

Hedonic Durability

YANG YANG

CHRISTOPHER HSEE

OLEG URMINSKY

LI ZHANG

*** Draft version: Please inform authors before citing or circulating***

4/14/2014

AUTHORS' NOTE

Yang Yang (yy1@cmu.edu) is a doctoral student in marketing at the Tepper School of Business, Carnegie Mellon University. Christopher Hsee (Christopher.Hsee@chicagobooth.edu) is Theodore O. Yntema Professor of Behavioral Science and Marketing, and Oleg Urminsky (oleg.urminsky@chicagobooth.edu) is Associate Professor of Marketing at the University of Chicago Booth School of Business. Li Zhang (lzhangah@ust.hk) is a doctoral student at Hong Kong University of Science and Technology. The authors would like to thank Jeff Galak, Tony Lian, George Loewenstein, Joachim Vosgerau, and Hanwei Wang for their valuable comments and input.

Contribution Statement

Some products are more hedonically durable (keep consumers happy for a longer time) than other products. Hedonic durability is an important variable because it can help consumers make more-informed purchase decisions and researchers gain more knowledge about the psychology of hedonic adaptation. However, there does not exist a practical instrument to measure this important variable. This research is a first attempt to develop such an instrument. Across five studies, we show that our instrument (a brief and easy-to-administer survey) is psychometrically valid and capable of assessing the hedonic durability of many products at a single occasion.

Abstract

The hedonic durability of an item refers to how fast one's happiness with the item fades over time. It is an important variable because it influences how much happiness the item can bring to the person. Yet it is difficult for researchers to assess the hedonic durability of items because it is difficult to track consumers' happiness with items over time. This paper introduces a simple survey method - the Hedonic Durability Questionnaire (HDQ) to estimate hedonic durability. We test the validity of the HDQ by demonstrating that it produces results similar to those produced by more elaborate methods (Studies 1 and 2), and results consistent with predictions derived from existing literature on hedonic adaptation (Studies 3 and 4). Finally, we apply the HDQ to a variety of items in everyday life (Study 5) and discuss implications for measuring total happiness.

An individual's happiness with most items fades over time (Brickman et al. 1978; Di Tella et al. 2010; Frederick and Loewenstein 1999; Nelson et al. 2009; Redden 2008; Wang et al. 2009). For example, a person may feel ecstatic immediately after she acquires an iPad, but she may feel apathetic about the iPad a year later.

Not only does happiness with most items fade over time, it fades at different rates. In other words, some items are more *hedonically durable* than other items are. For example, suppose that a person feels equally happy when initially acquiring a real dog and when initially acquiring a stuffed toy dog. Further suppose that a year later she still feels happy with the real dog, but is indifferent about the stuffed toy dog. Then we would say that the individual's happiness with the real dog is more hedonically durable than her happiness with the stuffed toy dog.

Understanding the hedonic durability of an item is important because, holding the initial hedonic responses constant, more hedonically-durable items yield greater total happiness over time (Kahneman 2000). The total happiness of an item can be compared to the total effect of a medicine. If two allergy medicines produce the same initial allergy-relieving effect but the effect of one medicine lasts longer, its total effect is also greater.

While people are generally mindful of the action duration of a medicine, they may not be as cognizant of the hedonic durability of an item. A key reason for the discrepancy is that people can usually find out how long the effect of a medicine lasts from its description ("12 hour allergy relief"), yet they rarely have access to information about the hedonic durability of an item. The absence of such information arises, at least in part, from the difficulty researchers have in assessing hedonic durability. Ideally, to assess the hedonic durability of an item (e.g., an iPad), researchers would need to sample consumers who have just acquired an iPad and then resample

those same consumers repeatedly over time (i.e., real time measurement). This process is often expensive and logistically infeasible, may introduce mere-measurement biases (Dholakia and Morwitz 2002; Scollon et al. 2003), and suffers from a high rate of attrition, which may lead to biased results. The task is even more daunting if researchers wish to test and compare the hedonic durability of multiple items owned by different consumers for different periods of time.

In this paper, we introduce a simple survey method to estimate hedonic durability. Researchers can easily administer the survey on the internet or in paper-and-pencil forms and can assess the hedonic durability of many items at a single occasion. The method relies on recalled experiences from individuals who already possess the items.

Several caveats are in order here. First, the current research is merely intended to introduce a method for measuring hedonic durability and not to explain why different items differ in hedonic durability (for discussions on the topic, see e.g., Frederick and Loewenstein 1999; Galak et al. 2014; Lee, Cryder, and Nowlis 2014; Nelson et al. 2009; Nicolao et al. 2009; Redden 2008; Sheldon and Lyubomirsky 2012; Smith et al. 2009; Yang and Galak 2014). Second, the purpose of this research is not to introduce a “perfect” method that can find the precise hedonic durability of an item; as mentioned earlier, a “perfect” method would require real-time tracking of consumers’ hedonic experiences with the item over the duration of their ownership of the item and is highly impractical. In contrast, the purpose of this research is to introduce a practical method that would meet at least two criteria: (1) it can assess the hedonic durability of an item in a few minutes, and (2) it can assess the hedonic durability of multiple items on a single occasion. Therefore, in evaluating the quality of the HDQ, it is important to keep these criteria in mind.

Finally, the current research is the first attempt to develop a simple method to assess hedonic durability. As such, the method is highly preliminary and may be subject to important limitations. However, we believe that this approach has significant potential, particularly given the lack of existing methods. Furthermore, we hope that the current approach will also serve as a building block in the development of future methods.

THE HEDONIC DURABILITY QUESTIONNAIRE (HDQ)

The instrument, which we call the Hedonic Durability Questionnaire (HDQ), consists of only three questions: *How long have you had X [the item to be evaluated]? How do you feel about it now? How did you feel about it when you first acquired it?* Researchers can adjust the exact wording of these questions depending on the item to be evaluated. For example, if the item is a rented apartment, the first question can be worded as “How long have you lived in X?”

Figure 1 illustrates what each question measures. The x-axis is time, where the origin represents the time at which the respondent acquires the item. The y-axis is the elicited happiness, which represents happiness with the item at a given point in time. In the HDQ, the first question assesses T (the duration from acquisition of the item till the present time, when the respondent answers the HDQ). The remaining two questions measure H_T (happiness with the item now) and H_0 (the recalled happiness with the item right after acquisition). Because hedonic adaptation has been proposed by Drew and Abbott (2006) and observed by Nicolao et al. (2009) to follow a power function, the relationship between H_0 and H_T of the item for a given respondent can be assumed as shown in equation 1. It is worth noting, however, that the proposed method could be

easily adapted to accommodate a different declining function, such as a hyperbolic or an exponential function.

$$H_T = H_0 * (1 + T)^\eta \quad (1)$$

Based on this relationship, we can solve for the parameter of interest, η :

$$\eta = \frac{\text{LN}(H_T/H_0)}{\text{LN}(1+T)} \quad (2)$$

 Insert Figure 1 about here

In the equations above, η is the power constant of the item for a given respondent, and it determines how quickly happiness changes over time. In the current research, we call η the *hedonic durability index*. If $\eta < 0$, then happiness fades over time, while $\eta = 0$ indicates that happiness does not change over time, and $\eta > 0$ indicates that happiness increases over time.

Equation 2 has several noteworthy features. First, η is a function of the ratio of H_T to H_0 , and hence independent of the unit of the scale used to measure happiness. Second, η is assumed to be independent of time and therefore η s from different respondents can be compared with each other even if these respondents have possessed the item for different lengths of time. For example, suppose that two iPad owners are asked to fill out the HDQ for their iPad's, and their

η s are η_1 and η_2 , respectively. Then, if $\eta_1 > \eta_2$, we can say that the iPad is more hedonically durable for the first respondent than for the second respondent, even if they have owned their iPads for different amounts of time. We tested and validated this time independence assumption in Studies 4 and 5.

Third, the individual η s collected from different respondents for a given item can be pooled to calculate an average η for that item. Specifically, if n respondents have answered the HDQ for a given item and their η s are $\eta_1, \eta_2, \dots, \eta_n$ respectively, then the overall η for the item is:

$$\eta = \sum_{i=1}^n \eta_i / n \quad (3)$$

In addition, researchers can use the HDQ to compare the hedonic durability of items with different initial hedonic intensities. For example, suppose that a pen elicits more intense initial hedonic responses than a mug, and the pen has the same η as the mug, say -0.5 when the unit of time is year. We can say that the pen and the mug are equally hedonically durable, because their hedonic responses taper at the same rate over time.

Moreover, we expect the HDQ to be applicable to experiences with items for any length of time: for minutes, days, months, or years. However, we should mention three qualifications. First, when asking the time question in the HDQ, researchers should use a time unit commensurate to the estimated time duration of the given item, to reduce errors in judgment (Huttenlocher et al. 1990). If the item is likely to induce happiness for only a few minutes, the time unit should be minutes; if the item is likely to induce happiness for years, the time unit should be years. Second, if researchers intend to compare the hedonic durability of multiple items, they should use the same time unit in the HDQ to assess the η s for these items, to make

the estimates comparable. As a consequence of the first two qualifications, the third qualification is that researchers should not attempt to compare the hedonic durability of items with dramatically different time horizons (e.g., an item that can induce happiness for only a few minutes versus an item that can induce happiness for years), as such comparisons are likely to be unreliable.

In the proposed formula (2), η is well-defined as long as the item both provided initial positive utility ($H_0 > 0$) and does not currently cause unhappiness ($H_T \geq 0$). However, in practice, empirical data may include negative values and/or zeros. There are at least two possible approaches to handle these values. When researchers are interested in the durability of positive feelings rather than negative feelings or change in valence (i.e., from positive to negative or from negative to positive), one approach is to exclude from data analysis any negatively valued ratings, and then add a small constant (0.01) to all the zero ratings before calculating the η for each respondent (hereinafter, the *exclusion* approach). A second approach is to exclude negative initial happiness (H_0) values, and code current happiness H_T as .01 if it was negative or zero (hereinafter, the *truncation* approach). These two approaches yield the same conclusions in all our studies. We report the results of the first approach in the main text and the results of the second approach in Appendix A.

As mentioned earlier, the purpose of this research is to introduce a simple and practical method that can provide an estimate of the hedonic durability of multiple items in a short order on a single occasion. Therefore, the HDQ must rely on recalled happiness rather than real-time happiness. However, for the method to be valid, the recalled happiness has to be similar to the corresponding real-time happiness.

As we will demonstrate in our studies, the results produced by HDQ that rely on recalled happiness are indeed quite similar to results produced by more elaborate methods that rely on real-time happiness. Therefore, based on the simplicity and the practicality criteria we set earlier, the HDQ is a sufficiently valid method. Our intention to develop a simple and practical measurement by using recalled happiness rather than real-time happiness is inspired by Kahneman et al. (2004)'s Day Reconstruction Method (DRM). The DRM is designed to measure experienced utility. Ideally, experienced utility should be measured via real-time feelings. But procedures to measure real-time feelings are costly and intrusive. The DRM instead uses recalled feelings (recollections of the previous day's experiences) to estimate one's experienced utility, and the researchers have deemed the method valid by showing that it yields similar results to more elaborate methods that rely on real-time feelings (Dockray et al. 2010; Kahneman et al. 2004; Stone et al. 2006).

In the next sections we report five studies. Studies 1-4 provide evidence validating the HDQ. Specifically, Study 1 and Study 2 showed that the HDQ, relying on recall, produced similar results to using real time measures. Studies 3-4 demonstrated that the HDQ yielded different η s in situations which, according to the existing literature, should indeed yield different levels of hedonic durability. Studies 3-4 also tested the time-independence assumption and examined the test-retest reliability of the method. Finally, Study 5 applied the HDQ to assessing the hedonic durability of a variety of commonly experienced items in real life.

STUDY 1

To measure the hedonic durability of multiple items at a single occasion, the HDQ relies on recalled (rather than real-time) initial happiness of respondents at the time of acquiring the

items. Thus, the validity of the HDQ depends to a large extent on the accuracy of the recollection. Recent research suggests that recalled experience is largely accurate (in both the short-term and the long-term) and that recalled emotions and immediately measured emotions are usually highly correlated (Kahneman et al. 2004; Levine and Safer 2002; Safer et al. 2001).

Nevertheless, previous studies have also documented systematic biases in certain kinds of recalled experiences. For example, the accuracy of recalled emotions may depend on people's implicit theories on the stability of the emotions (Ross, 1989), on personality and post event knowledge (Safer et al. 2002), and on the nature of the emotions (Aaker et al. 2008). We suggest that the existence of these biases does not mean recalled emotions are biased in general, and does not mean that these biases apply to the kinds of measures collected in the HDQ. Nevertheless, it is incumbent on us to show that the HDQ, which relies on recalled emotions, is able to produce results similar to those produced by measures that rely on real-time measures. Study 1 and Study 2 were designed to carry out this task. Study 1 did so over a short time span (minutes), where Study 2 did so over a longer time span (months).

Method

Participants. One hundred and ninety respondents (60 females, 130 males; $M_{\text{age}} = 30.39$, $SD = 10.39$) from the Amazon Mechanical Turk online panel each completed a survey in exchange for \$0.50. To ensure that participants paid attention to the instructions, we administered an instructional manipulation check (Oppenheimer et al. 2009) at the beginning of the experiment. Participants could commence with the experiment only after they had correctly answered the instructional manipulation check.

Procedure. This study consisted of two between-participants conditions: real-time measure vs. HDQ-only. In both conditions, participants watched a single 30-second video of a cat playing a shell game 5 times in a row (i.e., 5 trials). Immediately afterwards, participants in the HDQ-only condition reported how they felt about the video at that time (i.e., after the fifth trial) and recalled how they felt about the video when they had first watched the video. Participants in the real-time measure condition reported their real-time happiness with the video twice, once after the first trial and once after the fifth trial. Next, participants in the real-time measure condition completed an unrelated filler task and then recalled how they felt about the video when they first watched it. This procedure allowed us to examine whether participants in the real-time measure condition could accurately recall their happiness with the video when they first watched it, in a within-participants design. Comparing η in the HDQ-only condition to η in the real-time measure condition allowed us to examine whether the HDQ could produce similar results to those produced by real time measures (in a between-participants design).

Participants indicated their happiness on either a Likert scale (i.e., an 11-point scale, -5 = *Very Unhappy*, 5 = *Very Happy*) or a sunny-day scale (between-participants). We included both the Likert scale and a sunny-day scale in this study to test the validity of HDQ on different scales. We used the Likert scale because it is one of the most widely used scales in marketing and psychology, and we used the sunny-day scale because it has proven to be more immune to the cross-category scale recalibration problem (Hsee and Tang 2007) and may thus be more appropriate when studying multiple product categories (as in Studies 4 and 5). The sunny-day scale is a modulus-based scale, in which respondents rate their feelings about an item by comparing it with their feelings when seeing a sunny day after a week of rain (Hsee and Tang 2007). If a participant's feeling about an item is R times as strong as her feeling when seeing a sunny day after a week of rain, she should rate her feeling about the item as R . R could be greater or less than 1; if the item aroused a negative feeling in her, she should assign a negative number to R . For instructions regarding the sunny-day scale, see Appendix B.

Results and Discussion

Four participants reported initially disliking the video (i.e., gave a negative initial rating, either actual or recalled). Given that these participants did not report an initial positive hedonic experience, we cannot measure the durability of happiness, and thus the participants were excluded. An additional 43 participants gave a negative evaluation after their final viewing. Below we first reported the results of analysis which excluded these participants (i.e., the exclusion approach mentioned earlier) and then reported the results of analysis which included these participants (i.e., the truncation approach).

Exclusion approach. Following this approach, we excluded participants who gave at least one negative rating, and then added a small constant (0.01) to all the zero ratings before calculating the η for each respondent, yielding 143 participants, of which 66 rated their initial happiness. Below, we first compared the η s produced by the HDQ in the HDQ-only condition and the η s produced by real-time measure (between-subjects), and then compared the recalled initial happiness ratings in the HDQ-only condition against the initial real-time happiness ratings in the real-time measure condition (between-subjects). Finally, we examined the η s and initial happiness ratings within the real-time measure condition.

First, we calculated the η for each participant, using minutes as the unit of time and found that the η s produced by the HDQ in the HDQ-only condition and the η s produced by real-time measure were remarkably similar on both the Likert scale and the “sunny-day” scale. When the Likert scale was used, $\eta_{HDQ} = -1.39$, $SD = 2.20$; $\eta_{real-time} = -1.33$, $SD = 1.97$; $t(80) = -.12$, $p = .90$ (see Figure 2A). When the sunny day scale was used, $\eta_{HDQ} = -.98$, $SD = 1.39$; $\eta_{real-time} = -1.26$, $SD = 1.74$; $t(59) = .70$, $p = .49$ (see Figure 2B).

Second, results showed that the recalled initial happiness in the HDQ-only condition did not differ from the initial happiness in the real-time measure condition, suggesting that the HDQ is not systematically biased. When the Likert scale was used, $H_{0HDQ} = 3.21$, $SD = 1.10$; $H_{0real-time} = 3.49$, $SD = 1.27$; $t(80) = -1.06$, $p = .29$. When the sunny day scale was used, $H_{0HDQ} = 1.33$, $SD = 0.67$; $H_{0real-time} = 1.22$, $SD = 1.02$; $t(59) = .48$, $p = .63$.

Third, looking at only the *real-time measure* condition, we calculated two η s, one based on real-time measures and one based on recollection (administered after the filler questions), and found that the two η s were highly correlated, Spearman rho $r_s = .98$, $p < .001$, and did not significantly differ from each other, $\eta_{real-time} = -1.30$, $SD = 1.86$; $\eta_{recalled} = -1.33$, $SD = 1.85$; $t(65)$

= 1.19, $p = .24$. Note that we use Spearman's rho in our analyses because of skew in the η -parameter estimates and happiness ratings. Moreover, participants' recalled happiness were highly correlated with their actual happiness about the video when they first watched it ($r_s = .97$, $p < .001$), indicating high test-retest reliability at the individual level. Furthermore, the majority of participants (88%) accurately recalled how they had felt. Specifically, when the Likert scale was used, $H_{0real-time} = 3.49$, $SD = 1.27$; $H_{0recalled} = 3.59$, $SD = 1.25$; $t(38) = -1.28$, $p = .21$; when the sunny day scale was used, $H_{0real-time} = 1.22$, $SD = 1.02$; $H_{0recalled} = 1.24$, $SD = 1.01$; $t(26) = -1.00$, $p = .33$. These results reinforced our belief that the recalled happiness measure can yield similar results to real-time happiness.

Truncation approach. We replicated all of the findings reported above, using a truncation approach to deal with negative values, instead of the exclusion approach. In these analyses, we only excluded participants who reported negative happiness (actual or recalled) after the first viewing and then coded the happiness ratings after the final viewing as .01 if they were negative or zero, yielding 186 cases. For both scales, the method did not affect η scores (Likert: $\eta_{HDQ} = -2.03$, $SD = 2.46$; $\eta_{real-time} = -1.88$, $SD = 2.26$; $t(109) = -.34$, $p = .73$; sunny day scale: $\eta_{HDQ} = -1.41$, $SD = 1.72$; $\eta_{real-time} = -1.73$, $SD = 1.83$; $t(73) = .77$, $p = .44$). In the real-time measure condition, the two η s were highly correlated, $r_s = .99$, $p < .001$, and did not significantly differ from each other, $\eta_{real-time} = -1.82$, $SD = 2.10$; $\eta_{recalled} = -1.84$, $SD = 2.08$; $t(65) = 1.19$, $p = .24$. In Studies 3 and 4, we will report analyses following the exclusion approach in the main text and report analyses following the truncation approach in Appendix A. In all of the cases, the two sets of analyses yielded similar results.

Insert Figure 2A and 2B about here

STUDY 2

In Study 1, we validated the HDQ by comparing HDQ results with real-time results for experiences over a rather short period of time (several minutes). However, the accessibility of memory may decrease over time (Schacter 2002), and thus we were curious whether the results of Study 1 would extend to experiences over a longer period of time. To address this question, we conducted Study 2 by testing whether the accuracy of recall and the resulting validity of the HDQ hold over a longer span of time (months rather than minutes).

Method

Participants. 48 respondents (24 females, 24 males; $M_{\text{age}} = 20.42$, $SD = .79$) from a large university in China completed both waves. They earned 3 yuan (around \$0.50) for completing the first wave and an opportunity to win an iPod Shuffle for completing the second wave.

Wave I. In January 2013, 114 undergraduate students (57 heterosexual romantic couples; $M_{\text{age}} = 20.55$, $SD = .83$) completed a study in which each participant answered some questions and then received a gift (reported in Yang and Galak 2014). Right after participants received the gifts, they indicated their happiness with the gift on a 9-point scale (0 = *Not At All*, 8 = *Very*

Much) as well as a few other questions that were unrelated to this research. At the end, we collected the participants' contact information.

Wave II. In October 2013, we contacted all participants who completed Wave I. Of the 114 individuals who completed Wave I, we were able to collect 51 (45%) responses. The response rate was low because many participants graduated in July. In Wave II, participants indicated their current happiness (H_T) with the gift, recalled their happiness with the gift when they had first received it (H_0) on 9-point scales (0 = *Not At All*, 8 = *Very Much*), and reported whether they still possessed the gift or not. They also answered a few questions that were unrelated to this research.

Results and Discussion

We excluded one participant who did not possess the gift at the time of Wave II from our data analysis, resulting in 50 usable responses.

First, we compared the hedonic durability based on HDQ and real-time measures. Specifically, we calculated two η s (using month as the unit of time), one based on HDQ (i.e., recalled initial happiness and actual current happiness) and one based on real-time happiness (i.e., actual initial and current happiness). Results showed that the two η s were not significantly different from each other, $\eta_{HDQ} = -.01$, $SD = .168$; $\eta_{real-time} = .01$, $SD = .392$; $t(49) = .36$, $p = .72$, bootstrap $p = .81$. Bootstrapped p-values were calculated for t-tests due to unequal variances and potentially correlated errors among the participants, in this study as well as in Studies 4 and 5. In fact, scores calculated from actual or recalled initial happiness were identical for most respondents (70%). The two η s were highly correlated, $r_s = .60$, $p < .001$.

Second, we examined the extent to which participants were able to accurately recall 9 months later their happiness with the item at the time of acquisition. Results showed that participants' recalled happiness (collected during Wave II) was strongly correlated with their actual happiness with the gift at the time (collected during Wave I), $r_s = .58, p < .001$. Furthermore, recalled happiness did not differ significantly from actual happiness at the time ($H_{0real-time} = 7.52, SD = 1.36; H_{0recalled} = 7.26, SD = 1.41; t(50) = 1.22, p = .23, \text{bootstrap } p = .21$). These results suggested that people were generally accurate at recalling their happiness when receiving the gift, even after 9 months had passed.

In sum, this study extended Study 1 by testing the validity of HDQ over a 9-month period and found that HDQ yielded similar results to real-time measures.

STUDY 3

Study 3 tested whether the HDQ is sensitive enough to detect significant differences in hedonic adaptation due to factors identified in the existing literature. Prior literature has shown that individuals adapt more slowly to high-variability experiences than to low-variability experiences (Epstein et al. 2009; Redden 2008; Temple et al. 2008). In this study, we tested whether η derived from the HDQ was greater for high-variability experiences than for low-variability experiences, *ceteris paribus*.

Method

Participants. Four hundred and thirty-three respondents (151 females, 282 males; $M_{\text{age}} = 28.31$, $SD = 8.81$) from the Amazon Mechanical Turk online panel each completed the study in exchange for \$1.80. We also administered an instructional manipulation check as we did in Study 1.

Procedure. This study consisted of two between-participants conditions: the high-variability condition and the low-variability condition. Each condition consisted of five within-participants phases, which took place on five consecutive days. The stimuli used in the study were five 20-second video clips about a cat named Maru. Each video clip featured a different theme, such as Maru cleaning himself, Maru playing with boxes, and Maru being brushed. These different video clips were rated as equally attractive by a group of pre-test participants ($N = 91$, between-subjects; $F(4, 86) = 1.41$, $p = .24$, see Appendix C Table 3).

Participants in the high-variability condition watched a different one of the five clips on each of the five days of the experiment. To determine the sequence of the video clips, we randomly picked a sequence from all possible combinations, and used this sequence and its reversed sequence (in order to counterbalance the order). Specifically, half of the participants watched Maru cleaning himself on Day 1, three other videos on Days 2-4, and Maru playing with boxes on Day 5; the other half watched Maru playing with boxes on Day 1, three other videos on Days 2-4, and Maru cleaning himself on Day 5. Participants in the low-variability condition watched the same video clip five days in a row. Specifically, half of them watched Maru cleaning himself on Days 1-5, and half of them watched Maru playing with boxes on Day 1-5. Finally, on Day 5, participants in both conditions completed the HDQ by recalling how they had

felt about the video on Day 1 and how they had felt about the video on Day 5 using an 11-point scale (-5 = *Very Unhappy*, 5 = *Very Happy*).

Results and Discussion

To test whether the recalled happiness on Day 1 was biased, we compared the ratings with the reported happiness for the same video clip collected in the pretest. We found no significant difference, $F(1, 468) = .56, p = .45$, which suggested that, on average, participants' recall of how they had felt about the video they had watched five days earlier was unbiased.

To calculate η , following the exclusion approach, we excluded from data analysis 23 participants who gave at least one negative value on their rated happiness, added a small constant (0.01) to all the zero ratings, and then calculated the η for each respondent. Consistent with what the existing literature on hedonic adaption would predict, η in the high-variability condition was significantly greater than that in the low-variability condition, $\eta_{high-variability} = .09, SD = 1.11; \eta_{low-variability} = -.38, SD = 1.38; t(408) = -3.79, p < .001$ (see Figure 3). The difference was driven by the effect of variability on happiness in Day 5 ($H_{1high-variability} = 3.42, SD = 1.56; H_{1low-variability} = 2.51, SD = 1.66; t(408) = 5.74, p < .001$), while variability had no effect on recalled initial happiness ($H_{0high-variability} = 3.43, SD = 1.63; H_{0low-variability} = 3.49, SD = 1.56; t(408) = 0.40, p = .69$). The results replicated if we used the truncation approach (see Appendix A).

 Insert Figure 3 about here

STUDY 4

Whereas Studies 1-3 tested the HDQ using experimenter-provided stimuli, Study 4 tested the HDQ using naturally-occurring stimuli – items participants actually possessed in real life, including pets, stuffed toys, and items of high or low sentimental value.

Study 4 sought to achieve multiple objectives. First, this study again tested whether the HDQ is sensitive to factors that have been shown to influence hedonic adaptation. Previous research suggests that people adapt to high-variability stimuli more slowly than to low-variability stimuli (Epstein et al. 2009; Redden 2008; Temple et al. 2008) and adapt to high sentimental-value items more slowly than to low-sentimental value items (Yang and Galak 2014). To test whether the HDQ can pick up these differences, we asked each participant to complete the HDQ for four items: a high-variability item (a pet they owned), a low-variability item (a stuffed toy they owned), a high-sentimental value item that they specified, and a low-sentimental value item that they specified. If the HDQ is a sensitive instrument, based on prior research it should yield a greater η for pets than for stuffed toys, and a greater η for high-sentimental value items than for low-sentimental value items.

Second, Study 4 measured the test-retest reliability of the HDQ. Specifically, we administered the HDQ twice, with a one-week interval in-between. If the HDQ is reliable, it would yield a high correlation between scores collected during the two administrations.

Third, Study 4 empirically tested the assumption that η is independent of time. This is an important assumption, without which η s from respondents who have possessed an item for different lengths of time would not be comparable to each other. To test this assumption, we

calculated the correlations between participants' η s of an item and the duration for which they had owned the item. If η is time-independent, then there would be no significant correlations.

Finally, Study 4 tested whether people's recollections are biased over a long period of time (i.e., months or years). If people's recollection of their initial happiness is not systematically biased by the passage of time, their recalled initial happiness with an item should remain constant regardless of how long they have owned the item. In other words, there would be no significant correlation between ratings of initial happiness and length of ownership.

Method

Participants. Eighty-one respondents (31 females, 50 males; $M_{\text{age}} = 29.31$, $SD = 9.10$) from the Amazon Mechanical Turk online panel each completed the study in exchange for \$1.80. We also administered an instructional manipulation check as we did in Study 1. The study consisted of two waves.

Wave 1. Participants were asked to consider four items they owned, one at a time, and completed the HDQ for those items. The four items were a pet, a stuffed toy, an item the respondent specified as having lots of sentimental value, and an item the respondent specified as having little sentimental value. The order in which these items were presented was counterbalanced. If a respondent did not have an item, he or she was instructed to skip the HDQ for this item. If a respondent had more than one count of a particular item, he or she was instructed to randomly pick one to answer the questions. In this study, we asked participants to answer all the feeling questions in the HDQ using the sunny-day scale. We used this scale rather

than the Likert scale because this study involved items from different categories and the sunny-day scale is more immune to the cross-category scale recalibration problem that may occur when people evaluate items from different categories (Hsee and Tang 2007). Finally, as manipulation checks, we asked participants to rate the experience of owning the pet and the stuffed toy, respectively (1 = *Definitely Static*, 7 = *Definitely Dynamic*), and rate the sentimental value of the low and the high-sentimental value items they listed (1 = *Not At All*, 7 = *Very Much*.)

Wave II. A week later, we sent each participant a customized email which listed the items they had reported during the first wave of the study and invited them to participate in the second wave of the study. In the second wave, participants once again completed the HDQ for the items they had reported during the first wave. Fifty participants (12 females, 38 males; $M_{\text{age}} = 29.70$, $SD = 8.25$) completed both the first and the second waves of the study.

Results and Discussion

Sensitivity of the HDQ. We first tested the sensitivity of the HDQ by analyzing the data from the first wave of the study. Because each participant completed the HDQ for up to four items, for ease of discussion, we referred to H_0 and H_T from a given respondent for a given item as one pair of happiness ratings. Following the exclusion approach, we excluded 23 pairs of happiness ratings (10.04% of 229 pairs of happiness ratings) that contained at least one negative value and then calculated the η of each pair of happiness ratings for each respondent using year as the unit of time. The results we report below are based on the remaining 206 pairs of happiness ratings among the 81 Wave 1 participants (31 females; $M_{\text{age}} = 29.31$, $SD = 9.10$).

Results for pets vs. stuffed toys: The responses to the manipulation check question confirmed that pets were more variable than stuffed toys ($M_{pet} = 6.63$, $SD = .91$; $M_{stuffed\ toys} = 1.85$, $SD = 1.43$; $t(63) = 16.42$, $p < .001$, bootstrap $p < .001$). Therefore, according to the prior literature, pets should be more hedonically durable than stuffed toys. Consistent with this prediction, we found that the η for pets was significantly greater than the η for stuffed toys ($\eta_{pet} = .11$, $SD = .62$; $\eta_{stuffed\ toys} = -.44$, $SD = .93$; $t(63) = 2.85$, $p = .006$, bootstrap $p = .005$), suggesting that people adapted more slowly to pets than to stuffed toys. These findings corroborated Study 3 by showing that, consistent with the existing literature on hedonic adaptation, the HDQ yielded higher hedonic durability estimates for high-variability items than for low-variability items, both in a controlled setting and in a natural setting.

Results for high vs. low sentimental value items: The responses to the manipulation check question confirmed that asking participants to list high and low-sentimental value items yielded a significant difference in sentimental value ($M_{high-sentimental\ value} = 6.48$, $SD = 0.80$; $M_{low-sentimental\ value} = 2.18$, $SD = 1.79$; $t(139) = 19.11$, $p < .001$, bootstrap $p < .001$). More importantly, replicating the findings of prior research, the η for high-sentimental value items was significantly greater than the η for low-sentimental value items ($\eta_{high-sentimental\ value} = -.07$, $SD = .77$; $\eta_{low-sentimental\ value} = -2.81$, $SD = 8.10$; $t(139) = 2.99$, $p = .003$, bootstrap $p < .001$), indicating that people adapted more slowly to high-sentimental value items than to low-sentimental value items.

In summary, the results from the HDQ were consistent with the existing literature showing that high-variability stimuli are more hedonically durable than low-variability stimuli (see Figure 4A) and that sentimental stimuli are more hedonically durable than non-sentimental stimuli (see Figure 4B). The results were similar if we used the truncation approach (see Appendix A).

Test-retest reliability. To assess the test-retest reliability of the HDQ, we compared the two waves of the survey. We excluded 26 pairs of happiness ratings (17% of 150 pairs of happiness ratings) from Wave II of the study: 18 because they contained at least one negative value and 8 because the product listed in Wave II was different from Wave I. We then calculated the η of each valid pair for each respondent. One hundred and twenty-four pairs of happiness ratings from 50 participants (12 females, 38 males; $M_{\text{age}} = 29.70$, $SD = 8.25$) remained. The η from the second wave was highly correlated with the η from the first wave ($r_s = .72$, $p < .001$, bootstrap $p < .001$), suggesting that implementing the HDQ at different times produces highly consistent measures of η . Thus, the moment-to-moment variation in current happiness with the items was low enough, indicating that the measure was reliable. Moreover, the recalled initial happiness did not differ across the two waves (44% of ratings matched exactly; $H_{\text{time1}} = 2.29$, $SD = 2.36$; $H_{\text{time2}} = 2.30$, $SD = 2.54$; $t(123) = .05$, $p = .96$, bootstrap $p = .99$). The recalled initial happiness with the item at the time of acquisitions elicited during Wave I were highly correlated with those elicited during Wave II ($r_s = .73$, $p < .001$, bootstrap $p < .001$), suggesting that memory of the initial feeling remained rather constant over time.

Insert Figure 4A and 4B about here

Time Independence Assumption. To test the time independence assumption, we calculated the correlations between the respondents' η scores and the duration for which they had owned

the item, separately for each item in each wave. Supporting the assumption, none of the correlations was significant (see Appendix C Table 4). Having confirmed this assumption, η s from different respondents can be compared with each other even if these respondents have possessed the item for different lengths of time. Furthermore, the lack of correlation also justified the use of a power function to denote the relationship between H_0 and H_T . While a significant correlation between time and η could have also been evidence that the functional form used was a poor descriptor of the adaptation process, we did not find any such result.

Effect of Time on Accuracy of Recollection. To more directly test for the accuracy of recollected initial happiness based on elapsed time since acquisition, we correlated participants' initial happiness (H_0) with an item with their duration of ownership of the item, separately for each of the four items in each wave. We found none of the correlations were significant when using a Bonferroni correction for multiple tests, which suggested that participants' memory of their initial happiness was not systematically biased by the passing of time.. Only initial happiness with the low sentimentality item was positively correlated with time since purchase (based on the uncorrected p-value, see Appendix C Table 5), and this correlation was no longer significant after correcting for multiple comparisons. To confirm that that the absence of significant correlation was not due to a non-linear relationship between initial happiness and time, we also compared the recollection of initial happiness across different quartiles of length of ownership, using ANOVA for each item. Results revealed that the initial happiness did not significantly differ across quartiles overall (see Appendix C Table 6), further supporting the conclusion that the recollection was not systematically biased.

STUDY 5

The studies reported so far suggested that the HDQ was a valid and reliable measure of hedonic durability. In Study 5, we applied the HDQ to measuring the hedonic durability of a broad sample of commonly-owned consumer products in real life.

Method

Materials. We intended for this study to encompass a wide range of items that the average consumer may possess. We started with items from two sources: items listed immediately under each department on Amazon.com, and items listed immediately under each category on ConsumerReports.org. Two research assistants then reviewed all the items and removed redundant items (i.e., items that were listed in both sources) and items that were either too vague (e.g., all beauty), non-durable (e.g., movie), for one-time use only (e.g., toilet paper), or accessories (e.g., car batteries). After this step, 76 items remained. In addition, we included in the survey cats and dogs (the two most popular pets in the U.S.) as well as stuffed toys. We expected to replicate the finding in Study 4 that pets were more hedonically durable, that is, had a greater η , than stuffed toys. Together, there were 79 items in this study.

Participants. Eight hundred and seven respondents (338 Females, 469 Males; $M_{\text{age}} = 32.21$, $SD = 11.80$) from the Amazon Mechanical Turk online panel each completed this study in exchange for \$1.00. We also administered an instructional manipulation check as we did in Study 1.

Procedures. The survey consisted of four parts. The first part introduced respondents to the sunny-day scale on which to answer all the happiness questions in the HDQ for all the items in this study. The sunny-day scale was used to avoid the cross-category scale recalibration problem (Hsee and Tang 2007).

The second part of the survey was the HDQ elicitation. Each respondent was asked to answer the HDQ using the sunny-day scale for around 20 items, which were randomly selected from the 79 items. The time question was asked using months as the unit. If a respondent did not own a particular item, he or she was instructed to skip the questions about that item. If a respondent had more than one count of a particular item (e.g, two cars), he or she was instructed to answer the questions about the one that he or she interacted with the most. We excluded twenty six items which had either just been purchased (time = 0) or for which the time of ownership was not provided.

The third part of the survey solicited additional information about the items. Each item was listed again and the respondents were asked how many times they had interacted with or used the item in the last month (frequency), and how much money they had spent on the item so far (expenditure). The fourth (last) part of the survey collected demographic information from the respondents, including age, gender, language, and income. These factors did not moderate any of the results and are not discussed further.

Results and Discussion

In this study, each participant completed the HDQ for up to 20 items. For ease of discussion, we referred to H_0 and H_T from a given respondent for a given item as one pair of happiness ratings. Following the exclusion approach, we excluded 802 pairs of happiness ratings

(out of a total of 7575 pairs of happiness ratings, or 10.59%) from 376 participants because they contained at least one negative value. We then calculated the η of each remaining pair for each respondent using year as the unit of time. The results we report below are based on the remaining 6773 pairs of happiness ratings from 804 participants (336 females; $M_{\text{age}} = 32.22$, $SD = 11.81$). Figure 5 summarizes the η s of these items. As the figure shows, the η for all the items were negative, suggesting that people's happiness, on average, faded over time for all the items. Items with the highest hedonic durability indices included wedding and engagement rings, cats, religious jewelry, and dogs, presumably because they were of high sentimental value or of high variability.

Results also revealed differences in η s of items of different variability. To test the validity of the HDQ, we compared the η for cats, dogs and stuffed toys. Replicating our previous finding, planned contrasts showed that cats and dogs, which were highly variable, were indeed more hedonically durable than stuffed toys, which were not variable ($\eta_{\text{pet}} = -.13$, $SD = .82$; $\eta_{\text{toy}} = -.78$, $SD = 2.41$; $t(195) = 2.83$, $p = .005$, bootstrap $p = .007$).

 Insert Figure 5 about here

We also tested the time independence assumption separately for each item by calculating the correlation between the η s from each participant and how long they report having owned the item. Of the 79 items, only nine have η s that are significantly correlated with time, similar to the number that would be expected by chance. Using a Bonferroni correction for the multiple

significance tests, we find no significant correlations. Therefore, we find no evidence that η s are systematically associated with time since purchase.

Lastly, we investigated whether people reliably differ in their level of adaptation. To test this, we computed a split-half reliability score, by randomly splitting each person's evaluated products into two subsets, and computing the average η for each person for both lists. The correlation between the average η scores for the two subsets was $r_s = .503$ ($p < .001$, $N = 799$), suggesting reliable individual differences in the tendency for adaptation.

General Discussion

In this research, we have introduced an easy-to-implement instrument to estimate the hedonic durability of an item by sampling individuals who own the item. We have designed the method so that responses from individuals with different ownership times for a given item can be pooled together to form a composite hedonic durability index for that item, and the hedonic durability indices of different items can be compared. Using both experimentally-manipulated items and naturally-occurring items, we have demonstrated the ability of this method to produce results similar to real-time measures and consistent with existing theories on hedonic adaption.

Limitations and Future Directions

The HDQ is an initial attempt at a simple method to measure hedonic durability. As such, it requires strong assumptions. As acknowledged earlier, the HDQ utilizes recalled experiences, assuming that recollections are unbiased, but this assumption may not always hold (e.g., Pearson

et al. 1992; Ross 1989; Safer et al. 2002), and in situations where the potential for such biases exists, further research would be needed to test the validity of the HDQ specific to the setting.

Moreover, the HDQ assumes that the happiness ratings are on a ratio scale and that a rating of 2 is twice as strong as a rating of 1. Simple Likert and bounded semantic scales only provide interval measurement and do not necessarily possess this property. In particular, bounded scales can contribute to biased estimates of adaptation because of ceiling effects, particularly when initial happiness is very high. In contrast, the sunny day scale, as used in Studies 1, 4 and 5, is not bounded and does possess the ratio scale property, and therefore has better statistical properties for the HDQ. However, because the sunny day scale is relatively complicated, respondents may find it difficult to fully follow the instructions.

There may be greater individual differences in the hedonic durability of some items than others, and a composite hedonic durability index averaged across people for a given item will be less meaningful for the former type of items than for the latter. In fact, understanding heterogeneity in people's η s, both across people and across items, may often be a useful analysis.

Total Happiness

While intended to measure hedonic durability, the three questions in the HDQ could also be used to estimate the total happiness from an item. We describe our approach below but note here that the estimate it yields is only a parsimonious approximation and relies on strong simplifying assumptions.

According to Kahneman et al. (1997) and Kahneman (2000), the total happiness an item generates is the temporal integral of one's moment-to-moment experience with the item over time. Thus, the total happiness of an item can be estimated by the total shaded area in Figure 6,

which is an extension of Figure 1. The time T is when the respondent answers the HDQ. The time T_+ is when the item would no longer evoke any happiness, even when it is still present. Based on the power function, we initially assume $T_+ = +\infty$.

 Insert Figure 6 about here

The shaded area to the left of T reflects the total happiness accrued from the item between the time of acquisition and the time of answering the HDQ; the area to the right of T reflects the total happiness expected to accrue between the time of answering the HDQ and the time of extinction of affective reactions toward the item, T_+ . Thus, the total happiness from the item is the sum of the two parts.

Formally, the total happiness (TH) can be expressed as:

$$TH = \int_0^{+\infty} H_0(1+T)^b dT = \begin{cases} -\frac{H_0}{b+1}, & b < -1 \\ 0, & b = -1 \\ +\infty, & b > -1 \end{cases} \quad (3)$$

However, the analysis above relies on several simplifying assumptions. First, we assume that happiness follows the same power function (i.e., shares the same η) after T as before T . This assumption is not necessarily always true in reality, but without knowledge of exactly how the η changes (e.g., a discrete event that changes how an item is enjoyed), this is a reasonable proxy when estimating average total happiness for a population. The lack of a relationship between

time and average η in Studies 4 and 5 provides some empirical validation for this assumption. That said, estimates of total happiness will be more robust when the η is measured a longer time after acquisition, as the tail end of the area, representing future total happiness, will be smaller.

A second and related assumption in this simple calculation is that the item does not wear out or die before happiness with the item fades away. In reality, every item wears off or dies eventually (i.e. before $T_+ = +\infty$), but for items with a long life relative to the decline in happiness, the bias in estimation should be small. When an expected life-span for the item, T_{life} , is known, the proposed model can be used to also estimate the total happiness conditional on the assumed end time point. Since T_{life} is shorter than T_+ (see Figure 6), we can truncate the shaded area at T_{life} , and calculate the total happiness as the truncated shaded area.

Formally, we revise Equation 3 by incorporating T_{life} in the calculation, as follows:

$$TH = \int_0^{T_{life}} H_0(1+T)^b dT = \begin{cases} \frac{H_0}{b+1} \left((1+T_{life})^{b+1} - 1 \right) & b \neq -1 \\ H_0 \ln(1+T_{life}), & b = -1 \end{cases} \quad (4)$$

When researchers have information about the life expectancy of an item, they should incorporate that information in estimating total happiness. If not, and if the items in the comparison list are durables (as is the case in Study 5), researchers may adopt a constant and relatively long estimated life expectancy, such as 5 years (as indicated by IRS depreciation schedules, for example).

Using the above method and assuming a T_{life} of 5 years, we calculated the TH s of all of the items in Study 5 and summarize the results in Figure 7. The items that topped the list were cats, wedding and engagement rings, dogs, and motorcycles. The results are fairly robust to the assumed life expectancy, and do not change substantially if we adopt either a shorter life

expectancy, such as 2.5 years, or a longer life expectancy, such as 10 years. However, if we use a different life expectancy for the different items, this could yield different results for total estimated happiness.

Insert Figure 7 about here

Total happiness can also be used to estimate the per-dollar happiness return on the purchase of an item and thereby guide consumers to make hedonically wise decisions. If we know the life-time cost of an item, C , we can estimate its per-dollar happiness return as TH/C . Suppose, for example, that on the basis of existing consumers' answers on the HDQ, we know that the total happiness of a rug is 10 and the total happiness of a mug is 5, yet the rug costs \$100 and the mug costs only \$5. Then we say that the per-dollar happiness return of the mug, $5/\$5 = 1$, is greater than that of the rug, $10/\$100 = 0.1$. An even more sophisticated analysis could take into account the timing of the costs and experienced happiness and incorporate a discount function to reflect a lower weight on costs and happiness that occur farther in the future.

Concluding Remarks

Like the action duration of a medicine, the hedonic durability of a product is an important variable. The availability of a method to measure hedonic durability can both help consumers make more informed purchase decisions, and allow researchers to make further progress on hedonic durability. Yet we are not aware of a simple and practical prior method to measure

hedonic durability. The HDQ is an initial attempt at such a method. This brief instrument can generate a wealth of information, including hedonic durability indices, total happiness estimates, and per-dollar happiness returns. For that, we hope that the HDQ has a place in the literature and will serve as a basis for developing more refined measurements in the future.

REFERENCES

- Aaker, Jennifer, Aimee Drolet, and Dale Griffin (2008), "Recalling Mixed Emotions," *Journal of Consumer Research*, 35 (2), 268-78.
- Brickman, Philip, Dan Coates, and Ronnie Janoff-Bulman (1978), "Lottery Winners and Accident Victims: Is Happiness Relative?," *Journal of Personality and Social Psychology*, 36 (8), 917-27.
- Dholakia, Utpal M and Vicki G Morwitz (2002), "The Scope and Persistence of Mere-Measurement Effects: Evidence from a Field Study of Customer Satisfaction Measurement," *Journal of Consumer Research*, 29 (2), 159-67.
- Di Tella, Rafael, John Haisken-De New, and Robert MacCulloch (2010), "Happiness Adaptation to Income and to Status in an Individual Panel," *Journal of Economic Behavior & Organization*, 76 (3), 834-52.
- Dockray, Samantha, Nina Grant, Arthur A Stone, Daniel Kahneman, Jane Wardle, and Andrew Steptoe (2010), "A Comparison of Affect Ratings Obtained with Ecological Momentary Assessment and the Day Reconstruction Method," *Social Indicators Research*, 99 (2), 269-83.
- Drew, Patrick J and Larry F Abbott (2006), "Models and properties of power-law adaptation in neural systems," *Journal of neurophysiology*, 96 (2), 826-33.
- Epstein, Leonard H., Jennifer L. Temple, James N. Roemmich Roemmich, and Mark E. Bouton (2009), "Habituation as a Determinant of Human Food Intake," *Psychological Review*, 116 (2), 384-407.

- Frederick, Shane and George Loewenstein (1999), "Hedonic Adaptation," in *Well-being: The Foundations of Hedonic Psychology*, Daniel Kahneman and Ed Diener and Norbert Schwarz, eds. New York, NY, US: Russell Sage Foundation.
- Fredrickson, Barbara L and Daniel Kahneman (1993), "Duration Neglect in Retrospective Evaluations of Affective Episodes," *Journal of Personality and Social Psychology*, 65 (1), 45-55.
- Galak, Jeff, Joseph P. Redden, Yang Yang, and Ellie J. Kyung (2014), "How Perceptions of Temporal Distance Influence Satiation," *Journal of Experimental Social Psychology*, 52 (0), 118-23.
- Hsee, Christopher K and Judy Ningyu Tang (2007), "Sun and Water: On a Modulus-Based Measurement of Happiness," *Emotion*, 7 (1), 213-18.
- Huttenlocher, Janellen, Larry V Hedges, and Norman M Bradburn (1990), "Reports of Elapsed Time: Bounding and Rounding Processes in Estimation," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16 (2), 196-213.
- Kahneman, Daniel (2000), "Experienced Utility and Objective Happiness: A Moment-Based Approach," in *Choices, Values and Frames* Daniel Kahneman and Amos Tversky, eds. Cambridge, England: Cambridge University Press.
- Kahneman, Daniel, Alan B Krueger, David A Schkade, Norbert Schwarz, and Arthur A Stone (2004), "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method," *Science*, 306 (5702), 1776-80.
- Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin (1997), "Back to Bentham? Explorations of Experienced Utility," *The Quarterly Journal of Economics*, 112 (2), 375-406.

- Lee, Kelly Kiyeon, Cynthia E. Cryder, and Stephen M. Nowlis (2014), "Jimmy Choo vs. Nike: Experienced Adaptation for Hedonic vs. Utilitarian Products," Working Paper. Washington University in St. Louis.
- Levine, Linda J and Martin A Safer (2002), "Sources of Bias in Memory for Emotions," *Current Directions in Psychological Science*, 11 (5), 169-73.
- Miron-Shatz, Talya, Arthur Stone, and Daniel Kahneman (2009), "Memories of Yesterday's Emotions: Does the Valence of Experience Affect the Memory-Experience Gap?," *Emotion*, 9 (6), 885-91.
- Nelson, Leif D., Tom Meyvis, and Jeff Galak (2009), "Enhancing the Television-Viewing Experience Through Commercial Interruptions," *Journal of Consumer Research*, 36 (2), 160-72.
- Nicolao, Leonardo, Julie R Irwin, and Joseph K Goodman (2009), "Happiness for Sale: Do Experiential Purchases Make Consumers Happier than Material Purchases?," *Journal of Consumer Research*, 36 (2), 188-98.
- Oppenheimer, Daniel M., Tom Meyvis, and Nicolas Davidenko (2009), "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power," *Journal of Experimental Social Psychology*, 45 (4), 867-72.
- Pearson, Robert W, Michael Ross, and Robyn M Dawes (1992), "Personal Recall and the Limits of Retrospective Questions in Surveys," *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, 65-94.
- Redden, Joseph P. (2008), "Reducing Satiation: The Role of Categorization Level," *Journal of Consumer Research*, 34 (5), 624-34.

Ross, Michael (1989), "Relation of Implicit Theories to the Construction of Personal Histories," *Psychological Review*, 96 (2), 341-57.

Safer, Martin A, George A Bonanno, and Nigel P Field (2001), "It was Never that Bad": Biased Recall of Grief and Long-Term Adjustment to the Death of a Spouse," *Memory*, 9 (3), 195-203.

Safer, Martin A, Linda J Levine, and Amy L Drapalski (2002), "Distortion in Memory for Emotions: The Contributions of Personality and Post-Event Knowledge," *Personality and Social Psychology Bulletin*, 28 (11), 1495-507.

Schacter, Daniel L (2002), *The Seven Sins of Memory: How the Mind Forgets and Remembers*: Houghton Mifflin Harcourt.

Scollon, Christie N, Chu Kim-Prieto, and Ed Diener (2003), "Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses," *Journal of Happiness studies*, 4 (1), 5-34.

Sheldon, Kennon M and Sonja Lyubomirsky (2012), "The Challenge of Staying Happier Testing the Hedonic Adaptation Prevention Model," *Personality and Social Psychology Bulletin*, 38 (5), 670-80.

Smith, Dylan M, George Loewenstein, Aleksandra Jankovic, and Peter A Ubel (2009), "Happily Hopeless: Adaptation to a Permanent, but not to a Temporary Disability," *Health Psychology*, 28 (6), 787-91.

Stone, Arthur A, Joseph E Schwartz, David Schkade, Norbert Schwarz, Alan Krueger, and Daniel Kahneman (2006), "A Population Approach to the Study of Emotion: Diurnal Rhythms of A Working Day Examined with the Day Reconstruction Method," *Emotion*, 6 (1), 139-49.

- Temple, Jennifer L, April M Giacomelli, James N Roemmich, and Leonard H Epstein (2008), "Dietary Variety Impairs Habituation in Children," *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, 27 (1 Suppl), S10-S19.
- Wang, Jing, Nathan Novemsky, and Ravi Dhar (2009), "Anticipating Adaptation to Products," *Journal of Consumer Research*, 36 (2), 149-59.
- Yang, Yang and Jeff Galak (2014), "Love It Longer: Sentimental Value Slows Hedonic Adaptation," Working Paper. Carnegie Mellon University.

Figure 1. The variables measured in the hedonic durability questionnaire (HDQ). The x-axis represents time, and the y-axis represents happiness.

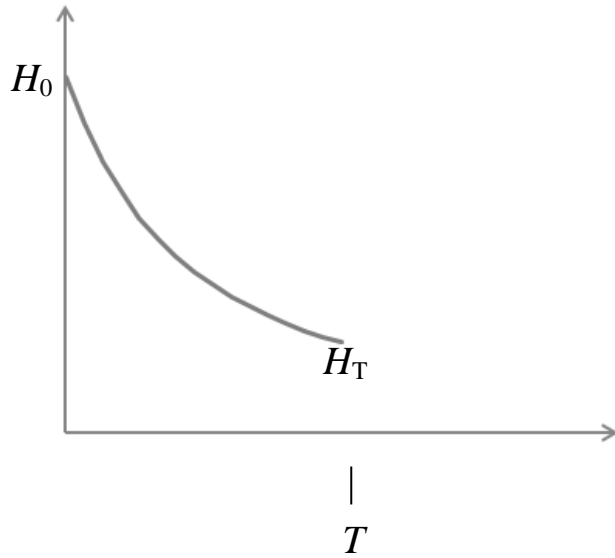
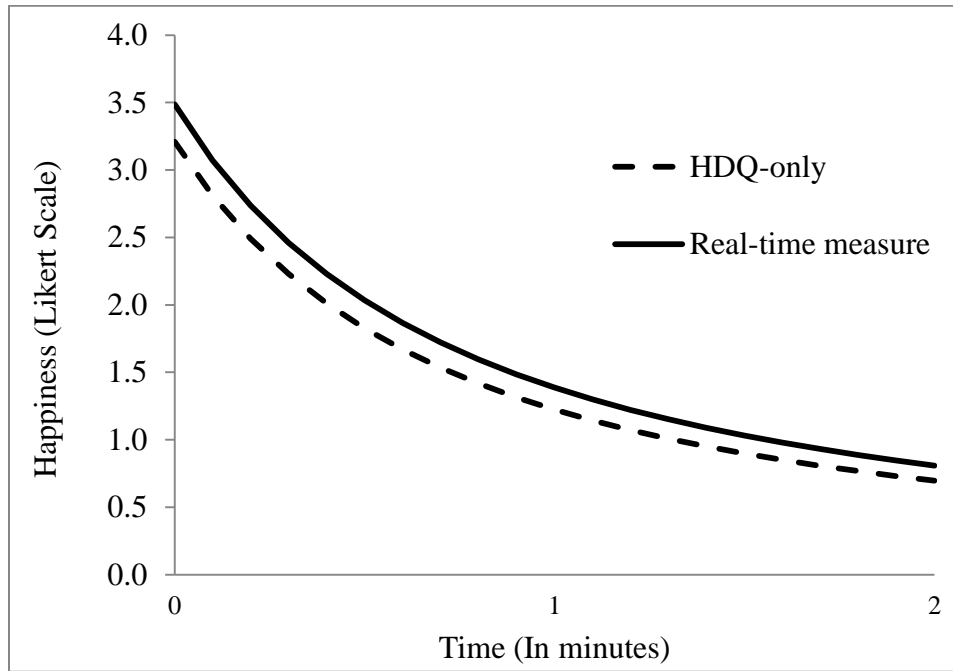
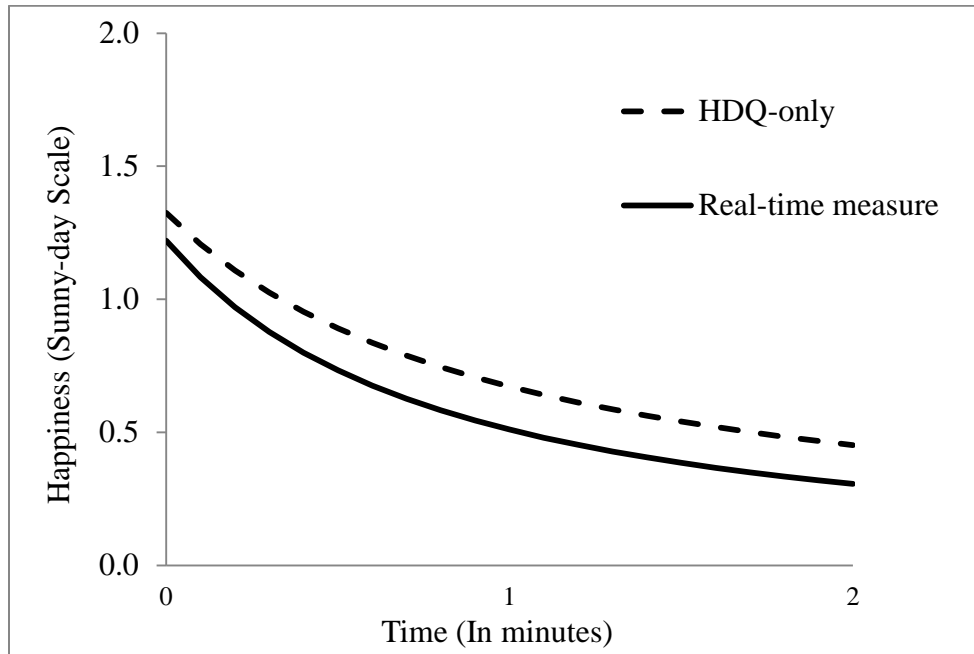


Figure 2A. The estimated happiness with the video clip over time in the HDQ-only vs. real-time measure conditions in Study 1, when the Likert scale was used.



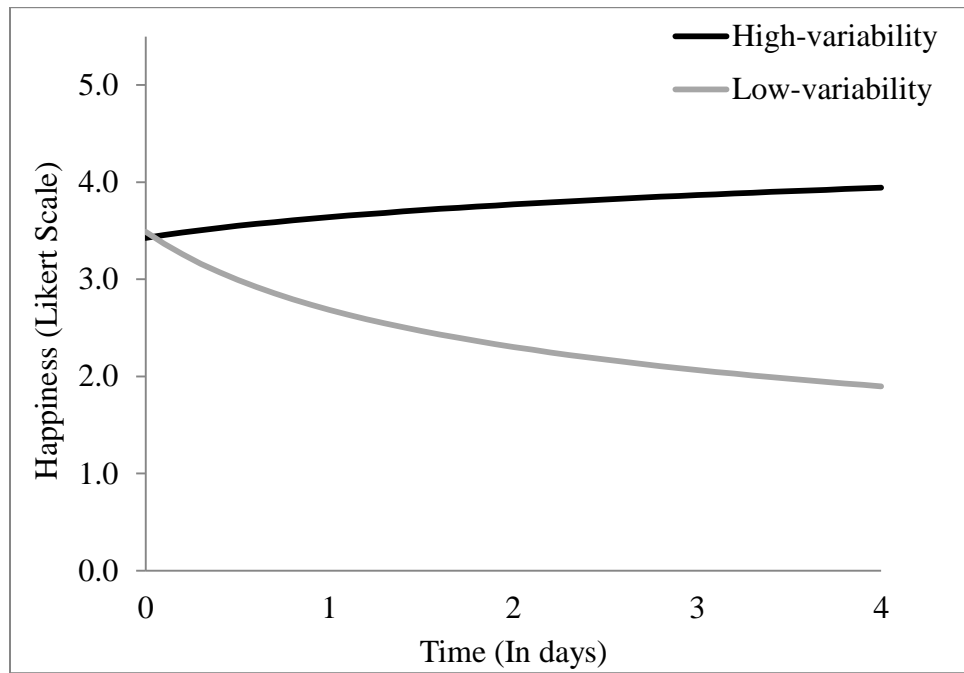
Note: The y-axis is hedonic rating on a Likert scale, ranging from -5 (very unhappy) to 5 (very happy).

Figure 2B. The estimated happiness with the video clip over time in the HDQ-only vs. real-time measure conditions in Study 1, when the “Sunny-day scale” was used.



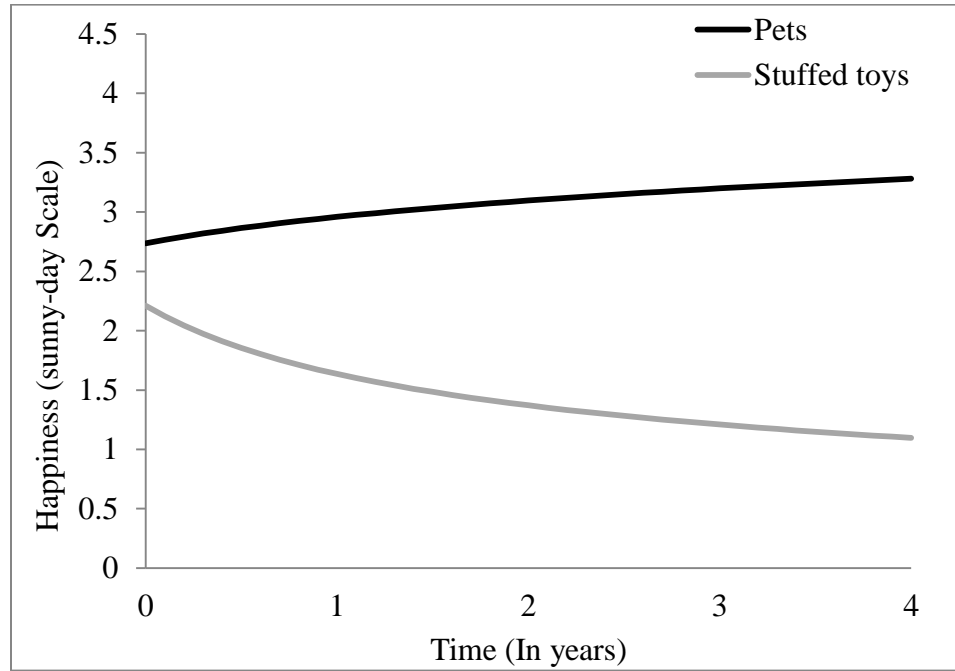
Note: The y-axis is hedonic rating on the “Sunny-day scale”.

Figure 3. The happiness with the video clips over time in Study 3.



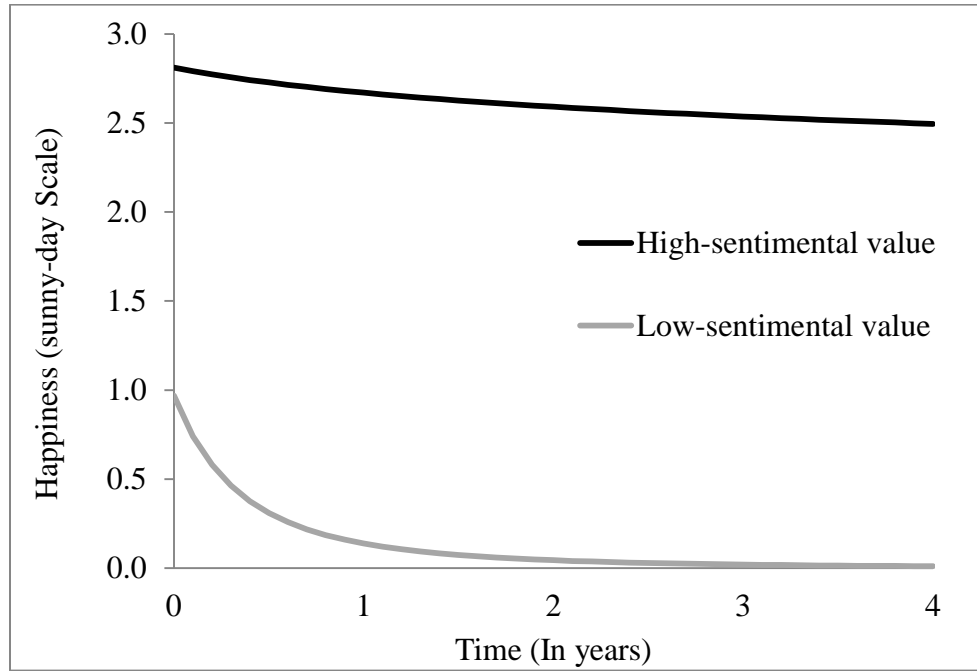
Note: The y-axis is hedonic rating on a Likert scale, ranging from -5 (very unhappy) to 5 (very happy).

Figure 4A. The happiness with pets and stuffed toys over time in Study 4.



Note: The y-axis is hedonic rating on the “sunny-day scale”.

Figure 4B. The happiness with high/low-sentimental value items over time in Study 4.



Note: The y-axis is hedonic rating on the “sunny-day scale”.

Figure 5. The hedonic durability indices (η_s) of the items in Study 5.

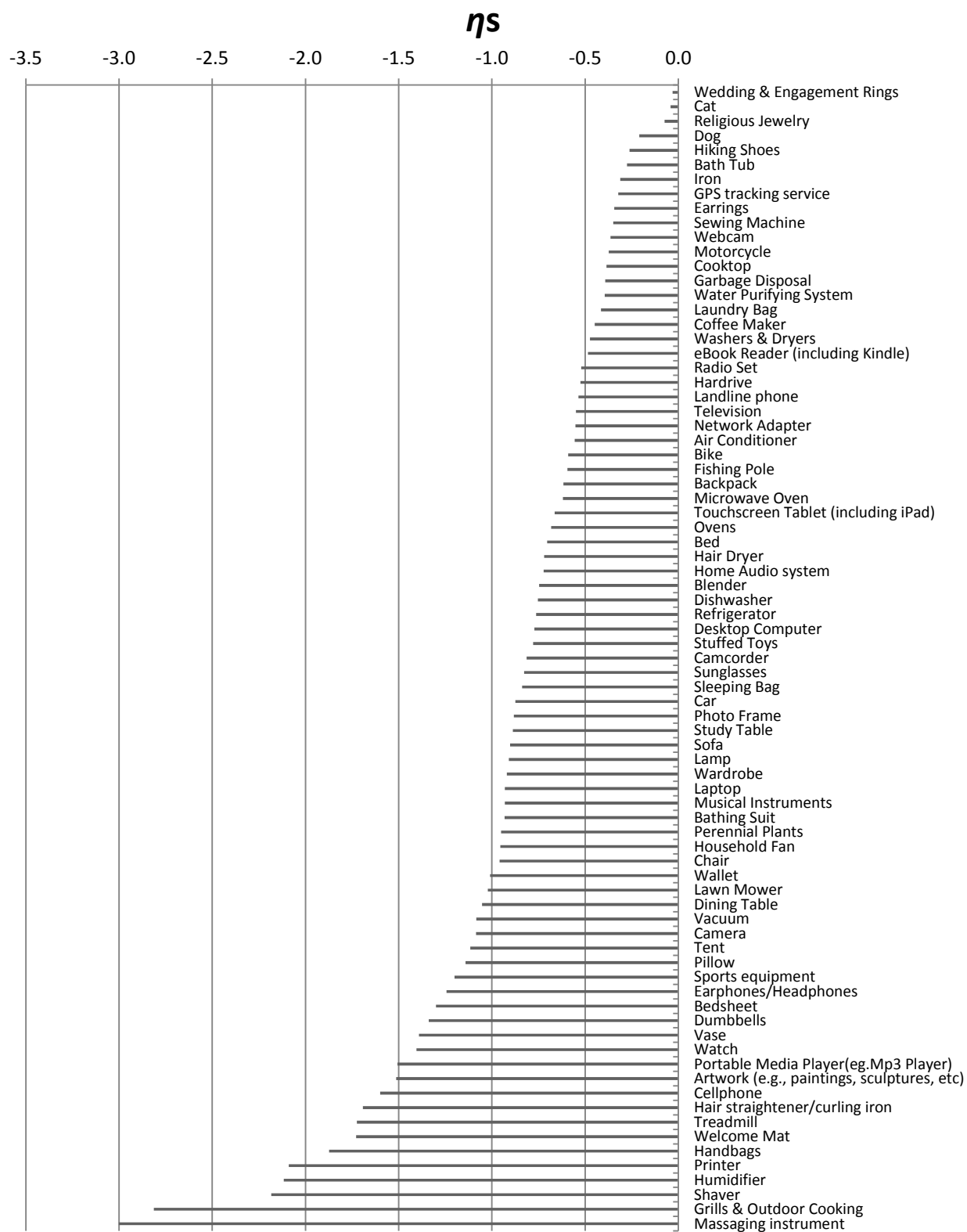


Figure 6. Total happiness (TH) as represented by the shaded areas.

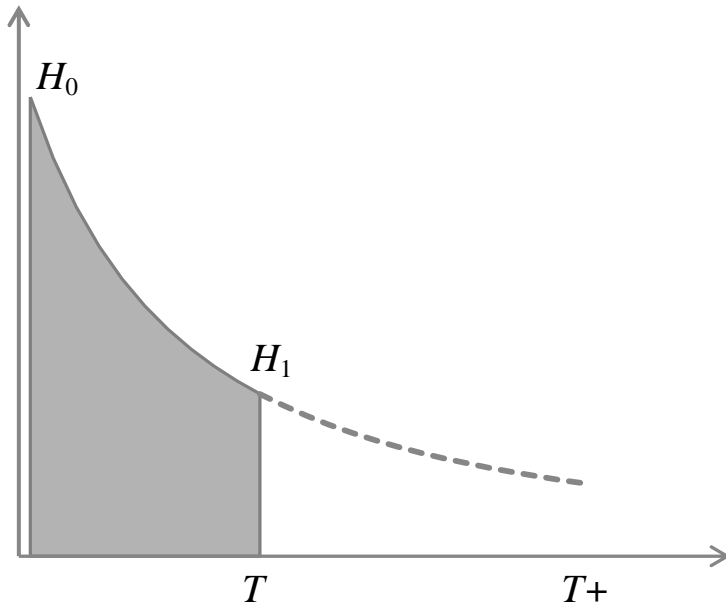
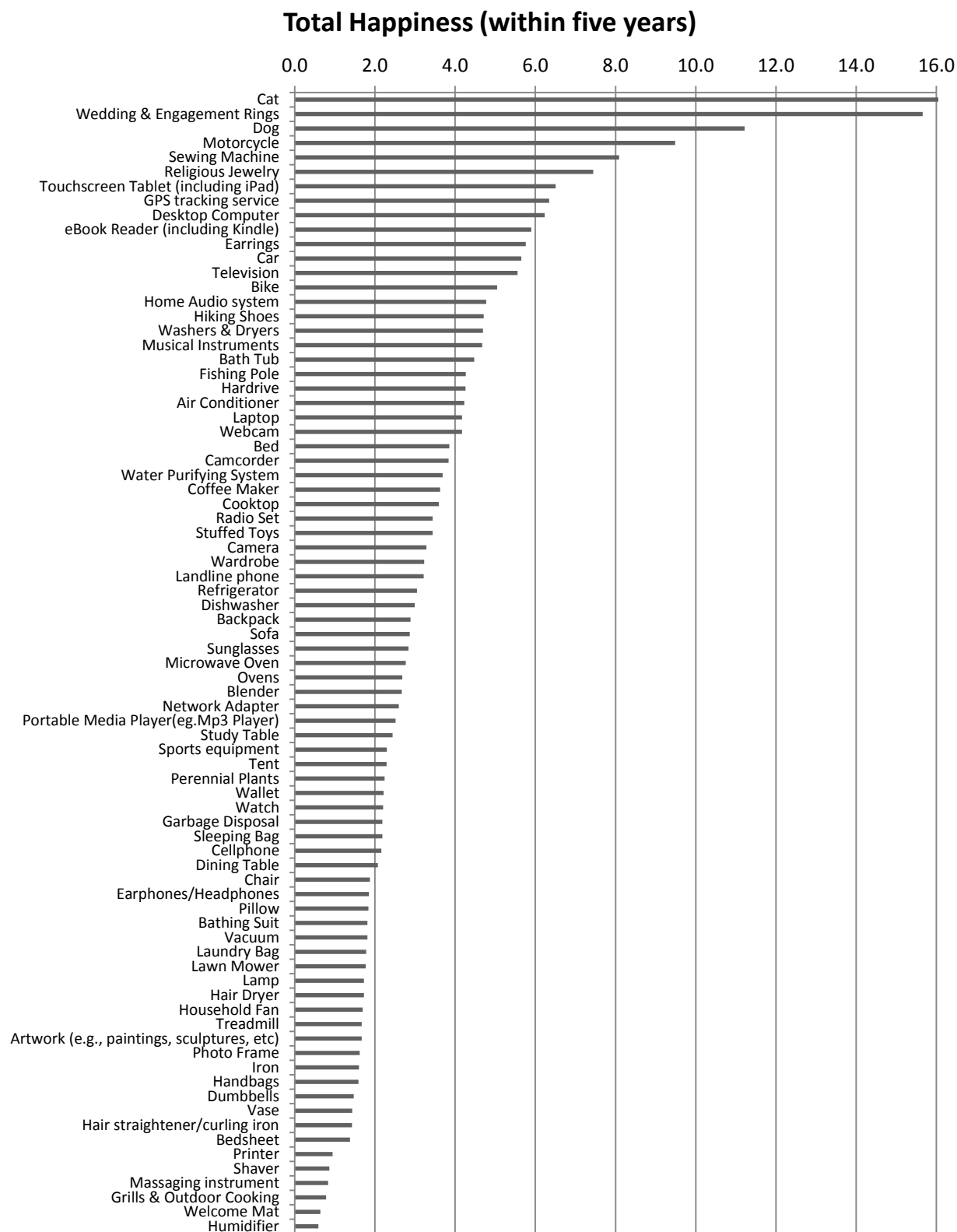


Figure 7. The total happiness (TH) of the items within five years in Study 5.



Appendix A

**RESULTS FOLLOWING THE TRUNCATION APPROACH (RECODING NEGATIVE
FINAL RATINGS TO .01)**

Study	Condition	H_0	H
Study 1 (N=186)	Real-time (Likert Scale)	2.96 (1.55)	-1.88 (2.26)
	HDQ-only (Likert Scale)	3.00 (1.33)	-2.03 (2.46)
	Real-time (Sunny Day Scale)	1.09 (.95)	-1.73 (1.83)
	HDQ-only (Sunny Day Scale)	1.28 (.71)	-1.41 (1.72)
Study 3 (N= 428)	High-variability	3.42 (1.63)	.02*** (1.21)
	Low-variability	3.50 (1.56)	-.57 (1.54)
Study 4 (N = 220)	Pet	2.72 (1.53)	-.01 [†] (1.00)
	Toy	2.21 (1.92)	-.44 (.93)
	High-sentimental value	2.81*** (2.65)	-.07** (.77)
	Low-sentimental value	1.00 (1.03)	-3.65 (9.83)

Note: Standard deviations are in parentheses. When the number in a cell is significantly different from the number below it, it is denoted by [†] for $p < .1$, * for $p < .05$, ** for $p < .01$ or *** for $p < .001$.

Appendix B

THE SUNNY-DAY SCALE INSTRUCTION USED IN STUDY 1

“In a moment, we will ask you to evaluate a video clip using a special scale, called the Sunny Day Scale. Here is how it works.

Imagine that it has been raining for a week and when you wake up today, you see a sunny day and feel happy. Let us define your degree of happiness at this moment as \hat{S} . To rate your feeling toward the video, just compare it with \hat{S} . For example, if your feeling toward the video is as happy as \hat{S} (i.e., as happy as your feeling toward seeing a sunny day after a week of rain), rate it as 1. If your feeling toward the video is half as happy as \hat{S} , then rate it as 0.5. If your feeling toward the video is twice as happy as \hat{S} , rate it as 2.

If you have no feeling toward the video, rate it as 0. If you feel unhappy about the video and the degree of your unhappiness is as intense as \hat{S} (i.e., as intense as your happiness toward seeing a sunny day after a week of rain), rate it as -1. If your degree of unhappiness is half as intense as \hat{S} , rate it as -0.5. If your degree of unhappiness is twice as intense as \hat{S} , rate it as -2.

You can use any numbers as you wish, such as 3, -0.8, 1.5, etc. However, please do not use fractions, such as $1/3$, $1\ 1/2$, etc.

Don't forget to put a negative sign if the feeling is negative (unhappy). You don't need to put a positive sign if the feeling is positive (happy).

To test whether you understand the Sunny Day Scale, please answer the following questions.

If your feeling toward the video is as happy as \hat{S} (i.e., as happy as your feeling at seeing a sunny day after a week of rain), how should you rate the video?

If your feeling toward the video is twice as happy as \hat{S} (i.e., twice as happy as your feeling at seeing a sunny day after a week of rain), how should you rate the video?

If you feel unhappy about the video and the degree of your unhappiness is half as intense as \hat{S} (i.e., half as intense as your happiness at seeing a sunny day after a week of rain), how should you rate the video?”

Participants will advance to the next page only if they answer the above questions correctly.

Appendix C

Results of Supplemental Statistical Analyses

Table 1: Linear Regression Predicting η
(Study 1, following the exclusion approach, N=143)

	β	Std Error	t	<i>p</i>
Constant	-1.25	0.16	7.88	.000
Method ^a	-0.07	0.16	0.43	.667
Scale ^b	0.12	0.16	0.78	.439
Method x Scale	-0.08	0.16	0.52	.605

Table 2: Linear Regression Predicting η
(Study 1, following the truncation approach, N=186)

	β	Std Error	t	<i>P</i>
Constant	-1.772	0.16	11.05	.000
Method ^a	-0.051	0.16	0.32	.749
Scale ^b	0.195	0.16	1.22	.226
Method x Scale	-0.116	0.16	0.72	.470

Table 3: ANOVA Predicting Rating of Video
(Study 3 pre-test, N=91)

	<i>SS</i>	df	F	<i>P</i>
Intercept	7636.7	1	2764.5	.000
Version of video	15.6	4	1.41	.236
Error	237.6	86		
Total	7915.0	91		

^a -1 = Likert, 1 = Sunny Day scale

^b -1 = HDQ, 1 = Real-time measurement

Table 4: Spearman correlation between η scores and duration of ownership
(Study 4, N=206)

Wave	Product	r_s	p	N
Wave I	Pet	-.05	.79	38
	Toy	.20	.32	27
	High sentimental	.07	.56	79
	Low sentimental	.11	.42	62
Wave II	Pet	-.32	.11	26
	Toy	-.09	.76	15
	High sentimental	.26	.08	48
	Low sentimental	-.11	.53	35

Table 5: Spearman correlation between initial happiness and duration of ownership (Study 4, N=206)

Wave	Product	r_s	P	N
Wave I	Pet	-.02	.90	38
	Toy	.12	.57	27
	High sentimental	-.17	.13	79
	Low sentimental	.28	.03	62
Wave II	Pet	.30	.14	26
	Toy	.24	.39	15
	High sentimental	-.25	.09	48
	Low sentimental	.34	.05	35

Table 6: ANOVA Predicting Initial Happiness (Study 4, N=206)

Wave I:

		SS	df	F	P
Pet	Intercept	274.255	1	115.986	.000
	quartile	8.474	3	1.195	.326
	Error	80.395	34		
	Total	373.500	38		
Toy	Intercept	129.948	1	34.232	.000
	quartile	8.432	3	.740	.539
	Error	87.311	23		
	Total	227.790	27		

High sentimental	Intercept	611.421	1	89.861	.000
	quartile	37.736	3	1.849	.146
	Error	510.304	75		
	Total	1172.000	79		
Low sentimental	Intercept	58.375	1	73.565	.000
	quartile	1.990	3	.836	.480
	Error	46.024	58		
	Total	106.136	62		

Wave II:

		SS	df	F	P
Pet	Intercept	211.609	1	74.779	.000
	quartile	15.783	3	1.859	.166
	Error	62.256	22		
	Total	283.000	26		
Toy	Intercept	80.852	1	94.114	.000
	quartile	5.883	3	2.283	.136
	Error	9.450	11		
	Total	97.000	15		
High sentimental	Intercept	420.261	1	33.079	.000
	quartile	38.724	3	1.016	.395
	Error	559.015	44		
	Total	1018.000	48		
Low sentimental	Intercept	32.859	1	58.756	.000
	quartile	.484	3	.288	.833
	Error	17.337	31		
	Total	53.001	35		
	Total	211.609	1	74.779	.000