaccurate and inconsistent happiness information because it fails to provide an applicable method for respondents to precisely measure their subjective wellbeing, these suspicions are to be justified. Surveys that based on FSM and the happiness measuring method recommended by this paper needs to be conduct to help determine how different data collected from the two approaches can be and whether it is worthwhile to replace the simple FSM by the relative complicated method.

Thirdly, while "just perceivable increment" of happiness is believed to be interpersonally comparable, how exchange rates between these subjective happiness units are determined is a question that must be carefully treated. While there are various approaches to make interpersonal comparisons, it is a question open to all sociologists and policy makers. Especially, if the Purchase-Power-Parity approach is adopted, what events and activities should be included into a standardized happiness basket, will be of most significance.

References

Armstrong, W. (1951). Utility and the theory of welfare. Wxford Economic Papers, 3: 259-271.

Blanton, H. and Jaccard, J. (2006). Arbitrary Metrics in Psychology. *American Psychologist*, 61: 27–41.

Colander, D. (2007), 'Edgeworth's Hedonimeter and the Quest to Measure Utility', Journal of Economic Perspectives, 21, 215–25.

Conti, G. and Pudney, S. (2008). If You're Happy and You Know it, Clap Your Hands! Survey Design and the Analysis of Satisfaction. Available from: http://www.iser.essex. ac.uk / publications /working-papers/iser/2008-39.pdf

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. New York: Harper and Row.

Edgeworth, F. Y. (1881). *Mathematical Psychics*. London: Kegan Paul.

Fredrickson, B.L. and Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, 65:45-55.

Goodman, L. and Markowitz, H. (1952). Social welfare functions based on individual rankings,. American Journal of Sociology, 58: 257-262.

Kahneman, D., Fredrickson, D. L., Schreiber , C. A. and Redelemeier, D.A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4: 401-405.

Kahneman, D. and Krueger, A. (2006). Developments in the Measurement of Subjective Wellbeing. *Journal of Economic Perspective*, 20: 3-24.

Kahneman, D., Krueger, A., Schkade, D., Schwarz, N. and Stone, A. (2004a). A Survey Method for Characterizing Daily Life Experiences: The Day Reconstruction Method. *Science*, 306: 1776–80.

Kahneman, D., Krueger, A., Schkade, D., Schwarz, N. and Stone, A. (2004b). Toward National Wellbeing Accounts. *American Economic Review*, 94: 429-434.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47: 263–92.

Kristoffersen, I. (2010). The metrics of subjective wellbeing: cardinality, neutrality and additivity. *Economic Record*, 86: 98-123.

Loh, J. and Wackernagel, M. (editors) (2004). *The Living Planet Report 2004.* World Wildlife Fund for Nature, Gland, Switzerland.

Ng, Y.-K. (1975). Bentham or Bergson? Finite Sensibility, Utility Functions and Social Welfare Functions. *Review of Economic Studies*, 42: 545-569.

Ng, Y.-K. (1984). Expected subjective utility: is the Neumann-Morgenstern utility the same as the neoclassical's? Social Choice and Welfare, 1: 177-186.

Ng, Y.-K. (1996). Happiness surveys: some comparability issues and an exploratory survey based on just perceivable increments. *Social Indicators Research*, 38: 1-29.

Ng, Y.-K. (2000). Efficiency, Equality, and Public Policy: With a case for Higher Public Spending, London: Macmillan.

Ng, Y.-K. (2008). Happiness studies: ways to improve comparability and some public policy implications. *The Economic Record*, 84: 253-266.

Rothenberg, J. (1961). The Measurement of Social Welfare. Englewood Cliffs: Precntice-Hall.

Schwarz, N. (1987). Stimmung als Informa- tion: Untersuchungen zum Einflufl von Stimmungen auf die Bewertung des eigenen Lebens. Heidelberg: Springer Verlag.

Sen, A. K. (1970a). Interpersonal Aggregation and Partial Comparability. *Econometrica*, 38: 393-409.

Sen, A. (1970b). *Collective Choice and Social Welfare*. San Francisco :Holden-Day.

Sen, A. K. (1977). On Weights and Measures: Informational Constraints in Social Welfare Analysis. *Econometrica*, 45:1539-1572.

Sen, A. (1999). The Possibility of Social Choice. *American Economic Review*, 89: 349-378.

Sichelstiel, G. and Sollner, F. (1996). Finite sensibility and utility functions. *Social Choice and Welfare*, 13: 25-41.

Stone, A. A. and Shiffman, S. (1994). Ecological momentary assessment (EMA) in Behavioral Medicine. *Annals of Behavioral Medicine*, 16: 199-202.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211: 453-58.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5: 297–323.

Veenhoven, R.(1996). Happy life-expectancy: a comprehensive measure of quality-of-life in nations. *Social Indicators Research*, 39: 1-58.