Emotion
2007, Vol. 7, No. 1, 213–218

1528-3542/07/$12.00 DOI: 10.1037/1528-3542.7.1.213

Sun and Water: On a Modulus-Based Measurement of Happiness

Christopher K. Hsee
The University of Chicago

Judy Ningyu Tang
Shanghai Jiaotong University

Most happiness researchers use semantic differential or Likert scales to assess happiness. Such conventionally used scales are susceptible to scale renorming (interpretation of scales differently in different contexts) and can produce a specious relativism effect (e.g., rating a low-income person happier than a high-income person in situations where the low-income person is not happier). Building on related psychophysical measurements, the authors propose a simple, survey-friendly, modulus-based scale of happiness and show that it is less susceptible to specious relativism than conventional rating scales but can still catch genuine relativism (e.g., rating a low-income person to be happier than a high-income person in situations where the low-income person is indeed happier).

Keywords: measurement scale, renorming, subjective well-being, happiness

Research on happiness and subjective well-being has generated many intriguing findings, among which is that happiness is context dependent and relative (e.g., Brickman & Campbell, 1971; Easterlin, 1974, 2001; Parducci, 1995; Ubel, Loewenstein, & Jepson, 2005; see Diener et al., 2006; Hsee & Hastie, 2006, for reviews). For example, paraplegics can be nearly as happy as lottery winners (Brickman et al., 1978). However, to assess happiness, researchers commonly use bounded labeled scales, such as a 7-point semantic differential scale ranging from *very unhappy* (1) to *very happy* (7) or a three-category scale consisting of *not too satisfied* (1), *satisfied* (2), and *very satisfied* (3). These conventionally used rating scales are susceptible to a measurement bias: the tendency to renorm—that is, interpret the scale differently in different contexts. Therefore, they cannot distinguish between genuine and specious relativisms.

To illustrate, consider two individuals: one earning $15,000 a year and living in a country where most others earn only $10,000 a year and the other earning $25,000 a year and living in a country where most others earn $30,000 a year. When asked to report their happiness with their income on the 7-point scale described above, the $15,000 earner reports greater happiness than the $25,000 earner. We refer to this phenomenon as an *outcome–happiness reversal*.

There are two explanations for this reversal. One is genuine relativism: The $15,000 earner indeed feels happier than the $25,000 earner. This may arise if the two individuals compare their income with those of their respective compatriots. The other explanation is scale renorming or specious relativism: This may occur if the two individuals norm (interpret) the labels on the scale as descriptions of relative happiness between themselves and their respective compatriots. Because the $15,000 earner earns more than his compatriots and the $25,000 earner earns less than his compatriots, the former gives a higher rating, even though in reality the former does not feel happier.

Specious relativism has been noted by many researchers (e.g., Ariely & Gneezy, 2004; Baron et al., 2003; Birnbaum, 1999; Cacioppo et al., 1989; Kahneman, Ritov, & Schkade, 1999; Lacey et al., 2006; Ostrom & Upshaw, 1968; Vosgerau & Gatingnon, 2006). For example, Birnbaum (1999) reported that, in a between-subjects design, the number 9 was rated greater than the number 221. Presumably, respondents who judged 9 interpreted the rating scale as descriptions of single-digit numbers, of which 9 is great, and respondents who judged 221 interpreted the scale as descriptions of three-digit numbers, of which 221 is small.

The notion that judgments are relative is not new. What we try to highlight is that relativism can be either genuine or specious. Conventional rating scales cannot discriminate these two types of relativisms.

Proposing a Modulus-Based Scale

The present research has two objectives: (a) to develop a simple scale that can be used in paper-and-pencil surveys and can potentially reduce specious relativism (scale renorming) and (b) to demonstrate that the scale is indeed superior to conventional rating scales in avoiding scale renorming. Two clarifications are in order here. First, it is not an objective of this research to answer such questions as whether paraplegics or lottery winners are happier, or whether wealthy individuals in poor countries or poor individuals in wealthy countries are happier. If our scale proves promising in its ability to reduce scale renorming, researchers may use it to address those questions in the future. Second, our scale is designed only to minimize scale renorming and not to minimize other possible measurement biases.

In this section, first we describe our scale, then we review related literature, and finally we specify the criteria for the scale. To use the scale, respondents are first asked to rate having no feeling as 0 and to rate their happiness with a particular positive

Christopher K. Hsee, Graduate School of Business, The University of Chicago; Judy Ningyu Tang, Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai, China.

The authors thank the University of Chicago, Shanghai Jiaotong University, and CEIBS for research support or assistance.

Correspondence concerning this article should be addressed to Christopher K. Hsee, Graduate School of Business, The University of Chicago, 5807 South Woodlawn Avenue, Chicago IL 60637. E-mail: chris.hsee@chicagogsb.edu

event (e.g., drinking a cup of water after having traveled for half a day without water and feeling thirsty) as 10. The event is called a *modulus*. Then they are asked to rate their feelings with other events relative to their feeling toward the modulus. These other events are what the researcher is interested in studying and are called *targets*. If a target makes the respondents as happy as the modulus does, they should rate it 10. If a target makes them half as happy, they should rate it 5. If a target makes them twice as happy, they should rate it 20, and so forth. If a target makes the respondents unhappy, they should give it a negative rating; negative ratings are symmetrical in meaning to positive ratings.

Because our scale uses a modulus as a common yardstick and does not allow respondents to make arbitrary interpretations of the scale, we expect it to be less vulnerable to scale renorming than conventional rating scales.

Our scale is inspired by two methods in psychophysical measurement: cross-modality matching and magnitude estimation. In cross-modality matching, respondents adjust the level of one modality (e.g., brightness) to match the level of a stimulus of another modality (e.g., loudness; Stevens & Greenbaum, 1966) or to make magnitude estimates of stimuli of different modalities (Stevens & Marks, 1980). In magnitude estimation (Stevens, 1975), respondents first consider the intensity of a stimulus as a modulus (e.g., a particular light source) and rate the intensity of other stimuli of the same modality (e.g., other light sources) relative to the modulus.

Our method is also similar to scales with concrete anchors adopted by Baron et al. (2003) and Lacey et al. (2006) in judgment of health conditions and by Smith and Kendall (1963) in performance measurement (see also Nathan & Alexander, 1985; Benson, Buckley, & Hall, 1988). For example, to rate the undesirability of having bad knees, respondents were asked to rate not having bad knees as 0 and immediate death as 100 and then to rate bad knees within that range. However, unlike those scales, our scale does not have an upper bound and is intended to be a ratio scale.

In addition, we designed the modulus of our scale so that it meets the following criteria:

1. The modulus is an event which the respondents have experienced or to which they can easily relate.

2. The modulus evokes happiness rather than other sensations, so that respondents can compare their happiness from the modulus with their happiness from the target.

3. The happiness generated by the modulus is comparable in degree with the happiness generated from the target. For example, if the target is "finding a penny on the street," the modulus should not be "the best sexual experience you have ever had." If the happiness generated from the target is too small or too large compared with the happiness from the modulus, the rating is likely to be inaccurate.

4. The modulus per se is different from and incompatible with the target. For example, if the target is money, the modulus should not be money; if the target is GPA, the modulus should not be GPA. If the modulus were similar to or compatible with the target, respondents might directly compare their objective magnitudes rather than the feelings they evoke. For example, if the target is "winning $100" and the modulus is "winning $50," respondents may simply say winning $100 is twice as happy an event as winning $50, even though they may not feel that way.

If researchers intend to compare happiness between groups of individuals, the modulus should meet an additional criterion:

5. The modulus does not generate systematically different degrees of happiness between these groups. For example, when comparing happiness between girls and boys, one should not use "getting a toy train" as the modulus, because usually boys react more positively to this event than girls. Furthermore, one should not use an unidentified event such as "the happiest moment in your life" or "what you did this morning" as the modulus. These events may differ systematically between groups.

It is important to note that, whereas we advocate the use of the modulus-based scale, we do not propose the use of one universal modulus for all purposes. What modulus to adopt depends on what one intends to test. In the experiments we report in this article, we used two different moduli—drinking water when thirsty and seeing a sunny day after a week of rain. For what we intended to test, these moduli met all the criteria outlined above.

Testing the Modulus-Based Scale

Testing the superiority of the modulus-based scale over conventional rating scales is not an easy task. Unlike psychophysicists who can test the validity of a scale by fitting data to a pre-established function, such as Stevens's power function (e.g., Bartoshuk et al., 2002), we do not have in our disposal such pre-established happiness functions. In most situations, if we find ratings by the modulus-based scale to be different from ratings by a conventional scale, we cannot tell which scale is better. For example, in the rich person–poor person example introduced earlier, if using a conventional rating scale, the $15,000 earner reports greater happiness than the $25,000 earner but using a modulus-based scale, the $25,000 earner reports greater happiness, we would not know which is true because genuine relativism may or may not exist here.

Here is how we test the modulus-based scale: We ask respondents to predict the feelings of recipients of different outcomes using either the modulus-based scale or a conventional 7-point scale. These outcomes are organized in two separate sets: Set S = $[s_S, s_L]$ and Set L = $[l_S, l_L]$, where $s_S < s_L < l_S < l_L$ (< means "is worse than"). We construct the scenarios so that genuine relativism would not exist; that is, in reality, the recipient of s_L would not be happier than the recipient of l_S. We expect that, using the 7-point scale, participants will renorm within each set, and because s_L is the better outcome in Set S and l_S is the worse outcome in Set L, they will rate the recipient of s_L as happier than the recipient of l_S, but with the modulus-based scale, this effect will disappear.

Both of our studies adopted the above method, but they did so by using different contexts and different moduli.

Study 1

Method

Research participants were 195 students recruited from several classes at a large university in the United States; they completed a questionnaire in class before their midterm exam. The questionnaire had four between-subjects versions, which constituted a 2 (scale: 7 point vs. modulus based) × 2 (set: small vs. large) factorial design.

Scale manipulation. In the 7-point scale condition, respondents were instructed to use a 7-point scale on which 1 = *very unhappy* and 7 = *very happy*. In the modulus-based scale condition, respondents were instructed to use the "drinking water when thirsty" scale as we mentioned above.

Set manipulation. After receiving their respective rating instructions, respondents were asked to assume that, after the midterm exam, their professor would tell them their percentile rank in the class. They were reminded that their percentile rank could range from the 99th (best) to the 1st (worst). They were then asked to consider two specific scenarios and rate their feelings in each scenario. There were two sets of scenarios to be rated:

Set 1.

Scenario 1: "You are at the 10th percentile."

Scenario 2: "You are at the 90th percentile."

Set 2.

Scenario 1: "You are at the 91st percentile."

Scenario 2: "You are at the 99th percentile."

Half of the respondents were asked to rate the two scenarios in Set 1, and the other half rated the two scenarios in Set 2.

Predictions, Results, and Discussion

We were interested in whether the respondents exhibited an outcome–happiness reversal across the boundary of the two sets; namely, whether respondents exposed to Set 1 reported greater happiness with achieving a percentile rank of 90th than respondents exposed to Set 2 with achieving a percentile rank of 91st.

In reality, this reversal should not exist. Respondents in the Set 1 and Set 2 conditions were in the same class and were both told that the percentile rank was calculated on the basis of the entire class and could range from the 1st (worst) to the 99th (best). Therefore, they should not have felt happier with the 90th percentile than with the 91st percentile; they would either feel worse or feel the same.

However, we predicted that people using the 7-point scale would renorm their responses within their respective set and exhibit the outcome–happiness reversal and that people using the modulus-based scale would not.

The results confirmed the prediction (see Figure 1). In the 7-point scale condition, happiness ratings were significantly higher for the 90th percentile than for the 91st percentile (Ms = 6.46 vs. 5.89, respectively), $t(102) = 3.25$, $p < .0016$. In the modulus-based scale condition, the reversal disappeared ($t < 1$, *ns*).[1] An analysis of variance on z scores of the ratings secured a significant 2 (scale: 7 point vs. modulus based) × 2 (outcome: 90th vs. 91st) interaction effect, $F(1, 190) = 4.49$, $p = .036$. Compared with the 7-point scale, the modulus-based scale was indeed less susceptible to scale renorming.

Study 2

Study 2 differed from Study 1 in several aspects. First, it used a different modulus: seeing a sunny day instead of drinking water. Second, it involved a different context: finding money instead of receiving scores. Most important, Study 2 introduced another independent variable: the presence or absence of social comparison. We created our scenarios so that in the without-social-comparison condition, real relativism would not occur, and the only relativism that might occur was specious relativism, whereas in the with-social-comparison condition, real relativism would occur. We intended to demonstrate that the conventional 7-point scale could not discriminate these two types of relativisms, whereas the modulus-based scale could; namely, it could screen out specious relativism yet still capture real relativism.

Method

Participants were 124 college students from a large university in China. They answered one of four versions of a questionnaire. The four versions constituted a 2 (scale: 7 point vs. modulus based) × 2 (social comparison: with vs. without) between-subjects design. Another independent variable was set (small vs. large); it was manipulated within subjects in this study.

Scale manipulation. In the 7-point scale condition, respondents were instructed to use the same 7-point scale as in Study 1. In the modulus-based scale condition, respondents received the same instructions as in Study 1, except that they were asked to imagine the following: "It has been raining for a week. When you wake up this morning, you see a sunny day." Then they were asked to treat their feeling with that event as 10 and rate feelings with other events relative to that feeling.

Set manipulation. After reading the instructions about scales, respondents read two short stories, each describing two protagonists (two students). The two stories represented Set S and Set L, respectively. According to the first story, the two protagonists had each found a sum of money on their way home after school, and the amounts they had found were ¥1 and ¥20, respectively. (¥1 was approximately $0.12 at the time the study was conducted.) According to the second story, the two protagonists had also each found a sum of money on their way home after school, and the amounts were ¥30 and ¥500, respectively.

Social comparison. In the without-social-comparison condition respondents were told that the two protagonists in each story lived in different cities, that they did not know each other, and that they found the money while walking home alone. In the with-social-comparison condition, the respondents were told that the protagonists in each story were classmates but not friends, that they found the money while walking home together, and that each protagonist knew how much the other had found.

In each story, respondents predicted the feelings of the two protagonists using either the 7-point scale or the modulus-based scale.

[1] One respondent in the modulus-based measurement condition of Study 1 rated the 91st percentile as 900 and the 99th percentile as 1000. These ratings were much higher than the highest ratings given by other respondents, which were only 145 for the 91st percentile and 150 for the 99th percentile. We considered that respondent an outlier and excluded him/her prior to our analysis.

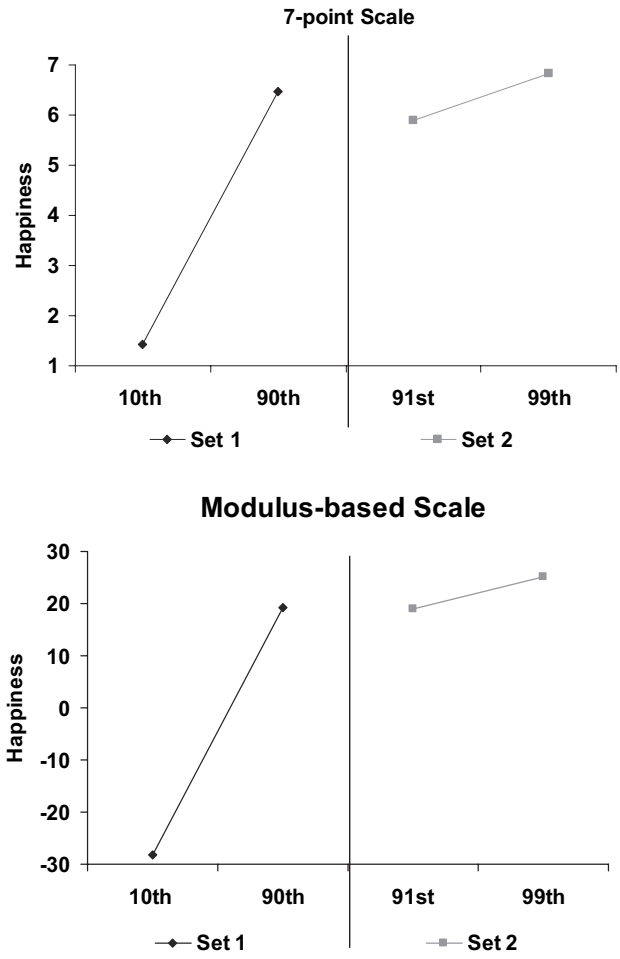

Figure 1. Study 1 results. Happiness ratings of two sets of percentile ranks (manipulated between subjects) using either a 7-point or a modulus-based scale. The modulus is drinking water when thirsty. As the graphs show, ratings using the 7-point scale exhibit an outcome–happiness reversal between the 90th percentile in Set 1 and the 91st percentile in Set 2; ratings using the modulus-based scale do not.

The reader may wonder why we asked research participants to predict the feelings of others rather than had them assume the role of one of the money finders and report their own feelings. The reason is as follows: If subjects assumed the role of one of the money finders, we had to tell them how much money the others found. If they exhibited an income–happiness reversal, we would not know whether it was due to scale renorming or genuine relativism. The advantage of our prediction method is that the subjects knew how much the other finders found, but in the without-social-comparison condition, the protagonists in the stories did not know. Therefore, if we found an income–happiness reversal in that condition, we could attribute it to specious relativism.

Predictions, Results, and Discussion

As in Study 1, our interest was in whether the respondents exhibited an outcome–happiness reversal around the boundary of the two sets (stories); namely, whether they rated the ¥20 finder in Story 1 as happier than the ¥30 finder in Story 2.

In reality, the outcome–happiness reversal was impossible in the without-social-comparison condition and highly possible in the with-social-comparison condition. In the without-social-comparison condition, each protagonist was from a different school and did not know how much others found. Thus, finding less money could not be happier than finding more. In the with-social-comparison condition, the ¥20 finder knew that her classmate found only ¥1 and the ¥30 finder knew that her classmate found ¥500. Thus, the ¥20 finder may well be happier than the ¥30 finder.

We predicted that ratings using the modulus-based scale would match the reality better than ratings using the conventional 7-point scale. This was indeed what we found (see Figure 2). When there was no social comparison, participants using the 7-point scale rated the ¥20 finder as happier than the ¥30 finder (Ms = 5.26 and 4.97, respectively), $t(34) = 1.83$, $p = .076$, but those using the modulus-based scale did not (Ms = 6.67 and 7.38, respectively), $t(28) = 1.19$, *ns*. An analysis of variance on z scores of the ratings revealed a significant 2 (scale: 7 point vs. modulus based) × 2 (outcome: ¥20 vs. ¥30) interaction effect, $F(1, 62) = 4.24$, $p = .043$. This result was a replication of the finding of Study 1.

When there was social comparison, participants in both scale conditions rated the ¥20 finder as happier than the ¥30 finder ($t > 4$, $p < .001$ in both conditions). Also, there was no Scale × Outcome interaction effect ($F < 1$).

The results suggest that the conventional rating scale could not distinguish between real and specious relativisms: It yields an outcome–happiness reversal effect regardless of whether social comparison warrants this effect. In contrast, the modulus-based scale filters out specious relativism yet still captures real relativism. It throws out the bath water yet keeps the baby.

General Discussion

Many people wonder whether happiness can be studied scientifically. To study happiness scientifically, one first needs to measure happiness accurately, but to measure happiness accurately has proven a daunting task. Scholars have identified numerous biases in happiness measurements (e.g., Schwarz & Strack, 1999; Schwarz, Strack, & Mai, 1991; Vosgerau & Gatingnon, 2006). Among them is scale renorming. This problem is particularly pronounced when using conventional rating scales such as 7-point semantic differential scales. To minimize the bias, we proposed a paper-and-pencil-friendly scale and empirically demonstrated its superiority.

Our scale is far from perfect. For example, respondents may not be able to precisely report their feelings using this scale, and the resulting ratings may entail a large error variance. Moreover, one universally applicable modulus does not exist. For instance, if one intends to compare happiness between Vancouverians and Angelenos, one could not use seeing-a-sunny-day-after-a-week-of-rain as the modulus, because seeing a sunny day is not as unusual or does not evoke as much happiness for Angelenos as for Vancouverians. Instead, one needs to adopt a modulus that evokes the same feelings between the two samples, such as kissing for the first time or seeing a long-separated friend. Unlike traditionally used 7-point scales, which are not context specific, the modulus-based

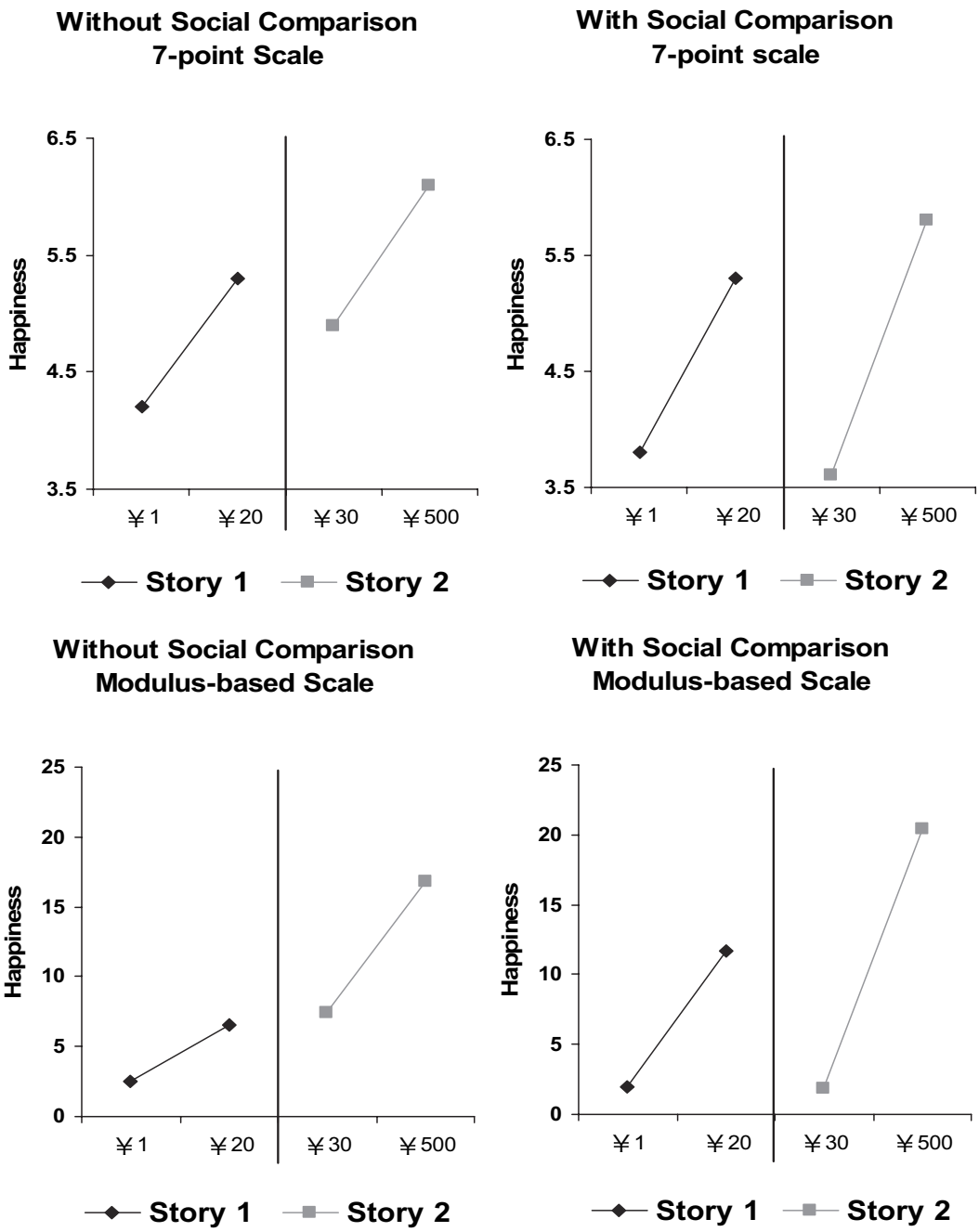

Figure 2. Study 2 results. Happiness ratings of four money finders described in two stories, using either a 7-point or a modulus-based scale, with or without social comparison. The modulus is seeing a sunny day after a week of rain. As the graphs show, in the without-social-comparison conditions, ratings using the 7-point scale exhibit an outcome–happiness reversal between the ¥20 finder in Story 1 and the ¥30 finder in Story 2, but ratings using the modulus-based scale do not; in the social-comparison conditions, ratings using either scale exhibit a reversal.

scale calls for different moduli in different contexts. This makes the scale onerous to use, but it is also precisely this context specificity that makes the modulus-based scale less prone to scale renorming, because respondents do not have the leeway to interpret the scale differently in different contexts.

Our research is not intended to undermine existing findings showing relativism in happiness judgment. Genuine relativism does exist. For example, whether using scales with vague anchors or concrete anchors, people with a particular disease tend to rate it as less severe than people without the disease (Baron et al., 2003),

and they also tend to rank it as less severe among other health conditions (Lacey et al., 2006). To distinguish genuine relativism from specious relativism in other happiness judgments, including life satisfaction and experience with income, future researchers should use scales with better defined anchors, or modulus-based scales, which the research shows are less vulnerable to scale renorming.

References

Ariely, D., & Gneezy, U. (2004). *A critique on the measurement of happiness.* Paper presented at the Society for Judgment and Decision Making Annual Conference, Minneapolis, MN.

Baron, J., Asch, D. A., Fagerlin, A. Jepson, C., Loewenstein, G., Riis, J., et al. (2003). The effect of assessment method on the discrepancy between judgments of health disorders people have and do not have: A web study. *Medical Decision Making, 23,* 422–432.

Bartoshuk, L. M., Duffy, V. B., & Fast, K. (2002). Labeled scales (e.g., category, Likert and VAS) and invalid cross-group comparisons: What we have learned in genetic variations in taste, *Food, Quality and Preference, 14,* 125–138.

Benson, P. G., Buckley, M. R., Hall, S. (1988). The impact of rating scale format on rater accuracy: An evaluation of mixed standard scale. *Journal of Management, 14,* 415–423.

Birnbaum, M. (1999). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods, 1,* 243–249.

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation-level theory: A symposium.* New York: Academic Press.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology, 37,* 917–927.

Cacioppo, J., Andersen, B. L., Turnquist, D. C., & Tassinary, L. G. (1989). Psychophysiological comparison theory: On the experience, description and assessment of signs and symptoms. *Patient Education and Counseling, 13,* 257–270.

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist, 61,* 305–314.

Easterlin, R. A. (1974). Does economic growth improve the human lot? In P. A. David & M. W. Reder (Eds.), *Nations and households in economic growth: Essays in honor of Moses Abramovitz* (pp. 89–125). New York: Academic Press.

Easterlin, R. A. (2001). Income and happiness: Towards a unified theory. *The Economic Journal, 111,* 465–484.

Hsee, C., & Hastie, R. (2006). *Hedonomics.* [Working paper.] University of Chicago, Graduate School of Business.

Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty, 19,* 203–235.

Lacey, H. P., Fagerlin, A., Loewenstein, G., Smith, D., Riis, J., & Ubel, P. A. (2006). *Are they really that happy? Exploring scale recalibration among patients' and non-patients' quality of life estimates.* Unpublished manuscript. Carnegie Mellon University.

Nathan, B. R., Alexander, R. A. (1985). The role of inferential accuracy in performance rating. *Academy of Management Review, 10,* 109–115.

Ostrom, T. M., & Upshaw, S. H. (1968). Psychological perspective and attitude change. In A.G. Greenwald, T. C. Broch, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 217–242). New York: Academic Press.

Parducci, A. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications.* Hillsdale, NJ: Erlbaum.

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Sage.

Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55,* 3–23.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* 149–155.

Stevens, J. C., & Marks, L. E. (1980). Cross modality matching functions generated by magnitude estimation. *Perception and Psychophysics, 27,* 379–389.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New York: Wiley.

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception and Psychophysics, 1,* 439–446.

Ubel, P. A., Loewenstein, G., & Jepson, C. (2005). Disability and sunshine: Can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation? *Journal of Experimental Psychology: Applied, 11,* 111–122.

Vosgerau, J., & Gatingnon, H. (2006). A new method for comparing subjective wellbeing across countries. *Association of Consumer Research Conference Proceeding.*

Received January 26, 2006
Revision received August 3, 2006
Accepted August 28, 2006 ■