MARKETING SCIENCE
Vol. 27, No. 1, January–February 2008, pp. 15–25
ISSN 0732-2399 | EISSN 1526-548X | 08 | 2701 | 0015

DOI 10.1287/mksc.1070.0330
© 2008 INFORMS

Mental Accounting and Consumer Choice

Richard H. Thaler
Johnson Graduate School of Management, Cornell University, Ithaca, New York 14853, thaler@chicagogsb.edu

A new model of consumer behavior is developed using a hybrid of cognitive psychology and microeconomics. The development of the model starts with the mental coding of combinations of gains and losses using the prospect theory value function. Then the evaluation of purchases is modeled using the new concept of "transaction utility." The household budgeting process is also incorporated to complete the characterization of mental accounting. Several implications to marketing, particularly in the area of pricing, are developed.

This article was originally published in *Marketing Science*, Volume 4, Issue 3, pages 199–214, in 1985.

Key words: mental accounting; consumer choice; pricing
History: This paper was received June 1983 and was with the author for 2 revisions.

1. Introduction

Consider the following anecdotes:

1. Mr. and Mrs. L and Mr. and Mrs. H went on a fishing trip in the northwest and caught some salmon. They packed the fish and sent it home on an airline, but the fish were lost in transit. They received $300 from the airline. The couples take the money, go out to dinner and spend $225. They had never spent that much at a restaurant before.

2. Mr. X is up $50 in a monthly poker game. He has a queen high flush and calls a $10 bet. Mr. Y owns 100 shares of IBM which went up $\frac{1}{2}$ today and is even in the poker game. He has a king high flush but he folds. When X wins, Y thinks to himself, "If I had been up $50 I would have called too."

3. Mr. and Mrs. J have saved $15,000 toward their dream vacation home. They hope to buy the home in five years. The money earns 10% in a money market account. They just bought a new car for $11,000 which they financed with a three-year car loan at 15%.

4. Mr. S admires a $125 cashmere sweater at the department store. He declines to buy it, feeling that it is too extravagant. Later that month he receives the same sweater from his wife for a birthday present. He is very happy. Mr. and Mrs. S have only joint bank accounts.

All organizations, from General Motors down to single person households, have explicit and/or implicit accounting systems. The accounting systems often influence decisions in unexpected ways. This paper characterizes some aspects of the implicit mental accounting system used by individuals and households. The goal of the paper is to develop a richer theory of consumer behavior than standard economic theory. The new theory is capable of explaining (and predicting) the kinds of behavior illustrated by the four anecdotes above. Each of these anecdotes illustrate a type of behavior where a mental accounting system induces an individual to violate a simple economic principle. Example 1 violates the principle of fungibility. Money is not supposed to have labels attached to it. Yet the couples behaved the way they did because the $300 was put into both "windfall gain" and "food" accounts. The extravagant dinner would not have occurred had each couple received a yearly salary increase of $150, even though that would have been worth more in present value terms. Example 2 illustrates that accounts may be both topically and temporally specific. A player's behavior in a poker game is altered by his current position in that evening's game, but not by either his lifetime winnings or losings nor by some event allocated to a different account altogether such as a paper gain in the stock market. In example 3 the violation of fungibility (at obvious economic costs) is caused by the household's appreciation for their own self-control problems. They are afraid that if the vacation home account is drawn down it will not be repaid, while the bank will see to it that the car loan is paid off on schedule. Example 4 illustrates the curious fact that people tend to give as gifts items that the recipients would not buy for themselves, and that the recipients by and large approve of the strategy. As is shown in §4.3, this also violates a microeconomic principle.

The theory of consumer behavior to which the current theory is offered as a substitute is the standard economic theory of the consumer. That theory, of course, is based on normative principles. In fact, the paradigm of economic theory is to first characterize the solution to some problem, and then to assume the relevant agents (on average) act accordingly.

The decision problem which consumers are supposed to solve can be characterized in a simple fashion. Let $z = \{z_1, \ldots, z_i, \ldots, z_n\}$ be the vector of goods

available in the economy at prices given by the corresponding vector $p = \{p_1, \ldots, p_i, \ldots, p_n\}$. Let the consumer's utility function be defined as $U(z)$ and his income (or wealth) be given as I. Then the consumer should try to solve the following problem:

$$\max_z \; U(z)$$
$$\text{s.t.} \; \sum p_i z_i \le I.$$

Or, using Lagrange multipliers

$$\max_z U(z) - \lambda\left(\sum p_i z_i - I\right). \tag{1}$$

The first order conditions to this problem are, in essence, the economic theory of the consumer. Lancaster (1971) has extended the model by having utility depend on the characteristics of the goods. Similarly, Becker (1965) has introduced the role of time and other factors using the concept of household production. These extended theories are richer than the original model, and, as a result, have more to offer marketing. Nevertheless, the economic theory of the consumer, even so extended, has not found widespread application in marketing. Why not? One reason is that all such models omit virtually all marketing variables except price and product characteristics. Many marketing variables fall into the category that Tversky and Kahneman (1981) refer to as *framing*. These authors have shown that often choices depend on the way a problem is posed as much as on the objective features of a problem. Yet within economic theory, framing cannot alter behavior.

To help describe individual choice under uncertainty in a way capable of capturing "mere" framing effects as well as other anomalies, Kahneman and Tversky (1979) have developed "prospect theory" as an alternative to expected utility theory. Prospect theory's sole aim is to describe or predict behavior, not to characterize optimal behavior. Elsewhere (Thaler 1980), I have begun to develop a similar descriptive alternative to the deterministic economic theory of consumer choice. There I argue that consumers often fail to behave in accordance with the normative prescriptions of economic theory. For example, consumers often pay attention to sunk costs when they shouldn't, and underweight opportunity costs as compared to out-of-pocket costs.[1]

This paper uses the concept of mental accounting to move further toward a behaviorally based theory of consumer choice. Compared to the model in Equation (1) the alternative theory has three key features. First, the utility function $U(x)$ is replaced with the *value function* $v(\cdot)$ from prospect theory. The characteristics of this value function are described and then extended to apply to compound outcomes. Second, price is introduced directly into the value function using the concept of a *reference* price. The new concept *of transaction utility* is developed as a result. Third, the normative principle of *fungibility* is relaxed. Numerous marketing implications of the theory are derived. The theory is also used to explain some empirical puzzles.

2. Mental Arithmetic

2.1. The Value Function

The first step in describing the behavior of the representative consumer is to replace the utility function from economic theory with the psychologically richer *value function* used by Kahneman and Tversky. The assumed shape of the value function incorporates three important behavioral principles that are used repeatedly in what follows. First, the function $v(\cdot)$ is defined over perceived gains and losses relative to some natural reference point, rather than wealth or consumption as in the standard theory. This feature reflects the fact that people appear to respond more to perceived changes than to absolute levels. (The individual in this model can be thought of as a pleasure machine with gains yielding pleasure and losses yielding pain.) By using a reference point the theory also permits framing effects to affect choices. The framing of a problem often involves the suggestion of a particular reference point. Second, the value function is assumed to be concave for gains and convex for losses ($v''(x) < 0, x > 0$; $v''(x) > 0, x < 0$). This feature captures the basic psychophysics of quantity. The difference between \$10 and \$20 seems greater than the difference between \$110 and \$120, irrespective of the signs of the amounts in question. Third, the loss function is steeper than the gain function ($v(x) < -v(-x)$, $x > 0$). This notion that losses loom larger than gains captures what I have elsewhere called the endowment effect: people generally will demand more to sell an item they own than they would be willing to pay to acquire the same item (Thaler 1980).

2.2. Coding Gains and Losses

The prospect theory value function is defined over single, unidimensional outcomes. For the present analysis it is useful to extend the analysis to incorporate compound outcomes where each outcome is measured along the same dimension (say dollars).[2]

[1] These propositions have recently been tested and confirmed in extensive studies by Arkes and Blumer (1985), Gregory (1982), and Knetch and Sinden (1984).

[2] Kahneman and Tversky are currently working on the single outcome, multi-attribute case. It is also possible to deal with the compound multi-attribute case but things get very messy. Since this paper is trying to extend economic theory which assumes that all

The question is how does the joint outcome (x, y) get coded? Two possibilities are considered. The outcomes could be valued jointly as $v(x + y)$ in which case they will be said to be *integrated*. Alternatively they may be valued separately as $v(x) + v(y)$ in which case they are said to be *segregated*. The issue to be investigated is whether segregation or integration produces greater utility. The issue is interesting from three different perspectives. First, if a situation is sufficiently ambiguous how will individuals choose to code outcomes? To some extent people try to frame outcomes in whatever way makes them happiest.[3] Second, individuals may have preferences about how their life is organized. Would most people rather have a salary of \$30,000 and a (certain) bonus of \$5,000 or a salary of \$35,000? Third, and most relevant to marketing, how would a seller want to describe (frame) the characteristics of a transaction? Which attributes should be combined and which should be separated? The analysis which follows can be applied to any of these perspectives.

For the joint outcome (x, y) there are four possible combinations to consider:

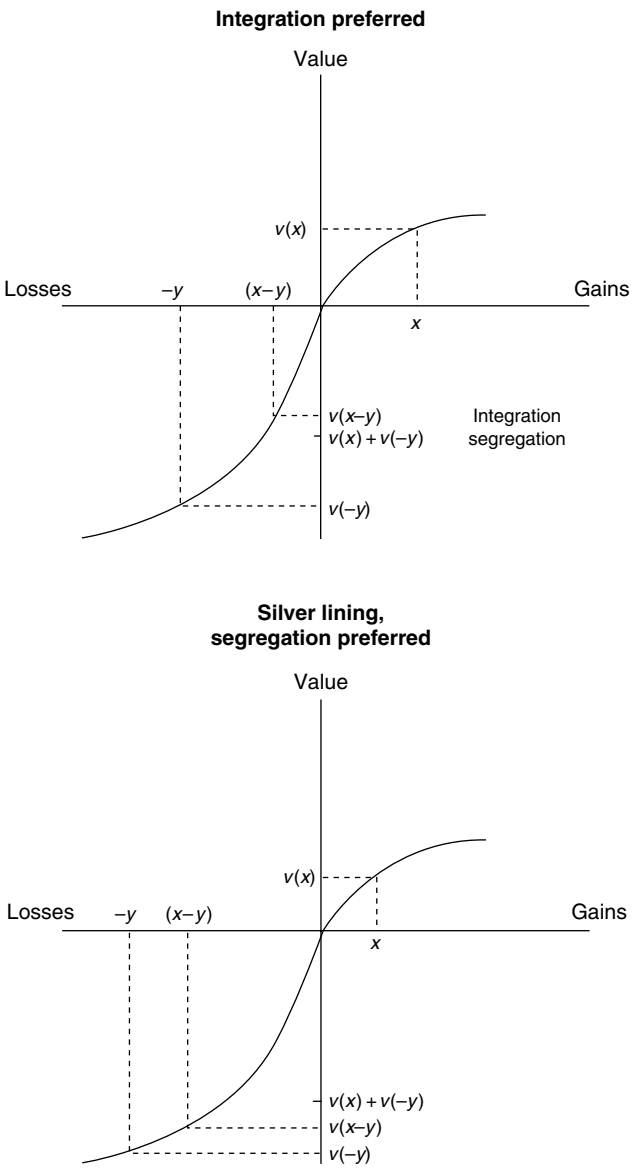
1. Multiple Gains. Let $x > 0$ and $y > 0$.[4] Since v is concave $v(x) + v(y) > v(x + y)$, so segregation is preferred. Moral: don't wrap all the Christmas presents in one box.

2. Multiple Losses. Let the outcomes be $-x$ and $-y$ where x and y are still positive. Then since $v(-x) + v(-y) < v(-(x + y))$ integration is preferred. For example, one desirable feature of credit cards is that they pool many small losses into one larger loss and in so doing reduce the total value lost.

3. Mixed Gain. Consider the outcome $(x, -y)$ where $x > y$ so there is a net gain. Here $v(x)+v(-y) < v(x - y)$ so integration is preferred. In fact, since the loss function is steeper than the gain function, it is possible that $v(x) + v(-y) < 0$ while $v(x - y)$ must be positive since $x > y$ by assumption. Thus, for mixed gains integration amounts to *cancellation*. Notice that all voluntarily executed trades fall into this category.

4. Mixed Loss. Consider the outcome $(x, -y)$ where $x < y$, a net loss. In this case we cannot determine without further information whether $v(x) + v(-y) \gtrless v(x - y)$. This is illustrated in Figure 1. Segregation is preferred if $v(x) > v(x-y)-v(-y)$. This is more likely the smaller is x relative to y. Intuitively, with a large loss and a small gain, e.g., (\$40, −\$6,000) segregation is preferred since v is relatively flat near −6,000. This will be referred to as the "silver lining" principle. On the other hand, for (\$40, −\$50) integration is probably preferred since the gain of the \$40 is likely to be valued less than the reduction of the loss from \$50 to \$10, nearly a case of cancellation.

Figure 1 Segregation or Integration of Gains and Losses



2.3. Evidence on Segregation and Integration

The previous analysis can be summarized by four principles: (a) segregate gains, (b) integrate losses, (c) cancel losses against larger gains, (d) segregate

outcomes can be collapsed into a single index (utils or money) sticking to the one-dimensional case seems like a reasonable first step.

[3] This is illustrated by the following true story. A group of friends who play poker together regularly had an outing in which they played poker in a large recreational vehicle while going to and from a race track. There were significant asymmetries in the way people (honestly) reported their winnings and losings from the two poker games and racetrack bets. Whether the outcomes were reported together or separately could largely be explained by the analysis that follows.

[4] For simplicity I will deal only with two-outcome events, but the principles generalize to cases with several outcomes.

"silver linings." To see whether these principles coincided with the intuition of others, a small experiment was conducted using 87 students in an undergraduate statistics class at Cornell University. The idea was to present subjects with pairs of outcomes either segregated or integrated and to ask them which frame was preferable. Four scenarios were used, one corresponding to each of the above principles.

The instructions given to the students were:

> Below you will find four pairs of scenarios. In each case two events occur in Mr. A's life and one event occurs in Mr. B's life. You are asked to judge whether Mr. A or Mr. B is happier. Would most people rather be A or B? If you think the two scenarios are emotionally equivalent, check "no difference." In all cases the events are intended to be financially equivalent.

The four items used and the number of responses of each type follow.

> 1. Mr. A was given tickets to lotteries involving the World Series. He won \$50 in one lottery and \$25 in the other.
>
> Mr. B was given a ticket to a single, larger World Series lottery. He won \$75.
>
> Who was happier? 56 A 16 B 15 no difference
>
> 2. Mr. A received a letter from the IRS saying that he made a minor arithmetical mistake on his tax return and owed \$100. He received a similar letter the same day from his state income tax authority saying he owed \$50. There were no other repercussions from either mistake.
>
> Mr. B received a letter from the IRS saying that he made a minor arithmetical mistake on his tax return and owed \$150. There were no other repercussions from his mistake.
>
> Who was more upset? 66 A 14 B 7 no difference
>
> 3. Mr. A bought his first New York State lottery ticket and won \$100. Also, in a freak accident, he damaged the rug in his apartment and had to pay the landlord \$80.
>
> Mr. B bought his first New York State lottery ticket and won \$20.
>
> Who was happier? 22 A 61 B 4 no difference
>
> 4. Mr. A's car was damaged in a parking lot. He had to spend \$200 to repair the damage. The same day the car was damaged he won \$25 in the office football pool.
>
> Mr. B's car was damaged in a parking lot. He had to spend \$175 to repair the damage.
>
> Who was more upset? 19 A 63 B 5 no difference

For each item, a large majority of the subjects chose in a manner predicted by the theory.[5]

[5] Two caveats must be noted here. First, the analysis does not extend directly to the multi-attnbute (or multiple account) case. It is often cognitively impossible to integrate across accounts. Thus winning \$100 does not cancel a toothache. Second, even within the same account, individuals may be unable to integrate two losses that are framed separately. See Johnson and Thaler (1985).

2.4. Reference Outcomes

Suppose an individual is expecting some outcome x and instead obtains $x + \Delta x$. Define this as a reference outcome $(x + \Delta x : x)$. The question then arises how to value such an outcome. Assume that the expected outcome was fully anticipated and assimilated. This implies that $v(x:x) = 0$. A person who opens his monthly pay envelope and finds it to be the usual amount is unaffected. However, when $\Delta x \neq 0$ there is a choice of ways to frame the outcome corresponding to the segregation/integration analysis of simple compound outcomes. With reference outcomes the choice involves whether to value the unexpected component Δx alone (segregation) or in conjunction with the expected component (integration). An example, similar to those above, illustrates the difference:

- Mr. A expected a Christmas bonus of \$300. He received his check and the amount was indeed \$300. A week later he received a note saying that there had been an error in this bonus check. The check was \$50 too high. He must return the \$50.
- Mr. B expected a Christmas bonus of \$300. He received his check and found it was for \$250.

It is clear who is more upset in this story. Mr. A had his loss segregated and it would inevitably be coded as a loss of \$50. Mr. B's outcome can be integrated by viewing the news as a reduction in a gain $-[v(300) - v(250)]$. When the situation is structured in a neutral or ambiguous manner then the same four principles determine whether segregation or integration is preferred:

(1) An increase in a gain should be segregated.

(2) An increase in (the absolute value of) a loss should be integrated.

(3) A decrease in a gain should be integrated (cancellation).

(4) A small reduction in (the absolute value of) a loss should be segregated (silver lining).

The concept of a reference outcome is used below to model a buyer's reaction to a market price that differs from the price he expected.

3. Transaction Utility Theory

In the context of the pleasure machine metaphor suggested earlier, the previous section can be thought of as a description of the hard wiring. The machine responds to perceived gains and losses in the way described. The next step in the analysis is to use this structure to analyze transactions. A two-stage process is proposed. First, individuals evaluate potential transactions. Second, they approve or disapprove of each potential transaction. The first stage is a judgment process while the second is a decision process. They are analyzed in turn.

3.1. Evaluating Transactions

Consider the following excerpt from a movie review:

> My sister just found out that for a $235 per month sublet she shares with another woman, she pays $185 per month. The other woman justifies her $50 per month rent two ways: one, she is doing my sister a favor letting her live there given the housing situation in New York City, and, two, everyone with a room to sublet in NYC will cheat her at least as badly. Her reasons are undeniably true, and that makes them quadruply disgusting. (*Cornell Daily Sun,* Feb. 21, 1983)

Notice that the writer's sister is presumably getting a good value for her money (the room is worth $185 per month) but is still unhappy. To incorporate this aspect of the psychology of buying into the model, two kinds of utility are postulated: *acquisition utility* and *transaction utility*. The former depends on the value of the good received compared to the outlay, the latter depends solely on the perceived merits of the "deal."

For the analysis that follows, three price concepts are used. First, define p as the actual price charged for some good z. Then for some individual, define $\bar{p}$ as the *value equivalent* of z, that is, the amount of money which would leave the individual indifferent between receiving $\bar{p}$ or z as a gift.[6] Finally, let p^* be called the *reference* price for z. The reference price is an expected or "just" price for z. (More on p^* momentarily.)

Now define acquisition utility as the value of the compound outcome $(z, -p) = (\bar{p}, -p)$. This is designated as $v(\bar{p}, -p)$. Acquisition utility is the net utility that accrues from the trade of p to obtain z (which is valued at $\bar{p}$). Since $v(\bar{p}, -p)$ will generally be coded as the integrated outcome $v(\bar{p} - p)$, the cost of the good is not treated as a loss. Given the steepness of the loss function near the reference point, it is hedonically inefficient to code costs as losses, especially for routine transactions.

The measure of transaction utility depends on the price the individual pays compared to some reference price, p^*. Formally, it is defined as the reference outcome $v(-p : -p^*)$, that is, the value of paying p when the expected or reference price is p^*. Total utility from a purchase is just the sum of acquisition utility and transaction utility.[7] Thus the value of buying good z at price p with reference price p^* is defined as $w(z, p, p^*)$ where:

$$w(z, p, p^*) = v(\bar{p}, -p) + v(-p: -p^*). \tag{2}$$

Little has been said as to the determinants of p^*. The most important factor in determining p^* is fairness. Fairness, in turn, depends in large part on cost to the seller. This is illustrated by the following three questionnaires administered to first-year MBA students. (The phrases in brackets differed across the three groups.)

> Imagine that you are going to a sold-out Cornell hockey playoff game, and you have an extra ticket to sell or give away. The price marked on the ticket is $5 (but you were given your tickets for free by a friend) [which is what you paid for each ticket] {but you paid $10 each for your tickets when you bought them from another student}. You get to the game early to make sure you get rid of the ticket. An informal survey of people selling tickets indicates that the going price is $5. You find someone who wants the ticket and takes out his wallet to pay you. He asks how much you want for the ticket. Assume that there is no law against charging a price higher than that marked on the ticket. What price do you ask for if
>
> 1. he is a friend ——
> 2. he is a stranger ——
>
> What would you have said if instead you found the going market price was $10?
>
> 3. friend ——
> 4. stranger ——

The idea behind the questionnaire was that the price people would charge a friend would be a good proxy for their estimate of a fair price. For each question, three prices were available as possible anchors upon which people could base their answers: the price marked on the ticket, the market price, and the price paid by the seller, i.e., cost. As can be seen in Table 1, the modal answers in the friend condition are equal to the seller's costs except in the unusual case

Table 1 Percent of Subjects Giving Common Answers to Hockey Ticket Question

		Friend				Stranger			
Cost	Market value	0	5	10	Other	0	5	10	Other
	$N = 31$								
0	5	<u>68</u>	26	3	3	6	<u>77</u>	10	6
0	10	<u>65</u>	26	6	3	6	16	<u>58</u>	19
	$N = 28$								
5	5	14	<u>79</u>	0	7	0	<u>79</u>	7	14
5	10	7	<u>79</u>	4	9	0	14	<u>57</u>	29
	$N = 26$								
10	5	0	<u>69</u>	23	8	0	42	<u>46</u>	12
10	10	0	15	<u>69</u>	15	0	0	<u>73</u>	27

Note. Modal answer is underlined.

[6] In the standard theory, $\bar{p}$ equals the reservation price, the maximum the individual would pay. In this theory, $\bar{p}$ can differ from the reservation price because of positive or negative transaction utility. Acquisition utility is comparable in principle to consumer surplus.

[7] A more general formulation would be to allow differing weights on the two terms in (2). For example, Equation (2) could be written as

$$w(z, p, p^*) = v(p, -p) + \beta v(-p: -p^*),$$

where β is the weight given to transaction utility. If $\beta = 0$ then the standard theory applies. Pathological bargain hunters would have $\beta > 1$. This generalization was suggested by Jonathan Baron.

where seller's cost was above market price. In contrast, the modal answers in the stranger condition are equal to market price with the same lone exception. The implication of this is that buyers' perceptions of a seller's costs will strongly influence their judgments about what price is fair, and this in turn influences their value for p^*.

The next questionnaire, given to those participants in an executive development program who said they were regular beer drinkers, shows how transaction utility can influence willingness to pay (and therefore demand).

Consider the following scenario:

> You are lying on the beach on a hot day. All you have to drink is ice water. For the last hour you have been thinking about how much you would enjoy a nice cold bottle of your favorite brand of beer. A companion gets up to go make a phone call and offers to bring back a beer from the only nearby place where beer is sold (a fancy resort hotel) [a small, run-down grocery store]. He says that the beer might be expensive and so asks how much you are willing to pay for the beer. He says that he will buy the beer if it costs as much or less than the price you state. But if it costs more than the price you state he will not buy it. You trust your friend, and there is no possibility of bargaining with (the bartender) [store owner]. What price do you tell him?

The results from this survey were dramatic. The median price given in the fancy resort hotel version was \$2.65 while the median for the small run-down grocery store version was \$1.50. This difference occurs despite the following three features of this example:

1. In both versions the ultimate consumption act is the same—drinking one beer on the beach. The beer is the same in each case.

2. There is no possibility of strategic behavior in stating the reservation price.[8]

3. No "atmosphere" is consumed by the respondent.

The explanation offered for these choices is based on the concept of transaction utility. (Acquisition utility is constant between the two cases.) While paying \$2.50 for a beer is an expected annoyance at the resort hotel, it would be considered an outrageous "rip-off" in a grocery store. Paying \$2.50 a bottle is \$15.00 a six-pack, considerably above the reference price.

3.2. Purchase Decisions—Multiple Accounts

The introduction of $w(\cdot)$ as the purchase evaluation device requires additional changes to the standard theory described in the introduction. Since $w(\cdot)$ is defined over individual transactions it is convenient to give each unit of a specific good its own label. Optimization would then require the individual to select the set of purchases that would maximize $\sum w(\cdot)$ subject to the budget constraint $\sum p_i z_i \leq I$ where I is income. A solution to this integer programming problem would be to make purchases if and only if

$$\frac{w(z_i, p_i, p_i^*)}{p_i} \geq k \tag{3}$$

where k is a constant that serves a role similar to that of the Lagrange multiplier in the standard formulation.

Notice that if k is selected optimally then (3) can be applied sequentially without any explicit consideration of opportunity costs. This sort of sequential analysis seems to be a good description of behavior.[9] First, the consumer responds to local temporal budget constraints. That is, the budget constraint that most influences behavior is the current income flow rather than the present value of lifetime wealth. For many families, the most relevant time horizon is the month since many regular bills tend to be monthly. Thus, the budgeting process, either implicit or explicit, tends to occur on a month-to-month basis. Second, expenditures tend to be grouped into categories. Potential expenditures are then considered within their category. (Families that take their monthly pay and put it into various use-specific envelopes to be allocated during the month are explicitly behaving in the manner described here. Most families simply use a less explicit procedure.) The tendency to group purchases by category can violate the economic principle of fungibility.

Given the existence of time and category specific budget constraints, the consumer evaluates purchases as situations arise. For example, suppose a couple is called by friends who suggest going out to dinner on Saturday night at a particular restaurant. The couple would have to decide whether such an expenditure would violate either the monthly or the entertainment constraints. Formally, the decision process can be modelled by saying the consumer will buy a good z at price p if

$$\frac{w(z, p, p^*)}{p} > k_{it}$$

where k_{it} is the budget constraint for category i in time period t.

[8] The question is what economists would call "incentive compatable". The respondent's best strategy is to state his or her true reservation price. Subjects given extensive explanations of this feature nevertheless still display a large disparity in answers to the two versions of the problem.

[9] The model that follows is based, in part, on some extensive, open-ended interviews of families conducted in 1982. The families were asked detailed questions about how they regulate their day-to-day expenditures, and what they have done in various specific situations such as those involving a large windfall gain or loss.

Of course, global optimization would lead all the k_{it}'s to be equal which would render irrelevant the budgeting process described here. However, there is evidence that individuals do not act as if all the k's were equal. As discussed elsewhere (Thaler and Shefrin 1981), individuals face self-control problems in regulating eating, drinking, smoking, and consumption generally. The whole mental accounting apparatus being presented here can be thought of as part of an individual's solution to these problems. For example, the rule of thumb to restrict monthly expenditures to no more than monthly income is clearly nonoptimal. Yet, when borrowing is permitted as a method of smoothing out monthly k's, some families find themselves heavily in debt. Restrictions on borrowing are then adopted as a second-best strategy. The technology of self-control often implies outright prohibitions because allowing a little bit eventually leads to excesses. (Although smoking cigarettes is undoubtedly subject to diminishing marginal utility, almost no one smokes between 1 and 5 cigarettes a day. That level, while probably preferred by many smokers and former smokers to either zero or 20, is just unattainable.)

Unusually high category specific k's are most likely to be observed for goods that are particularly seductive or addictive. Unusually low k's are observed for goods viewed to be particularly desirable in the long run such as exercise or education. Application of these ideas to gift giving behavior is discussed below.

4. Marketing Implications

The previous sections have outlined a theory and presented some survey evidence to support its various components. The following sections discuss the implications of this theory to marketing. There are two types of implications presented here. First, the theory is used to explain some empirical puzzles such as why some markets fail to clear. Second, some advice for sellers is derived, based on the presumption that buyers behave according to the theory. This advice is illustrated with actual examples. The implications are derived from each of the three main components of the theory: compounding principles, transaction utility, and budgetary rules.

4.1. Compounding Rule Implications

This section will illustrate how the results from the analysis of mental arithmetic can influence marketing decisions either in the design or products or in the choice of how products are described. The results of §2.2 can be summarized by two principles: segregate gains and integrate losses. Each principle also has a corollary: segregate "silver linings" (small gains combined with large losses) and integrate (or cancel) losses when combined with larger gains.

Segregate Gains. The basic principle of segregating gains is simple and needs little elaboration or illustration. When a seller has a product with more than one dimension it is desirable to have each dimension evaluated separately. The most vivid examples of this are the late-night television advertisements for kitchen utensils. The principle is used at two levels. First, each of the items sold is said to have a multitude of uses, each of which is demonstrated. Second, several "bonus" items are included "if you call right now." These ads all seem to use the same basic format and are almost a caricature of the segregation principle.

The silver lining principle can be used to understand the widespread use of rebates as a form of price promotion. It is generally believed that rebates were first widely used because of the threat of government price controls. By having an explicitly temporary rebate it was hoped that the old price would be the one for which new regulations might apply. Rebates for small items have the additional feature that not all consumers send in the form to collect the rebate. However, rebates continue to be widely used in the automobile industry in spite of the following considerations:

(1) Price controls seem very unlikely during the Reagan administration, especially with inflation receding.

(2) All purchasers claim the rebate since it is processed by the dealer and is worth several hundred dollars.

(3) Consumers must pay sales tax on the rebate. This can raise the cost of the purchase by 8% of the rebate in New York City. While this is not a large amount of money relative to the price of the car, it nonetheless provides an incentive to adopt the seemingly equivalent procedure of announcing a temporary sale.

Why then are rebates used in the automobile industry? The silver lining principle suggest one reason. A rebate strongly suggests segregating the saving. This can be further strengthened for those consumers who elect to have the rebate mailed to them from the corporate headquarters rather than applied to the down payment.[10]

Integrate Losses. When possible, consumers would prefer to integrate losses. The concavity of the loss function implies that adding \$50 less to an existing \$1,000 loss will have little impact if it is integrated. This means that sellers have a distinct advantage

[10] In the first year that rebates were widely used, one manufacturer reported (to me in personal communication) that about one-third of the customers receiving rebates chose the option of having the check sent separately. My impression is that this has become less common as rebates have become widespread.

in selling something if its cost can be added on to another larger purchase. Adding options to an automobile or house purchase are classic, well-known examples. More generally, whenever a seller is dealing with an expensive item the seller should consider whether additional options can be created since the buyers will have temporarily inelastic demands for these options. The principle also applies to insurance purchases. Insurance companies frequently sell riders to home or car insurance policies that are attractive (I believe) only because of this principle. One company has been advertising a "paint spill" rider for its homeowner policy. (This is apparently designed for do-it-yourselfers who have not yet discovered drop cloths.) Another example is credit card insurance which pays for the first \$50 of charges against a credit card if it is lost or stolen. (Claims over \$50 are absorbed by the credit card company.)

The principle of cancellation states that losses will be integrated with larger gains where plausible. The best example of this is withholding from paychecks. In the present framework the least aversive type of loss is the reduction of a large gain. This concept seems to have been widely applied by governments. Income taxes would be perceived as much more aversive (in addition to being harder to collect) if the whole tax bill were due in April. The implication for sellers is that every effort should be made to set up a payroll withdrawal payment option. Probably the best way to market dental insurance, for example, would be to sell it as an option to group health insurance through employers. If the employee already pays for some share of the health insurance then the extra premium would be framed as an increase in an existing deduction; this is the ultimate arrangement for a seller.

4.2. Transaction Utility Implications

Sellouts and Scalping. The tool in the economist's bag in which most economists place the greatest trust is the supply and demand analysis of simple commodity markets. The theory stipulates that prices adjust over time until supply equals demand. While the confidence put in that analysis is generally well founded, there are some markets which consistently fail to clear. One widely discussed example is labor markets where large numbers of unemployed workers coexist with wages that are not falling. Unemployment occurs because a price (the wage) is too high. Another set of markets features the opposite problem, prices that are too low. I refer to the class of goods and services for which demand exceeds supply: Cabbage Patch dolls in December 1983 and 1984, tickets to any Super Bowl, World Series, World Cup Final, Vladimir Horowitz or Rolling Stones concert, or even dinner reservations for 8:00 P.M. Saturday evening at the most popular restaurant in any major city. Why are these prices too low? Once the Cabbage Patch rage started, the going black market price for a doll was over \$100. Why did Coleco continue to sell the dolls it had at list price? Why did some discount stores sell their allotted number at less than list price? Tickets for the 1984 Super Bowl were selling on the black market for \$300 and up. Seats on the 50-yard line were worth considerably more. Why did the National Football League sell all of the tickets at the same \$60 price?

There are no satisfactory answers to these questions within the confines of standard microeconomic theory. In the case of the Super Bowl, the league surely does not need the extra publicity generated by the ticket scarcity. (The argument that long lines create publicity is sometimes given for why prices aren't higher during first week's showing of the latest *Star Wars* epic.) The ticket scarcity occurs every year so (unlike the Cabbage Patch Doll case) there is no possible surprise factor. Rather, it is quite clear that the league *knowingly* sets the prices "too low." Why?

The concept of transaction utility provides a coherent, parsimonious answer. The key to understanding the puzzle is to note that the under-pricing only occurs when two conditions are present. First, the market clearing price is much higher than some well-established normal (reference) price. Second, there is an ongoing pecuniary relationship between the buyer and the seller. Pure scarcity is not enough. Rare art works, beachfront property, and 25-carat diamonds all sell at (very high) market clearing prices.

Once the notion of transaction (dis)utility is introduced, then the role of the normal or reference price becomes transparent. The goods and services listed earlier all have such norms: prices of other dolls similar to Cabbage Patch dolls, regular season ticket prices, prices of other concerts, dinner prices at other times or on other days, etc. These well-established reference prices create significant transaction disutility if a much higher price is charged.

The ongoing relationship between the buyer and the seller is necessary (unless the seller is altruistic), else the seller would not care if transaction disutility were generated. Again that ongoing relationship is present in all the cases described. Coleco couldn't charge more for the dolls because it had plans for future sales to doll customers and even nondoll buyers who would simply be offended by an unusually high price. Musical performers want to sell record albums. Restaurants want to sell dinners at other times and days. When a well-established reference price exists, a seller has to weigh the short-run gain associated with a higher price against the long-run loss of good will and thus sales.

The pricing of sporting events provides a simple test of this analysis. For major sporting events, the

Table 2 Recent Prices for Major Sporting Events

1983 World Series	$25–$30
1984 Super Bowl	All seats $60
1984 Indianapolis 500	Top price $75
1981 Holmes–Cooney fight	Top price $600

price of tickets should be closer to the market clearing price, the larger is the share of total revenues the seller captures from the event in question. At one extreme are league championships such as the World Series and the Super Bowl. Ticket sales for these events are a tiny share of total league revenue. An intermediate case is the Indianapolis 500. This is an annual event, and is the sponsor's major revenue source, but racegoers frequently come year after year so some ongoing relationship exists. At the other extreme is a major championship fight. A boxing championship is a one-time affair involving a promoter and two fighters. Those three parties are unlikely to be a partnership again. (Even a rematch is usually held in a different city.) There is no significant long-run relationship between the sellers and boxing fans.

While it is impossible to say what the actual market clearing prices would be, the figures in Table 2 indicate that the predictions are pretty well confirmed. Good seats for the Super Bowl are probably the single item in greatest demand and are obviously underpriced since even the worst seats sell out at $60.

Of course, some Super Bowl tickets and Cabbage Patch dolls do change hands at high prices through scalpers. Since the black market price does rise to the market clearing level, why do the sellers permit the scalpers to appropriate these revenues? There are two reasons. First, the transaction disutility generated by a high black market price is not attributed to the original seller. The NFL sets a "fair" price; it is the scalper who is obtaining the immoral rents.[11] Second, in many cases the seller is really getting more than the face value of the tickets. Tickets to the Super Bowl are distributed to team owners in large numbers. Many of these tickets are resold to tour operators (see the next section) at prices which are not made public. Similarly, tickets to the NCAA basketball tournament finals are distributed in part to the qualifying teams. These tickets are sold or given to loyal alumni. The implicit price for such tickets is probably in the thousands of dollars.

Methods of Raising Price. A seller who has a monopoly over some popular product may find that the price being charged is substantially less than the market clearing price. How can price be raised without generating excessive negative transaction utility (and thus loss of good will)? The theory provides three kinds of strategies that can be tried. First, steps can be taken to increase the perceived reference price. This can be done in several ways. One way is to explicitly suggest a high reference price (see next section). Another way is to increase the perceived costs of the product, perhaps by providing excessive luxury. As the hockey question showed, perceptions of fairness are affected by costs. In the beer on the beach example, the owner of the run-down grocery store could install a fancy bar. Notice that the extra luxury need not increase the value of the product to the buyer; as long as p^* is increased then demand will increase holding acquisition utility constant. An illustration of this principle is that short best-selling books tend to have fewer words per page (i.e., larger type and wider margins) than longer books. This helps to raise p^*.

A second general strategy is to increase the minimum purchase required and/or to tie the sale of the product to something else. Because of the shape of the value function in the domain of losses, a given price movement seems smaller the larger is the quantity with which it is being integrated. The Super Bowl provides two illustrations of this phenomenon. Tickets are usually sold by tour operators who sell a package including air fare, hotel and game ticket. Thus the premium price for the ticket is attached to a considerably larger purchase. Also, hotels in the city of the Super Bowl (and in college towns on graduation weekend) usually impose a three-night minimum. Since the peak demand is for only one or two nights this allows the hotel to spread the premium room rate over a larger purchase.

The third strategy is to try to obscure p^* and thus make the transaction disutility less salient. One simple way to do this is to sell the product in an unusual size or format, one for which no well-established p^* exists. Both of the last two strategies are used by candy counters in movie theaters. Candy is typically sold only in large containers rarely seen in other circumstances.

Suggested Retail Price.[12] Many manufacturers offer a "suggested retail price" (SRP) for their products. In the absence of fair trade laws, SRP's must be only suggestions, but there are distinct differences across products in the relationship between market prices and SRPs. In some cases the SRP is usually equal to the market price. In other cases the SRP exceeds the market price by as much as 100% or more. What is the role of an SRP that is twice the typical retail price? One possibility is that the SRP

[11] Transferring the transaction disutility is often a good strategy. One way this can be done is to turn over an item for sale to an agent who will sell it at auction. The seller then bears less responsibility for the price.

[12] This paragraph was motivated by a discussion with Dan Horsky several years ago.

is being offered by the seller as a "suggested reference price." Then a lower selling price will provide positive transaction utility. In addition, inexperienced buyers may use the SRP as an index of quality. We would expect to observe a large differential between price and the SRP when both factors are present. The SRP will be more successful as a reference price the less often the good is purchased. The SRP is most likely to serve as a proxy for quality when the consumer has trouble determining quality in other ways (such as by inspection). Thus, deep discounting relative to SRP should usually be observed for infrequently purchased goods whose quality is hard to judge. Some examples include phonograph cartridges which usually sell at discounts of at least 50%, home furniture which is almost always "on sale," and silver flatware where "deep discounting—selling merchandise to consumers at 40% to 85% below the manufacturer's 'suggested retail price' has become widespread in the industry."[13]

4.3. Budgeting Implications: A Theory of Gift Giving

The analysis of budgeting rules suggests that category and time specific shadow prices can vary. This implies that individuals fail to undertake some internal arbitrage operations that in principle could increase utility. In contrast, the standard theory implies that all goods that are consumed in positive quantities have the same marginal utility per dollar, and in the absence of capital market constraints, variations over time are limited by real interest rates. Observed patterns of gift giving lend support to the current theory. Suppose an individual G wants to give some recipient R a gift. Assume that G would like to choose that gift which would yield the highest level of utility to R for a given expenditure. (Other nonaltruistic motives are possible, but it seems reasonable to start with this case.) Then the standard theory implies that G should choose something that is already being consumed in positive quantities by R.

How does this compare with common practice? Casual observation and some informal survey evidence suggest that many people try to do just the opposite, namely buy something R would not buy for himself. Flowers and boxed candy are items that are primarily purchased as gifts. "Gift shops" are filled with items that are purchased almost exclusively as gifts. Did anyone buy a pet rock for himself?

Once the restriction that all shadow prices be equal is relaxed, the apparent anomaly is easily understood. Categories that are viewed as luxuries will tend to have high k's. An individual would like to have a small portion of the forbidden fruit, but self-control problems prevent that. The gift of a small portion solves the problem neatly.

A simple test of the model can be conducted by the reader via the following thought experiment. Suppose you have collected \$100 for a group gift to a departing employee. It is decided to give the employee some wine since that is something the employee enjoys. Suppose the employee typically spends \$5 per bottle on wine. How expensive should the gift wine be? The standard theory says you should buy the same type of wine currently being purchased. The current theory says you should buy fewer bottles of more expensive wine, the kind of wine the employee wouldn't usually treat himself to.

One implication of this analysis is that goods which are priced at the high end of the market should be marketed in part as potential gifts. This suggests aiming the advertising at the giver rather than the receiver. "Promise her anything but give her Arpege."

The gift-giving anomaly refers to those goods in categories with high k's. Individuals may also have categories with low k's. Suppose I like to drink expensive imported beer but feel it is too costly to buy on a regular basis. I might then adopt the rule of drinking the expensive beer only on specific occasions, such as at restaurants or while on vacation.[14] Advertisers may wish to suggest other occasions that should qualify as legitimate excuses for indulgence. One example is Michelob's theme: "Weekends are made for Michelob." However, their follow-up campaign may have taken a good idea too far: "Put a little weekend in your week." Lowenbrau's ads stress a different category, namely, what beer to serve to company. "Here's too good friends, tonight is something special . . . ." While impressing your friends is also involved here, again the theme is to designate specific occasions when the beer k should be relaxed enough to purchase a high cost beer.

Another result of this analysis is that people may sometimes prefer to receive a gift in kind over a gift in cash, again violating a simple principle of microeconomic theory. This can happen if the gift is on a "forbidden list." One implication is that employers might want to use gifts as part of their incentive packages. Some organizations (e.g., Tupperware) rely on this type of compensation very heavily. Dealers are paid both in cash and with a multitude of gift-type items: trips, furniture, appliances, kitchen utensils, etc. Since most Tupperware dealers are women who are second-income earners, the gifts may be a way for a dealer to:

[13] See *Business Week*, March 29, 1982. This example was suggested by Leigh McAlister.

[14] One bit of evidence that people on vacation adopt temporarily low k's is that all resorts seem to have an abundance of gift and candy shops. Some of their business, of course, is for gifts to bring home, but while on vacation, people also seem to buy for themselves at these shops.

(1) mentally segregate her earnings from total family income;

(2) direct the extra income toward luxuries; and

(3) increase her control over the spending of the extra income.[15]

Another similar example comes from the National Football League. For years the league had trouble getting players to come to the year-end All-Star game. Many players would beg off, reporting injuries. A few years ago the game was switched to Hawaii and a free trip for the player's wife or girlfriend was included. Since then, no-shows have been rare.

Conclusion. This paper has developed new concepts in three distinct areas: coding gains and losses, evaluating purchases (transaction utility), and budgetary rules. In this section I will review the evidence presented for each, describe some research in progress, and suggest where additional evidence might be found.

The evidence on the coding of gains and losses comes from two kinds of sources. The "who is happier" questions presented here are a rather direct test, though of a somewhat soft variety. More research along these lines is under way using slightly different questions such as "two events are going to happen to you, would you rather they occurred on the same day or two weeks apart?" The two paradigms do not always lead to the same results, particularly in the domain of losses (Johnson and Thaler 1985). The reasons for the differences are interesting and subtle, and need further investigation. The other source for data on these issues comes from the investigation of choices under uncertainty. Kahneman and Tversky originally formulated their value function based on such choices. In Johnson and Thaler (1985) we investigate how choices under uncertainty are influenced by very recent previous gains or losses. We find that previous gains and losses do influence subsequent choices in ways that complicate any interpretation of the loss function. Some of our data comes from experiments with real money and so are in some sense "harder" than the who is happier data. Kahneman and Tversky are also investigating the multi-attribute extension of prospect theory, and their results suggest caution in extending the single attribute results.

The evidence presented on transaction utility was the beer on the beach and hockey ticket questionnaires, and the data on sports pricing. The role of fairness is obviously quite important in determining reference prices. A large-scale telephone survey undertaken by Daniel Kahneman, Jack Knetch and myself is under way and we hope it will provide additional evidence on two important issues in this area. First, what are the determinants of people's perceptions of fairness? Second, how are market prices influenced by these perceptions? Evidence on the former comes directly from the survery research, while evidence on the latter must come from aggregate economic data. The latter evidence is much more difficult to obtain.

Both the theory and the evidence on the budgetary processes are less well developed than the other topics presented here. The evidence comes from a small sample of households that will not support statistical tests. A more systematic study of household decision making, perhaps utilizing UPC scanner data, should be a high priority.

More generally, the theory presented here represents a hybrid of economics and psychology that has heretofore seen little attention. I feel that marketing is the most logical field for this combination to be developed. Aside from those topics just mentioned there are other extensions that seem promising. On the theory side, adding uncertainty and multiple attributes are obviously worth pursuing. Regarding empirical tests, I would personally like to see some field experiments which attempt to implement the ideas suggested here in an actual marketing environment.

[15] Tax evasion may be another incentive if recipients (illegally) fail to declare these gifts as income.

Acknowledgments

An earlier version of a portion of this paper was presented at the Association for Consumer Research conference in San Francisco, October 1982, and is published in the proceedings of that meeting. The author wishes to thank Hersh Shefrin, Daniel Kahneman and Amos Tversky for many useful discussions. Financial support from the Alfred P. Sloan Foundation is gratefully acknowledged.

References

Arkes, H. R., C. Blumer. 1985. The psychology of sunk cost. *Organ. Behav. Human Decision Processes* **35**(1) 124–140.

Becker, G. S. 1965. A theory of the allocation of time. *Econom. J.* **75**(September) 493–517.

Gregory, R. 1982. Valuing non-market goods: An analysis of alternative approaches. Unpublished dissertation, University of British Columbia, Vancouver, B.C., Canada.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(March) 263–291.

Knetch, J. L., J. A. Sinden. 1984. Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *Quart. J. Econom.* **99** 507–521.

Lancaster, K. J. 1971. *Consumer Demand, A New Approach*. Columbia University Press, New York.

Thaler, R. 1980. Toward a positive theory of consumer choice. *J. Econom. Behav. Organ.* **1**(March) 39–60.

Thaler, R., E. J. Johnson. 1990. Gambling with the house money and trying to break even: The effects of prior outcomes or risky choice. *Management Sci.* **36**(6) 643–660. [Originally cited as Johnson, E., R. Thaler. 1985. Hedonic framing and the break-even effect. Working paper, Cornell University, Ithaca, NY.]

Thaler, R., H. M. Shefrin. 1981. An economic theory of self-control. *J. Political Econom.* **39**(April) 392–406.

Tversky, A., D. Kahneman. 1981. The framing of decisions and the rationality of choice. *Science* **211** 453–458.