Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2012-32 | June 6, 2012 | http://www.economics-ejournal.org/economics/discussionpapers/2012-32

Experimental Test of Utility Maximization

Yuqing He
Jinan University

Abstract The study tests the cardinal utility maximization hypothesis by an experimental procedure in a framework of utility scaling approach following the psychophysical-econometric paradigm, conceived in He (*Psychophysical Interpretation for Utility Measures*, 2011). It reveals (i) the utility maximization can be tested and has been supported by experimental results; (ii) the utility scaling approach following the psychophysical econometric paradigm offers a new foundation to discuss the utility concept; and (iii) it is necessary to distinguish the perception utility and emotion utility to respectively describe economic choices and enjoyment choices.

JEL A10, D01
Keywords Utility maximization; experiment; Klein–Rubin utility function; perception utility; emotion utility

Correspondence Yuqing He, Department of Public Administration, Jinan University, Guangzhou 510632, Guangdong Province, China: e-mail: hyq5.27@163.com

Introduction

The modern utility thought originated from Bentham's nominal measures of pleasure, happiness, and etc (Bentham, 1789). Bentham's outstanding contribution to the utility theory was to make the utility concept be viewed as a numerical magnitude, not by any successful measurement but by repeated fruitless discussions (e.g. Stigler, 1950). If no utility maximization hypothesis jointed from 1854 (Gossen, 1854), the utility thought would stay in economics mainly as an academic language decoration and would never become one of core thoughts in economics. It was the utility maximization reasoning to combine the nominal hedonic measures deeply with the marginal analysis. It changed the language of utility analysis into a more precise, complicated, and abstruse natural science form in its reasoning aspect but, in its pioneers' time, never approached to another characteristic of natural science: test a hypothesis in experiments.

Although Gossen had never been understood by the circles of economics during his lifetime, and obscurely died in 1858, his utility maximization thought represented the historical development in this field, and some gists addressed in his work, for example, the duality explanation that will be discussed right away, continues today. Gossen's earliest statement for the utility maximization is not as mathematical as his successors did:

A person maximizes his utility when he distributes his available money among the various goods so that he obtains the same amount of satisfaction from the last unit of money (Gendtom) spent upon each commodity. (see Stigler, 1950)

It contains such a meaning that there are two things, utility and money, and the comparison between them finally determines one's economic choice. This is a utility-money duality explanation, in which the utility concept follows Bentham's hedonic interpretation, representing pleasure, happiness, and satisfaction, or their opposite, is the psychological aspect, whereas, the money distribution adheres to the marginal analysis approach, is the economic reasoning aspect. This earliest duality explanation—perhaps had never formally been put forward in the history—was inherited by the successors of utility maximization thought and has yet a profound affection today.

One of its affections was the utility maximization analysis to be departed into two isolate domains today: one is the experimentation that concerns with the psychological hedonic aspect and terminates to change non-quantitative intuitive hedonic experiences into quantitative psychological analyses (e.g., Breaut, 1983; Kahneman et al., 1997), related to Bentham's tradition; and the other is the econometrics that concerns with the economic money aspect and starts from theoretical construction of utility function to explore the utility maximization model for describing market monetary distribution behaviors (e.g., Klein and Rubin, 1947; Stone, 1954; Liuch, 1973; Houthakker, 1960; Theil, 1965; Deaton and Muellbauer, 1980), related to the marginal analysis tradition. The former always interested on individual hedonic measures but appeared scattered and segmented in a complete theoretical context of utility maximization, whereas, the latter always concentrated on the systematic maximization model but lost in how to confirm the maximized utility function in individual observations.

This duality explanation is a methodological spell fascinating researchers unconsciously to slide into the duality demarcation, in which experimentalists never probed into how a consumer distributes the money in a commodity bundle for his utility maximization, while, econometricians never understood how a maximized utility function in their models represents a consumer's hedonic feelings. For experimentalists the utility seems a hedonic feeling somehow to relate with the money distribution in one's economic concern, while, for econometricians the utility is only a mathematical function "u", which can be maximized, somehow to involve one's hedonic feelings. This is of the realistic picture for the utility concept differently between experimentalists and econometricians. Such a demarcation cut the complete utility maximization issue into two irrelevant halves, and made either experimentation or modeling approach fail to deal with a complete behavioral process in the utility maximization. It truly disabled the test of utility maximization.

We need a utility maximization theory covering its experimental foundation and empirical application. The methodological spell must be broken away.

No long after the utility concept was extensively publicized, a number of researchers, including Gossen, continuously thought it immeasurable (Spiegel, 1991). Another development of the utility theory is the cardinal utility, including its maximization, being replaced by the ordinal utility in microeconomics. Based on the indifference curve description, the ordinal utility theory only involves the pair-wise comparison between a consumer's preferences, and assumes that a consumer is self-evident to rank any consumption bundles by order of preference.

Nonetheless, the ordinal utility theory might neglect a serious question that one's cognition to the indifference curve is suspicious as a self-evident experience, and needs the experimental confirmation too. There has been however no clear experimental evidence to confirm the indifference curve that sufficiently satisfies all three strict standards (convexity, diminishing, and non-intersecting) for determining a utility maximization measure in subjects' performances (e.g., MacCrimmon and Toda, 1969). Experimental or empirical evidences always absented in the ordinal utility theory. In neo-classical economics, the best reason for keeping the utility maximization hypothesis might be merely that it was always neither testable nor falsifiable but seemed indispensable.

In fact, the cardinal utility is not so immeasurable, while, the experimental test of utility maximization is not so beyond our attainments. The successful psychophysical studies of appealing us to attempt the cardinal utility measurement had appeared for a long time (e.g., Galanter, 1962, 1990; Galanter and Pliner, 1974;

Galanter et al., 1977; Breaut, 1983; Parker and Schneider, 1988). In He, 2011, also basing on the psychophysical paradigm, Klein-Rubin utility function (Klein and Rubin, 1947) had been demonstrated as a linear combination of logarithmic laws for separate commodity utility measures in a bundle of commodities. It indicates a way to estimate experimental Klein-Rubin utility function in a bundle of goods, and further to derive a direct experimental test to the cardinal utility maximization. In this way, the experimentation and the econometric modeling approach will be firstly combined together to break away "the methodological spell".

We have come to a time to relook at the cardinal utility maximization hypothesis through experimental insights with a psychophysical-econometric paradigm. The following text will report in detail how such a test had been performed in an experimental procedure during March and April, 2012.

This study revealed that the cardinal utility maximization holds in subjects' choice behaviors in a three-commodity bundle at the acceptable level of $R^2 \geq 0.97$ for the curve fitting. Waiting for one hundred and fifty-eight years, Gossen's genius is firstly supported by experimental evidences. More important for resent economics, such an experimental study indicates a new analytical framework, which will be discussed in Section 3.

In this paper, Section 1 describes the design and performance of the experiment. Section 2 presents the experimental results. Section 3 summarizes and discusses the findings. The data sets and appended mathematical derivations are presented in Supplemental Files (please check on the journal web site).

1 Experimental design and performance

The following experimental designs refer to a series of probing experiments, conducted during April to October, 2010.

1.1 Analytical path

Denote U_{LES} as the Klein-Rubin utility function derived from Linear Expenditure System (LES) (Stone, 1954) in a three-commodity bundle, and U_{Est} the Klein-Rubin utility function estimated from the same commodity bundle by the utility scaling approach (He, 2011). The experiment is designed to test the utility maximization by comparing U_{LES} with U_{Est} in subjects' experimental choices in the same commodity bundle.

As well known, LES is the result of maximizing the Klein-Rubin utility function, and thus, U_{LES} is the utility function theoretically required by the utility maximization for the chosen commodity bundle. In the light of He, 2011, U_{Est} is the intuitive utility estimate for the same commodity bundle, and indicates subjects' intuitive utility judgment. If the comparison reveals an agreement between U_{LES} and U_{Est}, the theoretic utility maximization implied in LES agrees with the utility evaluation derived from subjects' intuitive judgment, that is, subjects' choices are just the result of their intuitive utility maximization, and thus, the experiment supports the utility maximization hypothesis; otherwise, does not.

In a three-commodity bundle, the Klein-Rubin utility function is

$$U = \sum b_k \ln(q_k - r_k), \quad b_k = \frac{p_k(q_k - r_k)}{\sum p_k(q_k - r_k)}, \quad k = 1,2,3.$$

Where, p_k is the price for commodity k, q_k the purchase quantity for commodity k, and, b_k and r_k are two undetermined parameters. To determine U_{LES} and U_{Est}, the experimental data of p_k, q_k, b_k, and r_k will be required.

p_k can be assigned as the experimental setting factor, and q_k is the subjects' quantity choices for commodity k in the experiment. The key problem for the experiment is how to determine b_k and r_k respectively for U_{LES} and U_{Est} in the experimental data of p_k and q_k.

(i) Determining U_{LES}

For a three-commodity bundle, LES is

$$p_k q_k = p_k r_k + b_k(\sum p_k q_k - \sum p_k r_k). \quad k = 1,2,3.$$

Solving this equation in the experimental data of p_k and q_k, b_k and r_k contained in U_{LES} will be derived, and $U_{LES} = b_k \sum (q_k - r_k)$ will be determined.

(ii) Determining U_{Est}

According to He, 2011, each r_k for determining U_{Est} is contained in the logarithmic laws

$$u_k = c \ln(q_k - r_k) + C, \quad k = 1,2,3,$$

which can be derived from subjects' utility estimations for each commodity k. And then, further determine b_k by

$$b_k = \frac{p_k(\overline{q}_k - r_k)}{\sum p_k(\overline{q}_k - r_k)}, \quad k = 1,2,3,$$

here $\overline{q}_k$ is the average of q_k. Finally, $U_{Est} = b_k \sum (q_k - r_k)$ will be determined.

The experiment consists of two sessions, labeled Session A and Session B. Session A provides data to determine b_k and r_k for U_{LES} by solving LES in the experimental data of p_k and q_k, and Session B offers data to estimate the logarithmic laws which contain r_k for U_{Est}. b_k for U_{Est} is determined by combining the data of q_k obtained from Session A with r_k derived from Session B.

1.2 Experimental designs

There are two experiments distinguished by their different types of commodity bundles, called Exp. 1 and Exp. 2. In Exp. 1, the trading goods are three kinds of between-meal nibble kernels, pistachio, almond, and cashew nut, they are mutually substitutable to most people and easy to compare with each other; and in Exp. 2, the trading items are different three types of goods, apple, pen, and facial tissue, they are mutually un-substitutable in everyday life and difficult to compare with each other. Beside of the differences between their trading goods, the designs for Exps. 1 and 2 are the same. Below, take Exp. 1 as an example to illustrate the experimental designs.

As mentioned above, an experiment includes two sessions, Session A and Session B. In Exp. 1, Session A includes three choices of commodity quantity combinations, labeled Choice I, Choice II, and Choice III, in each of which three kinds of kernels, pistachio, almond, and cashew nut as three trading items are available for subjects' purchases. Below, in Exp. 1, subscripts "I", "II", and "III" always indicate Choices I, II, and III, and subscripts "1", "2", and "3" the pistachio, almond, and cashew nut. In addition, the experiment uses Chinese money RMB to price the goods, for succinctness, price p_{ij}, $i =$I,II,III, $j = 1,2,3$, for example, "RMB0.50" will be briefly denoted as "0.50".

In Choice I, subjects are asked to report q_{Ij}, $j = 1,2,3$, the quantities of three kinds of kernels they are willing to buy at assigned prices $(p_{I1}, p_{I2}, p_{I3}) = (0.50, 0.50, 0.50)$; in Choices II and III, subjects are also asked to report q_{IIj} and q_{IIIj}, $j = 1,2,3$, the quantities of three kinds of kernels they are willing to buy at assigned prices $(p_{II1}, p_{II2}, p_{II3}) = (0.70, 0.30, 0.50)$ in Choice II, and at $(p_{III1}, p_{III2}, p_{III3}) = (0.50, 0.30, 0.70)$ in Choice III. They will deliver the data of q_{ij}, $i =$I,II,III, $j = 1,2,3$ in three sets of prices. Choices I, II, and III respectively provide experimental data to construct the Klein-Rubin utility function U_{LES} in three distinguished price combinations. That is, its maximization will be tested under three price combinations in Exp. 1. The Klein-Rubin utility function U_{LES} is specifically written as

$$U_{LES} = \sum b_{ij} \ln(q_{ij} - r_{ij}), \quad i = \text{I,II,III}, \quad j = 1,2,3.$$

Subjects consecutively finish Choices I-III on an answer sheet. This is a real pay-off test. Subjects will be told that they are engaged to pay their choices. Meanwhile, they are encouraged to freely decide to buy or reject one, two, or all three kinds of kernels, completely basing on their own interests.

To control subjects' purchases naturally following some budget constraints, in Session A, every subject is restricted to purchase no more than RMB6.60 for each of Choices I, II, and III in Exp. 1. Meanwhile, to prevent the subjects with too low purchasing desires entering the valid sample, only those who purchase at least 45% the maximal money amount, namely no less than RMB3.00, in a choice bundle will be valid. That is, every valid subject will purchase between RMB3.00-6.60 in a choice bundle in Exp. 1. This is called Valid Condition 1.

In addition, evidently, the derivation of utility maximization implies that a subject keeps his preference consistent in his purchase choices. It is called Valid Condition 2. An experiment for testing the utility maximization hypothesis should also provide clear information to guarantee it.

In the probing experiments conducted in 2010, observations discovered that a subject's choices might be motivated only by expending assigned maximal money amount RBM6.60 as more as possible but completely neglected his true consumption preference. Namely, Valid Condition 2 is uncertainly followed by a subject in the experiment. A special discrimination is inevitable. You will see in Subsection 1.3.2 that the design of consecutive Choices I-III can be used to reveal whether or not a subject follows Valid Condition 2.

Made up by six sets of price choices, Session B measures subjects' six utility scales (logarithmic laws) u_{ij}, $i =$I,II,III and $j = 1,2,3$, for assigned quantities of the kernels referring to the offered prices assigned in Choices I, II, or III. It provides data to estimate r_{ij}, $i =$I,II,III and $j = 1,2,3$, for U_{Est} from the logarithmic laws u_{ij}. The Klein-Rubin utility function U_{Est} is specifically written as

$$U_{Est} = \sum b_{ij} \ln(q_{ij} - r_{ij}), \quad b_{ij} = \frac{p_{ij}(\overline{q}_{ij} - r_{ij})}{\sum p_{ij}(\overline{q}_{ij} - r_{ij})}, \quad i = \text{I,II,III}, \quad j = 1,2,3.$$

The way of measuring the logarithmic laws is similar to that used in "Meas. 3" of the electrical-power massage experiment (He, 2011): five quantities for a kind of kernel with an offered unit price are randomly shown to subjects, and subjects are asked to report their bid unit prices. The bid unit price is the subject's utility estimate. To ensure the sufficient information necessary for presenting a utility scale, a valid subject at least reports the valid bid unit prices for three of five assigned quantities in each utility scale for two choice bundles. This is Valid Condition 3.

In order to compare U_{LES} with U_{Est}, a valid subject must provide valid data simultaneously in Choice I and one of Choices II and III in Session A and deliver corresponding utility scales in Session B. This is Valid Condition 4.

The data, at least satisfying Valid Conditions 1, 2, 3, and 4, will be used in the experimental analyses.

Utility judgments in Sessions A and B, with the identical character that quantities and prices are all informed to subjects, are the same type of single estimate (He, 2011). It ensures the utility scales obtained from Session B can be compared with subjects' choice results in Session A.

Now we have separate two ways to estimate the Klein-Rubin utility function in the experiment: using p_{ij} and q_{ij} of Session A to solve LES, the Klein-Rubin utility function U_{LES} will be obtained experimentally; and combining q_{ij} of Session A with u_{ij} of Session B, another Klein-Rubin utility function U_{Est} will be obtained experimentally (He, 2011). As mentioned above, if U_{LES} agrees with U_{Est}, it will indicate subjects' choices in the experiment agreeing with their utility maximization estimate. In other words, we are able to test the utility maximization hypothesis by comparing the utility functions obtained from the above two ways.

1.3 Experimental performances

1.3.1 Participants

Two types of subjects participated in experiments: career persons working in the agents of business, government, and etc, called C-Sample, and full-time undergraduate students from Jinan University, called S-Sample. C-Sample only participated in Exp. 1, and S-Sample participated in Exps. 1 and 2. Table 1 outlines their status.

Table 1 Status of each sample

Sample	Subject	Male	Female	Age	Mean of ages
C-Sample	105	47	58	23-39	27.6
S-Sample	121	61	60	19-22	20.3

1.3.2 Performance of Exp. 1

In Exp. 1, the three kinds of kernels respectively contained in mini bags were used as goods traded in the experiment (see the photo pictures in Part 1 of Supplemental Files). All mini bags were transparent to show kernels to subjects. A mini bag was a trading unit with same quantities but often different prices between Choices I-III as shown in Table 2. To attract subjects to purchase in the experiment, all prices assigned in Table 2 are usually lower than 30% market prices.

Table 2 Prices for mini-bagged kernels (RMB)

	Pistachio	Almond	Cashew nut
Choice I	$p_{I1} = 0.50$	$p_{I2} = 0.50$	$p_{I3} = 0.50$
Choice II	$p_{II1} = 0.70$	$p_{II2} = 0.30$	$p_{II3} = 0.50$
Choice III	$p_{III1} = 0.50$	$p_{III2} = 0.30$	$p_{III3} = 0.70$

In Session A, the real objects of three kinds of mini-bagged kernels were shown to subjects. The three sets of prices assigned in Table 2 were shown on a sheet for every subject. The experimenter instructed to them: "This is a quantity-choice test. You are engaged to pay your choices at assigned prices. Everywhere, if you are reluctant to buy a trading item, please feel free to reject it by filling with a zero on the answer sheet. There will be three sets of quantity choices offered to you. In each choice set, there are three kinds of kernels with their assigned prices. Your purchase must be no more than RMB6.60 for every set. Otherwise, it will be invalid. You should choose in all Choices I, II, and III. But there will be only one choice set to be finally traded for you, because we shall randomly determine, by lot, only one of Choices I, II, and III to be executed. You don't worry about buying too many kernels in the experiment."

The designs in Table 2 can be used to examine Valid Condition 2 in subjects' choices. It is realized through two steps. The first, reveal a subject's true preference to three kinds of kernels in Choice I, and the second, survey whether this subject persists his preference in Choices II and III. In Choice I, the assigned unit prices for all three kinds of kernels are the same (all 0.50/bag), the motivation of "expending assigned maximal money amount 6.60 as more as possible" thereby does not disturb a subject's preference in his choice behaviors, accordingly, the differences between quantities chosen for the three kinds of kernels in Choice I will naturally indicate a subject's different preferences to three kinds of kernels. In other words, Choice I truly reveals a subject's preference and provide the comparable criteria for surveying a subject's preference performances.

In the practical performance, if Valid Condition 2 was held in a subject's Choices I-III, when the unit price of a kind of kernel in Choice II was, for instance, higher than that in Choice I and, simultaneously, the other two kinds of kernels had no increases in their unit prices, the quantity chosen for this kind of kernel in Choice II should be at least no more than that in Choice I; and so on. If the case was opposite, Valid Condition 2 failed to be fulfilled in choice behaviors, and the data should be discarded. And further, it was the same to the case from Choice II to Choice III. In this way, we could determine in the main whether a subject's behaviors satisfy Valid Condition 2. To reveal Valid Condition 2, a valid subject must buy all three kinds of kernels in Choice I and, simultaneously, at least buy all three kinds of kernels in one of Choices II and III. This requirement has been contained in Valid Condition 4.

Having concluded Session A, those who offered valid data in Session A were selected to proceed to Session B to measure utility scales u_{ij}, $i =$I,II,III and $j = 1,2,3$. In Session B, the experimenter instructed to subjects: "In the just finished choices, you were only allowed to choose quantities but not to choose prices. Now you are allowed to bid prices for each assigned quantity with reference to offered prices".

Subjects were asked to report their bid unit prices for five assigned quantities 4, 6, 8, 10, 12 (counted in mini bag number) with reference to an offered unit price for a kind of kernel. The offered unit prices were those assigned in Session A: 0.50/bag and 0.70/bag for pistachio, 0.50/bag and 0.30/bag for almond, and 0.50/bag and 0.70/bag for cashew nut (see Table 2). The reported bid unit prices were the utility estimates. They provided data to determine the logarithmic laws for estimating parameters r_{ij} and b_{ij} in U_{Est}.

There were three U_{LES}s in Session A respectively describing three Klein-Rubin utility functions in Choices I, II, and III. Every U_{LES} contained three logarithmic terms representing the utilities derived from three kinds of kernels, therefore, there were nine logarithmic terms in sum. To construct U_{Est} for comparing with U_{LES}, there were also nine utility scales (logarithmic laws) to be estimated. To simplify the measurement procedure, the cross affections between different kinds of kernels were not taken into account in the estimations of U_{Est}, and thus, subjects reported their utility judgments for the kernels only basing on the assigned quantities and offered unit prices for these kernels. The series of assigned quantities for three kinds of kernels in Session B are the same, and all are 4, 6, 8, 10, 12. With this simplification, if the assigned unit prices for a kind of kernel were the same in different choice bundles, they would be described by the same logarithmic law in the estimations of U_{Est}. According to the designs in Table 2, the assigned unit prices for pistachio were the same in Choices I and III, for almond were the same in Choices II and III, and for cashew nut were the same in Choices I and II. In other words, the pistachio was described by the same logarithmic law in Choices I and III, the almond by the same logarithmic law in Choices II and III, and the cashew nut by the same logarithmic law in Choices I and II. The number of estimated logarithmic laws in Session B was therefore reduced to six. It would greatly increase the valid rate in the experimental data (see Subsection 2.4).

From Table 2, the six logarithmic laws estimated in Session B were specified as follows:

1) for pistachio with unit price 0.50/bag (assigned in Choice I and III);
2) for pistachio with unit price 0.70/bag (assigned in Choice II);
3) for almond with unit price 0.50/bag (assigned in Choice I);
4) for almond with unit price 0.30/bag (assigned in Choices II and III);
5) for cashew nut with unit price 0.50/bag (assigned in Choices I and II);
6) for cashew nut with unit price 0.70/bag (assigned in Choice III).

They are summarized in Table 3.

Table 3 Assigned quantities and prices in Session B of Exp. 1

Kernel	Offered unit price	Assigned quantities	Measuring utility scales
Pistachio	0.50/bag	4-6-8-10-12	in Choices I and III
Pistachio	0.70/bag	4-6-8-10-12	in Choice II
Almond	0.50/bag	4-6-8-10-12	in Choice I
Almond	0.30/bag	4-6-8-10-12	in Choices II and III
Cashew nut	0.50/bag	4-6-8-10-12	in Choices I and II
Cashew nut	0.70/bag	4-6-8-10-12	in Choice III

There are totally six unit prices and thirty quantities assigned in Table 3 for three kinds of kernels.

To lessen the order effect in subjects' judgments, thirty quantities assigned in Table 3 were randomly ordered one by one to show to subjects in the experiment. Subjects were asked to report their bid unit prices one by one for assigned quantities with reference to offered unit prices. The measurement process was similar to that held in "Meas. 3" of the electrical-power power massage experiment (He, 2011): the experimenter presented to a subject a unit price inquiry card which indicated, for instance, "if you are asked to buy 6 bags of almond with offered unit price 0.30/bag, your bid unit price will be (   )", the subject wrote down his bid unit price in the bracket, and the experimenter collected the card; then the next inquiry card was presented to the subject, the subject reported his bid unit price, and so on, until all thirty inquiry cards (correspond to thirty assigned quantities in Table 3) had been randomly presented to the subject.

The data used in curve regressions for measuring utility scales were created by multiplying subjects' bid unit prices with corresponding assigned quantities. For example, if the subject's bid unit price for quantity "10" of "Almond (0.50/bag)" was 0.35 in Choice I, it would deliver a datum $0.35 \times 10 = 3.50$ at quantity "10" for the almond utility scale in Choice I. This data creating manner identified with subjects' judgment manner in Session A, in which beside of referring to their preference, subjects determined their purchase quantity by multiplying the unit price with the purchase quantity to judge the total purchase money amount no more than RMB6.60 in a choice bundle.

To get rid of subjects' worry about buying too many kernels in Session B, the experimenter declared the transaction regulation to them: "For every kind of kernels, only one quantity among assigned quantities will be selected, by lot, as executed trade, and if your bid price is between the mode price $\pm 10\%$ for the executed quantity, your trade will be successful, otherwise will not. With this regulation, at most, only three quantities among thirty priced quantities are possible to be traded. Therefore, you just independently bid for every assigned quantity and don't worry about paying too much for cumulative successful trades." Usually, the terminology "mode price" was easily addressed to subjects.

1.3.3 Performance of Exp. 2

With the similar designs to Exp. 1, Exp. 2 also contains three sets of choices in its Session A. In Exp. 2, the subscripts "I", "II", and "III" also always indicate Choice I, Choice II, and Choice III, and the subscripts "1", "2", and "3" the apple, pen, and facial tissue.

Table 4 presents the assigned prices of apple, pen, and facial tissue in Session A, and Table 5 presents the assigned quantities and offered prices in Session B.

Valid Conditions 2, 3, and 4 in Exp. 2 are the same to those in Exp. 1, but Valid Condition 1 requests that all valid subjects purchase between RMB3.60-8.00

in each of Choices I-III.

Table 4 Prices in Session A of Exp. 2 (RMB)

	Apple	Pen	Facial tissue
Choice I	$p_{I1} = 0.50$	$p_{I2} = 0.50$	$p_{I3} = 0.50$
Choice II	$p_{II1} = 0.70$	$p_{II2} = 0.90$	$p_{II3} = 0.50$
Choice III	$p_{III1} = 0.50$	$p_{III2} = 0.90$	$p_{III3} = 0.70$

Table 5 Assigned quantities and prices in Session B of Exp. 2

Goods	Offered unit price	Assigned quantities	Measuring utility scales
Apple	0.50/piece	4-6-8-10-12	in Choices I and II
Apple	0.70/piece	4-6-8-10-12	in Choice III
Pen	0.50/piece	4-6-8-10-12	in Choice I
Pen	0.90/piece	4-6-8-10-12	in Choices II and III
Facial tissue	0.50/set	4-6-8-10-12	in Choices I and III
Facial tissue	0.70/set	4-6-8-10-12	in Choice II

2 Results

2.1 Results in Exp. 1 of C-Sample

105 subjects participated in C-Sample, and 38 of them delivered valid data in Choice I, 37 in Choice II, and 36 in Choice III.

By solving LES in the category data of Session A, each U_{LES} for Choices I-III of C-Sample is derived as (1), (3), and (5) (for details please see Part 2 of Supplemental Files).

To derive U_{Est}, it needs to estimate b_{ij} and r_{ij} respectively by two steps. The first step, fit the logarithmic law $u_{ij} = c\ln(q_{ij} - r_{ij}) + C$, $i =$I,II,III and $j = 1,2,3$, in the average data of Session B to determine r_{ij}. It is realized by the curve regression in SPSS, in which the optimal values of c and C in the logarithmic law are created automatically, but the values of r_{ij} must be selected by hand. To isolate from the derivation of U_{LES}, the procedure of selecting r_{ij} is that select the values of r_{ij} to improve the regression results in SPSS until $R^2 \geq 0.97$. Except of the regression curve for the cashew nut in Choices I and II rounding its value of R^2 from 0.966 to 0.97, all other regression curves are rigorously satisfy $R^2 \geq 0.97$ (for details please see Part 3 of Supplemental Files). And the second step, use $\bar{q}_{ij}$ the average quantities subjects chose in Session A and the values of r_{ij} to determine

b_{ij} by $b_{ij} = \frac{p_{ij}(\bar{q}_{ij}-r_{ij})}{\sum(\bar{q}_{ij}-r_{ij})}$ (for details please see Part 3 of Supplemental Files). Using the values of b_{ij} and r_{ij}, each U_{Est} for Choices I-III of C-Sample is derived as (2), (4), and (6):

In Choice I,

$$U_{LES} = 0.31\ln(q_{I1} - 2.95) + 0.33\ln(q_{I2} - 2.56) + 0.36\ln(q_{I3} - 3.49) \quad (1)$$

$$U_{Est} = 0.32\ln(q_{I1} - 2.39) + 0.33\ln(q_{I2} - 2.30) + 0.34\ln(q_{I3} - 2.50) \quad (2)$$

In Choice II,

$$U_{LES} = 0.32\ln(q_{II1} - 0.96) + 0.41\ln(q_{II2} - 3.11) + 0.28\ln(q_{II3} - 2.79) \quad (3)$$

$$U_{Est} = 0.33\ln(q_{II1} - 0.96) + 0.40\ln(q_{II2} - 2.40) + 0.27\ln(q_{II3} - 2.50) \quad (4)$$

In Choice III,

$$U_{LES} = 0.32\ln(q_{III1} - 2.47) + 0.33\ln(q_{III2} - 3.45) + 0.35\ln(q_{III3} - 0.55) \quad (5)$$

$$U_{Est} = 0.26\ln(q_{III1} - 2.39) + 0.39\ln(q_{III2} - 2.40) + 0.34\ln(q_{III3} - 0.55) \quad (6)$$

To get more visual comparisons, Tables 6-8 collect the comparisons between U_{LES} and U_{Est} for the values of b_{ij} and r_{ij} contained in (1)-(6).

Table 6 Comparisons: U_{LES} and U_{Est} in Choice I, Exp. 1 of C-Sample

	b_{I1}	b_{I2}	b_{I3}	r_{I1}	r_{I2}	r_{I3}
U_{LES}	0.31	0.33	0.36	2.95	2.56	3.49
U_{Est}	0.32	0.33	0.34	2.39	2.30	2.50

Wilcoxon test: Z=1.75, p=0.08

Table 7 Comparisons: U_{LES} and U_{Est} in Choice II, Exp. 1 of C-Sample

	b_{II1}	b_{II2}	b_{II3}	r_{II1}	r_{II2}	r_{II3}
U_{LES}	0.32	0.41	0.28	0.96	3.11	2.79
U_{Est}	0.33	0.41	0.26	0.96	2.40	2.50

Wilcoxon test: Z=1.46, p=0.14

Table 8 Comparisons: U_{LES} and U_{Est} in Choice III, Exp. 1 of C-Sample

	b_{III1}	b_{III2}	b_{III3}	r_{III1}	r_{III2}	r_{III3}
U_{LES}	0.32	0.33	0.35	1.81	2.35	0.55
U_{Est}	0.26	0.39	0.34	2.39	2.40	0.55

Wilcoxon test: Z=0.81, p=0.42

The above comparisons reveal an approximate agreement between U_{LES} and U_{Est} in Choices I-III, and obviously support the utility maximization hypothesis in Exp. 1 of C-Sample. The average relative error ($\Sigma|U_{LES}$'s-U_{Est}'s$|\div\Sigma U_{LES}$'s) is 0.06 for b_{ij}, and 0.17 for r_{ij}.

2.2 Results in Exp. 1 of S-Sample

121 subjects participated in S-Sample, and among them, 41 delivered valid data in Choice I, 40 in Choice II, and 39 in Choice III, in Exp. 1.

By solving LES in the category data of Session A, each U_{LES} for Choices I, II, and III in Exp. 1 of S-Sample is derived as (7), (9), and (11) (for details please see Part 4 of Supplemental Files).

By the similar way to that deriving U_{Est} for C-Sample, each U_{Est} for Choices I, II, and III in Exp. 1 of S-Sample is derived as (8), (10), and (12) (for details please see Part 4 of Supplemental Files):

In Choice I,

$$U_{LES} = 0.46\ln(q_{I1} - 2.43) + 0.13\ln(q_{I2} - 2.97) + 0.38\ln(q_{I3} - 2.61) \quad (7)$$

$$U_{Est} = 0.49\ln(q_{I1} - 2.21) + 0.17\ln(q_{I2} - 2.97) + 0.34\ln(q_{I3} - 2.61) \quad (8)$$

In Choice II,

$$U_{LES} = 0.39\ln(q_{II1} - 3.14) + 0.25\ln(q_{II2} - 6.04) + 0.36\ln(q_{II3} - 3.98) \quad (9)$$

$$U_{Est} = 0.09\ln(q_{II1} - 2.90) + 0.51\ln(q_{II2} - 2.93) + 0.40\ln(q_{II3} - 2.61) \quad (10)$$

In Choice III,

$$U_{LES} = 0.41\ln(q_{III1} - 1.98) + 0.38\ln(q_{III2} - 2.93) + 0.20\ln(q_{III3} - 1.72) \quad (11)$$

$$U_{Est} = 0.42\ln(q_{III1} - 2.21) + 0.35\ln(q_{III2} - 2.93) + 0.12\ln(q_{III3} - 1.72) \quad (12)$$

Table 9 Comparisons: U_{LES} and U_{Est} in Choice I, Exp. 1 of S-Sample

	b_{I1}	b_{I2}	b_{I3}	r_{I1}	r_{I2}	r_{I3}
U_{LES}	0.46	0.13	0.38	2.43	2.97	2.61
U_{Est}	0.49	0.17	0.34	2.21	2.97	2.61

Wilcoxon test: Z=0.55, p=0.58

Table 10 Comparisons: U_{LES} and U_{Est} in Choice II, Exp. 1 of S-Sample

	b_{I1}	b_{II2}	b_{II3}	r_{II1}	r_{II2}	r_{II3}
U_{LES}	***0.39***	***0.25***	0.37	2.95	***6.04***	3.98
U_{Est}	***0.09***	***0.51***	0.40	2.90	***2.93***	2.61

Wilcoxon test: Z=1.36, p=0.17

Table 11 Comparisons: U_{LES} and U_{Est} in Choice III, Exp. 1 of S-Sample

	b_{III1}	b_{III2}	b_{III3}	r_{III1}	r_{III2}	r_{III3}
U_{LES}	0.38	0.41	0.20	1.98	2.93	1.72
U_{Est}	0.42	0.35	0.22	2.21	2.93	1.72

Wilcoxon test: Z=0.73, p=0.47

Tables 9-11 collect the comparisons between U_{LES} and U_{Est} for the values of b_{ij} and r_{ij} contained in (7)-(12).

Among the eighteen pairwise comparisons shown in Tables 9-11, except of three indicating disagreements (the difference is larger than 40%) in Table 10 (indicated by italic and boldfaced figures), fifteen parameter comparisons present an approximate agreement between U_{LES} and U_{Est}. About 83% of the comparisons support the utility maximization hypothesis in Exp. 1 of S-Sample. After omitting the three disagreeing comparisons in Table 10, the average relative error for the rest fifteen comparisons is 0.09 for b_{ij}, and 0.10 for r_{ij}.

The comparisons between U_{LES} and U_{Est} in Tables 9 and 11 support the utility maximization in Choices I and III of Exp. 1 for S-Sample. But in Choice II (Table 10), the disagreeing ones occupy a half of the comparisons. Choice II in Exp. 1 of S-Sample fails in the test.

2.3 Results in Exp. 2 of S-Sample

Among 121 subjects in S-Sample, 28 delivered valid data in Choices I, II, and III of Exp. 2. However, the calculation outcome for the estimation of Choice I diverges, and leads to Choice I failing in the test on Exp. 2 (see Part 5 of Supplemental Files). Thus, only Choices II and III in Exp. 2 will be tested below.

By solving LES in the category data of Session A, each U_{LES} for Choices II and III in Exp. 2 of S-Sample is derived as (13) and (15) (for details please see Part 5 of Supplemental Files).

By the similar way to that deriving U_{Est} for C-Sample, each U_{Est} for Choices II and III in Exp. 2 of S-Sample is derived as (14) and (16) (for details please see Part 5 of Supplemental Files):

In Choice II,

$$U_{LES} = 0.58\ln(q_{II1} - 2.59) + 0.18\ln(q_{II2} - 0.74) + 0.24\ln(q_{II3} - 0.77) \quad (13)$$

$$U_{Est} = 0.63\ln(q_{II1} - 2.59) + 0.13\ln(q_{II2} - 0.90) + 0.24\ln(q_{II3} - 0.77) \quad (14)$$

In Choice III,

$$U_{LES} = 0.52\ln(q_{III1} - 3.91) + 0.12\ln(q_{III2} - 1.46) + 0.34\ln(q_{III3} - 1.72) \quad (15)$$

$$U_{Est} = 0.50\ln(q_{III1} - 3.25) + 0.26\ln(q_{III2} - 0.90) + 0.24\ln(q_{III3} - 1.72) \quad (16)$$

Tables 12 and 13 collect the comparisons between U_{LES} and U_{Est} for the values of b_{ij} and r_{ij} contained in (13)-(16).

Table 12 Comparisons: U_{LES} and U_{Est} in Choice II of Exp. 2

	b_{II1}	b_{II2}	b_{II3}	r_{II1}	r_{II2}	r_{II3}
U_{LES}	0.58	0.18	0.24	2.59	0.74	0.77
U_{Est}	0.63	0.13	0.24	2.59	0.90	0.77

Wilcoxon test: Z=0.82, p=0.58

Table 13 Comparisons: U_{LES} and U_{Est} in Choice III of Exp. 2

	b_{III1}	b_{III2}	b_{III3}	r_{III1}	r_{III2}	r_{III3}
U_{LES}	0.52	***0.12***	0.34	3.91	1.46	1.72
U_{Est}	0.50	***0.26***	0.24	3.25	0.90	1.72

Wilcoxon test: Z=1.21, p=0.23

Except of the comparison of b_{III2} indicating a disagreement (the difference is larger than 40%) in Table 13, overall, they also reveal an approximate agreement between U_{LES} and U_{Est} in Choices II and III respectively, and thus support the utility maximization hypothesis in Exp. 2 of S-Sample. After omitting the disagreeing comparison of b_{III2} in Table 13, the average relative error is 0.12 for b_{ij}, and 0.14 for r_{ij}.

2.4 Summaries for the experimental results

This is an acceptable experimental test with an acceptable approximate level of $R^2 \geq 0.97$ for the curve regression, but not an optimal experimental test that requires the maximal value of R^2 for the curve regression.

In eight sets of tests (Tables 6-13), except of a Choice II in Exp. 1 (Table 10), the results reveal the approximate but systematic agreements between the paired U_{LES} and U_{Est} at the relative error levels 0.06-0.17 for parameter comparisons. The eight sets of tests contain forty-eight pairs of parameter comparisons in sum, and among them, forty-four pairs deliver agreeing outcomes, occupying a proportion of about 92%, four pairs deliver disagreeing results, occupying a proportion of about 8%. Overall, the experimental results definitively support the utility maximization hypothesis.

As far as the experiments have revealed, the utility maximization holds for both Exp. 1 using substitutable goods and Exp. 2 using un-substitutable goods. The utility maximization appears robustness in subjects' choice behaviors in the two kinds of commodity bundles.

In the above tests, cross affections between different kinds of kernels in a choice bundle were naturally reflected in U_{LES} but, to simplify the experimental procedure, these affections were not taken into account in U_{Est}. U_{Est} was determined by the measures of utility scales that only relied on the assigned quantities and unit prices but were irrelevant to the cross affections between different kinds of kernels in a choice bundle. For example, in Choices II and III of Exp. 1 or Exp. 2, r_{II2} and r_{III2} in corresponding U_{Est} are the same because their assigned prices are the same in Choices II and III. It should result errors in the comparisons between U_{LES} and U_{Est} in these cases. Among four disagreements (see the italic and boldfaced figures in Tables 10 and 13), three appear in the comparisons of r_{II2} or b_{II2}, which are directly or indirectly determined through utility scales delivering the estimated values of r_{II2} or r_{III2}. Therefore, it can be expected that if these cross affections were taken into account, the agreement in the test would be improved.

In the utility-scale estimate, an easy and feasible treatment for incorporating with the cross affections between different kinds of kernels in a choice bundle is that put the unit price information of other two kinds of kernels in the same choice bundle together with the inquired kernel in an inquiry card, and let the subject report his bid unit price by comparing with the other two kinds of kernels in the same choice bundle. For example, in Choice II, the contents of the inquiry card for "the almond with offered unit price 0.30/bag and assigned quantity 6" are changed into "the unit price for pistachio is 0.70/bag and for cashew nut is 0.50/bag. If you are asked to buy 6 bags of almond with offered unit price 0.30/bag, compared with the unit prices of pistachio and cashew nut and the offered unit price for almond, your bid unit price will be (   )". The probing experiments offered some evidences implying that the precision of U_{Est} would be improved greatly in such a treatment. However, the probing experiment also showed that the valid rate would greatly lower to about 6% if this change were introduced in the inquiry card in Exp. 1. The experiment may weaken its representativeness as a measure to normal purchase behaviors at so low a valid rate. The causes of low valid rate may include: 1) in this case, a subject are asked to estimate all nine utility scales that contain forty-five inquiry cards in sum, but usually the valid rate evidently gets to decrease when the number of inquiry cards is above twenty five; and 2) too many comparisons in the inquiry cards tire the subject.

Another way to improve the test may be that discriminate the preference types by Choice I, and then, respectively test the subsamples basing on different preference categories. It will greatly enhance the precise of U_{LES} and U_{Est} but require a very big sample.

As the first experimental test of utility maximization and a methodological exploration, I finally chose the simplified program but not the more precise one so that we can concentrate on the fundamental issues. In fact, the simplified experiments have contributed valuable clues to evaluate the test of utility maximization. With the mentioned imperfectness of the simplified test, this paper is of course only an initial probe but, meanwhile, an effective new beginning in the field of utility maximization test.

3 Concluding remarks

3.1 Major findings

The findings in the experimental test can be interpreted from three aspects.

First, we can test the utility maximization by the experiment, and the experimental results support the cardinal utility maximization. It is therefore concluded that the cardinal marginal utility theory has found its empirical foundation in an explicit experimental procedure. It is not only an inspiring evidence but also a methodological progress for the utility maximization thought. Even though it is late for more than one-hundred-fifty years, after all, Gossen's genius has been eventually combined with and preliminarily supported by the experimental observations.

Second, the linear combination of measurable logarithmic laws for economic judgments is a proper operational definition of the Klein-Rubin utility function, furthermore, the utility scaling approach following psychophysics paradigm (e.g., He, 2011; Galanter, 1962, 1990) offers an appropriate measure for the cardinal utility theory. The present study further confirms the conclusion presented in He, 2011: Utility is the subjective quantities of commodities in the utility maximization of LES.

Finally, economics should re-evaluate cardinal utility theory on the basis of behavioral observations. The cardinal utility maximization is not only measurable but also more measurable than the ordinal utility maximization. In evident, for multi-commodity choices the cardinal utility maximization can be more clearly, rigorously, and effectively tested in a psychophysical-econometric paradigm than the ordinal one in a pair-wise comparison way that may be disordered by elicitation effects, preference reversals, and etc (e.g., Fredrick and Fischhoff, 1998; Slovic and Lichtenstein, 1983). A new combining point associating positive behavioral studies with traditional theoretical analyses has emerged, in which classical and contemporary economic thoughts will together contribute their wisdoms on a common stage of positive theory.

3.2 Emotion utility and perception utility

Bentham's utility is described by pleasure, happiness, satisfaction, and so on, referring to a kind of emotion attribute, can be called "emotion utility"; combining the psychophysical analysis with the econometric modeling discussion, He, 2012 and the present study revealed that utility is the subjective quantity of commodity or evaluation, referring to a kind of perceptional attribute, can be called "perception utility". The utility research should deal with the two utility concepts but not solely Bentham's type.

A corollary derived from econometric models and the present study is that importance of the quantity perception exceeds the emotional evaluation in one's economic choices. Benthamists perhaps misunderstood an economic choice as an enjoyment choice. In an economic choice, such as purchase choice, exchange choice, and risk choice, the first determinant is "whether it is worth to pay", a comparison between subjective quantities, but not "whether I am pleasant" that is usually seen in an enjoyment choice, such as eating an apple or a bread, watching a football game or a movie, and accepting an unfair proposal or rejecting it, in which one seeks a physiological or psychological gain. The distinction between perception utility and emotion utility comes from and is in turn used to interpret the difference between economic choice and enjoyment choice. The utility analysis should base on the discrimination between economic choice and enjoyment choice.

Benthamists and econometricians had respectively worked in the two different domains long ago. The perception utility has broadly used in economic empirical and experimental studies to determine utility functions and value functions in various models, such as the above mentioned econometric models and risk-choice models (e.g., Tversky and Kahneman, 1992; Gonzalez and Wu, 1999). Is there the utility maximization model for the enjoyment choice?

The present study only discussed the Klein-Rubin utility maximization model in depth, are the utility functions contained in other econometric utility maximization models also construct-able and measurable in a perception utility framework?

Current ultimatum games used mixed utility judgments, could we conceive an ultimatum game mainly involving the emotion utility judgment or perception utility judgment to depart the emotion effect and perception effect in bargaining behaviors, in which unfair feelings (related with enjoyment choice) mixes with some monetary return (related with economic choice)? Furthermore, how do the emotion utility and perception utility affect the difference of responders' rejection behaviors between the moderate stake size and the very high stake size (Andersen et al., 2011)?

And so on.

We may just begin, and we will need more knowledge to clarify the attributes and functions of the emotion utility and perception utility.

References

Andersen, S., Ertac, S., Gneezy, U., Hoffman, M., and List, J. A. (2011). Stakes Matter in ultimatum games. *American Economic Review, 101*, 3427-3439.

Bentham, J. (1789). An Introduction to the Principle of Morals and Legislations. (reprinted) Oxford: Blackwell, 1948.

Breaut, K. D. (1983). Psychological measurement and the validity of the modern economic approach: A presentation of methods and preliminary experiments. *Social Science Research, 12*, 187-203.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review, 70*, 312-326.

Fredrick, S. and Fischhoff, B. (1998). Scope (in)sensitivity in elicited valuations. *Risk Decision and Policy, 3*, 109-123.

Galanter, E. (1962). The direct measurement of utility and subjective probability. *American Journal of Psychology, 75*, 208-220.

Galanter, E. (1990). Utility functions for nonmonetary events. *The American Journal of Psychology, 103*, 449-470.

Galanter, E. and Pliner, P. (1974). Cross-modality matching of money against other continua. In: Moskowitz, H., Sharf, B., and Stevens, J. C. (Eds), Sensation and Measurement: Papers in honor of S. Stevens. Netherland: Reidel. 65-76.

Galanter, E., Popper, R., and Perera, T. (1977). Annoyance scales for simulated VTOL and CTOL overflights. *Journal of the Acoustical Society of America, 8A* (Suppl.), 1.

Gonzalez, G. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*, 129-166.

Gossen, H. H. (1854). The Laws of Human Relations and the Rules of Human Action Derived Therefrom. Blitz, R. C. translated, Cambridge: MIT Press, 1983

He, Y. (2011). Psychophysical Interpretation for Utility Measures. Economics Discussion Papers, No 2011-50, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/ discussionpapers/2011-50

Houthakker, H. S. (1960). Additive Preferences. *Econometrica, 28*, 244-257.

Intriligator, M. D. (1980). Econometric Models, Techniques, and Applications. New Delhi: Prentice-Hall. 226-230.

Kahneman, D., Wakkar, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics, May* , 275-405.

Klein, L. R. and Rubin, H. (1947). A Constant-Utility Index of Cost of Living. *The Review of Economic Studies 15 (2)*, 78-83.

Liuch, C. (1973). The expended linear expenditure system. *European Economic Review, 4*, 21-32.

MacCrimmon, K. R. and Toda, M. (1969). The experimental determination of indifference curves. *Review of Economic Studies, 36*, 433-451.

Parker, S. and Schneider, B. (1988). Conjoint scaling of the utility of money using paired comparisons. *Social Science Research, 17*, 277-286.

Slovic, P. and Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review, 73*, 596-605.

Spiegel, H. W. (1991). The Growth of Economic Thought. Durham: Duke University Press. Chapter 22, Chapter 24.

Stigler, G. J., (1950). The development of utility theory, I. *The Journal of Political Economy 58 (4)*, 307-327; *58 (5)*, 373-396.

Stone, R. (1954). Linear expenditure system: An application to the pattern of British demand. *Economics Journal, 64* , 511-527.

Theil, H. (1965). The information approach to demand analysis. *Econometrica, 33*, 67-87.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297-324.