Will Products Look More Attractive When Presented Separately or Together?

CHRISTOPHER K. HSEE
FRANCE LECLERC*

This research examines whether each of two different options of comparable overall quality will be perceived more positively when presented in isolation and evaluated separately (*separate evaluation*) or when juxtaposed and evaluated side by side (*joint evaluation*). Six studies, involving either judgment or choice as the dependent variable, reveal a general principle: If the focal options are already attractive (relative to their natural reference) in separate evaluation, then subjecting these options to joint evaluation will hurt their attractiveness. If the focal options are unattractive (relative to their reference) in separate evaluation, subjecting them to joint evaluation will enhance their attractiveness.

Imagine the following scenario: Three local appliance stores sell coffeemakers. The first store offers two brands of comparable quality, Braun and Krups. The second store offers only one brand, Braun. The third store offers only one brand, Krups. Will consumers who visit the first store and evaluate the two brands of coffeemakers jointly value these two brands differently than consumers who go to the second or third store and evaluate either the Braun coffeemaker or the Krups coffeemaker independently? If so, in which store will these brands be valued more?

More generally, suppose that two options of comparable overall quality are presented and evaluated in two modes, separate and joint. In the *separate evaluation* mode, each one of these options is presented by itself and evaluated independently. In the *joint evaluation* mode, these options are juxtaposed and evaluated side by side. Will these options be perceived differently between the two evaluation modes? If so, in which condition will the options be perceived more positively? The present research seeks to answer these questions.

These questions address the fundamental issue of how preferences are constructed and how they are influenced by the nature of a given task. Recent research suggests that preferences are often constructed ad hoc and can be influenced by such subtle manipulations as the framing of the problem (see, e.g., Kahneman and Tversky 1979; Tversky and Kahneman 1981), the display of information (see, e.g., Johnson, Payne, and Bettman 1988), the addition of a dominated alternative to the choice set (see, e.g., Huber, Payne, and Puto 1982; Simonson 1989), and preexisting and developing preferences (Russo, Medvec, and Meloy 1996), to name just a few.

Despite the numerous descriptive theories of preferences, little is known about how judgments differ between joint and separate evaluation. (We shall examine the few lines of research on this topic later.) The distinction between evaluating products jointly or separately is basic in the sense that it applies to all judgments and decisions. Evaluations are done either in the joint evaluation mode where there are explicit alternatives or in the separate (independent) evaluation mode where there are no explicit alternatives. For example, if a consumer shops for a car at a dealership, she is usually in a joint evaluation mode since many alternative models are available for her to see and evaluate concurrently. Alternatively, if she owns a car and thinks about how much she likes it, she is probably in a separate evaluation mode. Likewise, if she does not own a car and considers buying a used car from a neighbor, she is likely to evaluate this car in the separate evaluation mode. Examining whether and how judgments and decisions differ between these two basic modes of evaluation will add to our knowledge of how evaluations are formed and what factors influence this process.

In our research, we argue that evaluation mode influences consumer decision making by either enhancing or muting the role of a natural reference for the product category. We propose that if the focal options are attractive relative to the natural reference used in separate evaluation, then the two options will be judged more posi-

*Christopher K. Hsee is associate professor of behavioral science and France Leclerc is associate professor of marketing, both at the Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637. The authors are listed in alphabetical order; they have made equal contributions. This research was funded in part by a grant from the Graduate School of Business to the authors. The authors thank Joseph Nunes, J. Edward Russo, Eldar Shafir, Itamar Simonson, Richard H. Thaler, the associate editor and three reviewers for helpful comments.

© 1998 by JOURNAL OF CONSUMER RESEARCH, Inc. • Vol. 25 • September 1998
All rights reserved. 0093-5301/99/2502-0006$03.00

tively and have a higher likelihood of purchase in separate evaluation than in joint evaluation. Conversely, if the focal options are unattractive relative to the reference, they will be perceived more favorably and have a higher likelihood of purchase in joint evaluation than in separate evaluation.

In what follows, we review the relevant research and propose a conceptual model along with a set of hypotheses. Experiments follow that test these hypotheses. Finally, we discuss theoretical and practical implications of our findings.[1]

REVIEW OF PREVIOUS RESEARCH

Several lines of research have examined the effect of joint versus separate evaluation on consumer behavior. One is on the effect of conflict in consumer choice (Shafir, Simonson, and Tversky 1993; Tversky and Shafir 1992). According to these authors, the addition (or removal) of an option to the offered set can influence people's choice by making the decision harder (easier) to justify. In one study, respondents were asked to imagine that they were shopping for a CD player. There were two conditions. In one, respondents were told about only one option—a popular Sony model that was on sale for $99. This resembles a separate evaluation condition. In the other condition, respondents were given a choice of two CD players—the Sony described above and a top-of-the-line Aiwa model on sale for $169. This resembles a joint evaluation condition. More participants chose to buy a CD player in the Sony-alone condition than in the Sony + Aiwa condition.

Tversky and Shafir (1992) explained their finding in terms of the degree of conflict generated by the choice set. Faced with an attractive option by itself, people had little conflict; they had a compelling reason to accept that option. The availability of competing alternatives of comparable attractiveness, however, created a conflict, because it did not present an immediate reason for choosing either alternative over the other. The conflict explanation implies that an option is more likely to be chosen when presented alone (separate evaluation) than when paired with a competing option of comparable value (joint evaluation). However, if one of the options is dominated by the other, for example, if the Aiwa is a model of inferior quality being sold at a more expensive price than the Sony, then the joint evaluation does not create a conflict but provides reasons for the participants to choose the dominant option. As a result, in the dominant option scenario, more participants chose the Sony in the Sony + Aiwa than in the Sony-alone condition (Tversky and Shafir 1992). What this research implies is that pairing an option with another option of comparable quality will create conflict in consumers and therefore will decrease the likelihood of purchase for either option.

In our research, we extend that work by showing that even when neither option dominates the other, pairing one option with the other can, under certain circumstances, increase the attractiveness of that option and thereby increase consumers' intention to purchase that option. We focus on the fact that in separate evaluation, people compare the option they evaluate to some reference information. As such, the relative attractiveness of the option compared to the reference will have an impact in a separate evaluation situation but less so in a joint evaluation situation since in this latter mode the focus shifts from the reference information to the other option in the set. Thus, we extend the findings of Tversky and Shafir (1992) by showing that having to choose between two comparable options can either decrease or increase the likelihood of choice of these options depending on how attractive they are relative to the reference information used in separate evaluation.

Another line of research related to our current work is on preference reversals between joint and separate evaluation (Bazerman, Loewenstein, and White 1992; Hsee 1996; Hsee et al. 1996). For example, in joint evaluation a used dictionary with a cosmetic defect was valued more highly than a used dictionary with fewer entries but no cosmetic defect; in separate evaluation, the reverse was true (Hsee 1996). These preference reversals have been explained in terms of the ''evaluability hypothesis'': attributes that are hard to evaluate independently (e.g., number of entries in a dictionary) loom larger in joint evaluation, and attributes that are easy to evaluate independently (e.g., cosmetic defect) loom larger in separate evaluation. Nowlis and Simonson (1997) documented a related but different type of preference reversal for consumer products, between direct choice and continuous ratings. Direct choice is always made in joint evaluation, and continuous ratings can be made in either separate or joint evaluation. Those authors showed that ''comparable attributes,'' which produce precise and easy-to-compute comparisons, loom larger in choice, and ''enriched attributes,'' which are more meaningful to the consumer, receive greater weight in ratings.

Our current work addresses a different issue regarding joint and separate evaluation than the joint-separate preference-reversal research. The preference-reversal work investigates whether the rank order of the two options changes between the two evaluation modes. The current research examines whether a given option will look more attractive when presented alone (i.e., in separate evaluation) or when paired with another option of comparable quality (i.e., in joint evaluation).[2] In other words, this

[1]Joint evaluation does not mean evaluation for joint purchases. It means evaluating options concurrently.

[2]It is important to stress that our goal is to study the evaluation of single alternatives depending on the context (whether they are considered alone or in the presence of another). This is very different from the question of whether two events, x and y, provide more satisfaction when aggregated (i.e., $u(x + y)$) or when segregated (i.e., $u(x)$

research is concerned with the relative value of all options when they are evaluated singly or together. In addition, the mechanism underlying the evaluation mode effect as studied in our current research is very different from the mechanism underlying the joint-separate preference-reversal phenomenon.

CONCEPTUAL FRAMEWORK

Before we introduce our specific predictions, let us examine two underlying assumptions. The first is that when evaluating only one option, people compare it to whatever reference information is available at the time. Previous research (see, e.g., Kahneman and Tversky 1979; Shafir, Osherson, and Smith 1993; Tversky and Kahneman 1991; see also Brickman and Campbell 1971; Helson 1964) shows that the attractiveness of a stimulus option depends not only on the absolute value of that option but also on its relative value, that is, relative to some reference value. For example, when a prospective home buyer assesses the worth of a house, she is likely to anchor on some reference and then make an assessment by comparing the current house against that reference. The reference can be either a specific and concrete case, for example, a house owned by her friend, or based on some abstract, distributional information, such as the ''average house'' or ''most other houses'' in the given region.

The second assumption underlying our current research, which is shared by the preference-reversal work mentioned previously, is that when evaluating two options jointly, people rely less on the reference information they would otherwise use in separate evaluation and make their evaluation by comparing one option against the other. For example, if the home buyer mentioned above is shown two new houses at the same time, she will be less likely to think of her friend's house or the average house; instead, she will compare across the two options presented. This assumption is based on prior research demonstrating that in judgment, people rely more on vivid and available information than on pallid background information, even if the latter information is more useful than the former (see, e.g., Tversky and Kahneman 1973, 1974). When a person judges an option in the joint evaluation mode, the alternative option is usually more salient and more available than the original reference information and therefore may replace the original reference and serve as the primary reference itself. In this case, one's attention is drawn to the differences between the focal options (Tversky 1969).

On the basis of these assumptions, we can briefly describe our hypotheses as follows. First, whether a set of stimulus options (let's say there are two stimulus options) will look more attractive in separate evaluation or in joint evaluation depends on the relationship of these options with the reference that people use in separate evaluation. More specifically, if these focal options are already attractive relative to their reference in separate evaluation, then subjecting these options to a joint evaluation will hurt their attractiveness. This is because joint evaluation enables people to compare across the options and through this comparison people will see weaknesses of each option relative to the other option. In contrast, if the focal options are unattractive relative to their reference in separate evaluation, then juxtaposing them in a joint evaluation will enhance their attractiveness. This is because joint evaluation enables people to discover advantages of each option relative to the other option.

The following is a more formal analysis of these propositions. Suppose that two stimulus options, options A and B, involve a trade-off along two attributes, attribute 1 and attribute 2:

	Attribute 1	*Attribute 2*
Option A	a1	a2
Option B	b1	b2

where a1 $>$ b1 and a2 $<$ b2. (We use the symbol ''$>$'' to mean ''better than'' and the symbol ''$<$'' to mean ''worse than.'')

Also assume that in separate evaluation both options are associated with some common reference, R, and that its values on the two attributes are r1 and r2, respectively. On the basis of the assumptions proposed above, we assume that in separate evaluation people compare the given option with R, and in joint evaluation they compare one option against the other. For the sake of simplicity and without loss of generality, we assume that in joint evaluation people totally ignore the original reference, R, and use exclusively the alternative option as their basis of comparison.[3]

Thus, the attractiveness of the stimulus options in the two evaluation models can be modeled as

$$V_{\mathrm{SEP}}(\mathrm{A}) = k + w \times u(\mathrm{a1} - \mathrm{r1}) + (1 - w) \times v(\mathrm{a2} - \mathrm{r2}); \quad (1)$$

$$V_{\mathrm{SEP}}(\mathrm{B}) = k + w \times u(\mathrm{b1} - \mathrm{r1}) + (1 - w) \times v(\mathrm{b2} - \mathrm{r2}); \quad (2)$$

$$V_{\mathrm{JNT}}(\mathrm{A}) = k + w \times u(\mathrm{a1} - \mathrm{b1}) + (1 - w) \times v(\mathrm{a2} - \mathrm{b2}); \quad (3)$$

and

$$V_{\mathrm{JNT}}(\mathrm{B}) = k + w \times u(\mathrm{b1} - \mathrm{a1}) + (1 - w) \times v(\mathrm{b2} - \mathrm{a2}), \quad (4)$$

where V_{SEP} and V_{JNT} stand for the attractiveness of the

$+ u(\mathrm{y}))$. On this, see Linville and Fisher (1991), Thaler (1985), and Thaler and Johnson (1990).

[3]In reality, people may use a combination of R and the alternative option as their basis of comparison in joint evaluation. Even so, the basic arguments will hold.

given option in the separate and the joint evaluation modes, respectively, k is the baseline attractiveness of a product in the given product category, which can be regarded as a constant, and w is the relative weight between the two attributes, where $0 < w < 1$.[4] Finally, u and v are utility functions for attribute 1 and attribute 2, respectively. Both u and v are monotonically increasing functions, with $u(0) = 0$ and $v(0) = 0$.

Equations 1–4 imply that the evaluation mode effect, whether the stimulus options are perceived more positively in separate evaluation or in joint evaluation, depends on the position of the stimulus options relative to the reference, which will be called the options-reference relationship. This leads to our first hypothesis:

H1: There will be an interaction between the options-reference relation and the evaluation mode. The options-reference relation will have a greater effect on the attractiveness of the stimulus options in separate evaluation than in joint evaluation.

There are many possible relationships between the stimulus options and the reference. Let us consider two extreme cases: (1) where A and B are each better than R on both attributes, which will be referred to as the ''better-than-reference'' condition, and (2) where A and B are each worse than R on both attributes, which will be referred to as the ''worse-than-reference'' condition. (We shall address other conditions in the General Discussion section.)

Let us first examine the better-than-reference case, where $\min(a1, b1) > r1$ and $\min(a2, b2) > r2$. Because $a1 > b1 > r1$, $b2 > a2 > r2$, and u and v are monotonically increasing functions, we have,

$$u(a1 - r1) > u(a1 - b1), \quad (5)$$

$$v(a2 - r2) > v(a2 - b2), \quad (6)$$

$$u(b1 - r1) > u(b1 - a1), \quad (7)$$

and

$$v(b2 - r2) > v(b2 - a2). \quad (8)$$

Applying Equations 5–8 to Equations 1–4, we obtain:

$$V_{SEP}(A) > V_{JNT}(A) \quad (9)$$

and

$$V_{SEP}(B) > V_{JNT}(B). \quad (10)$$

Equations 9 and 10 can be summarized in the following hypothesis, which is a special case of Hypothesis 1:

H1a: If both A and B are better than R, then they will look more attractive in separate evaluation than in joint evaluation.

Let us now turn to the worse-than-reference case, where $\max(a1, b1) < r1$ and $\max(a2, b2) < r2$. Given that $r1 > a1 > b1$, $r2 > b2 > a2$, and that u and v are monotonically increasing functions, we have,

$$u(a1 - r1) < u(a1 - b1), \quad (11)$$

$$v(a2 - r2) < v(a2 - b2), \quad (12)$$

$$u(b1 - r1) < u(b1 - a1), \quad (13)$$

and

$$v(b2 - r2) < v(b2 - a2). \quad (14)$$

Combining Equations 11–14 with Equations 1–4 leads to the following conclusions:

$$V_{SEP}(A) < V_{JNT}(A) \quad (15)$$

and

$$V_{SEP}(B) < V_{JNT}(B). \quad (16)$$

This is our Hypothesis 1b, which is another special case of Hypothesis 1.

H1b: If both A and B are worse than R, then they will look more attractive in joint evaluation than in separate evaluation.

Taken together, Hypotheses 1a and 1b suggest that the evaluation mode effect, whether the stimulus options will look more attractive in separate evaluation or in joint evaluation, is contingent on the relationship between the stimulus options and their reference in separate evaluation. If these options are attractive relative to their reference in separate evaluation, then joint evaluation will hurt them. If these options are unattractive relative to their reference in separate evaluation, then joint evaluation will help them.

EMPIRICAL EVIDENCE

Below we report on six empirical studies that tested these hypotheses. These studies tapped into different product categories, used different methods to manipulate

[4]We assume a constant relative attribute weight, w, for both the separate evaluation mode and the joint evaluation mode. In theory, the attribute weight can vary between the two modes. It can be easily proved that if the attribute weight varies, neither Equations 9 and 10 nor Equation 15 and 16 will hold. However, a relaxed version of these equations will still hold, namely, $V_{SEP}(A) + V_{SEP}(B) > V_{JNT}(A) + V_{JNT}(B)$ for the better-than-reference condition, and $V_{SEP}(A) + V_{SEP}(B) < V_{JNT}(A) + V_{JNT}(B)$ for the worse-than-reference condition. They imply that in the better-than-reference condition, A and B will on average look more attractive in separate evaluation than in joint evaluation, and in the worse-than-reference condition, A and B will on average look more attractive in joint evaluation than in separate evaluation. In all of the studies reported in this article, the results were consistent not only with these relaxed predictions, but also with the more stringent hypotheses of Equations 9 and 10 and Equations 15 and 16, suggesting that the relative attribute weight did not change significantly between the joint and the separate evaluation modes.

TABLE 1

SUMMARY OF THE STUDIES

Study	Product	Reference information	Dependent variable
1	Cordless phone	Externally given	Willingness to pay
2	Light bulbs	Externally given	Willingness to pay
3	CD changer	Naturally evoked	Willingness to pay
4	Printer	Naturally evoked	Willingness to pay
5	Cordless phone	Externally given	Choice
6	CD changer	Naturally evoked	Choice

the reference information, and involved different dependent variables, as summarized in Table 1.

Studies 1 and 2 used almost identical procedures; so did studies 3 and 4 and studies 5 and 6. To save space, we shall discuss only the first study in each pair in detail.

Studies 1 and 2

Method of Study 1

This study involved the evaluation of two cordless phones:

	Maximum operative distance	*Battery life per recharge*
Model A	150 feet	2 days
Model B	60 feet	10 days

The questionnaire for this study had six versions. They represented a 2 (options-reference relation) × 3 (evaluation mode) design. The two options-reference-relation conditions were better than reference and worse than reference. The three evaluation mode conditions were joint evaluation, separate evaluation A, and separate evaluation B. In all of the conditions, participants were asked to assume that they were shopping for a cordless phone and planned to spend between $50 and $150. (We included the price range in order to prevent participants from giving unreasonably high or low willingness-to-pay prices.) They were further told that when buying a cordless phone, they should check (*a*) its maximum operative distance, that is, how far from the base unit the handset can operate, and (*b*) its battery life per recharge, that is, how many days its battery can last after each recharge.

The reference information in this study was external to the focal options and explicitly given in the instructions. It was the cordless phone of a friend. In the better-than-reference condition, participants read, ''A friend of yours recently bought a cordless phone. His unit has a max operative distance of 30 feet, and a battery life of 1 day per recharge.'' Here, both model A and model B were better than the friend's phone (the reference).

In the worse-than-reference condition, participants read, ''A friend of yours recently bought a cordless phone. His unit has a max operative distance of 300 feet, and a battery life of 20 days per recharge.'' Here, both model A and model B were inferior to the reference.

TABLE 2

MEAN WILLINGNESS-TO-PAY VALUES ($) FOR EACH MODEL IN STUDY 1

Options-reference relation and evaluation mode	Model A	Model B
Better than reference:		
Joint	79.00	73.05
Separate	91.75	89.95
Worse than reference:		
Joint	80.93	66.33
Separate	65.37	60.04

NOTE.—N = 142.

In the joint evaluation condition, participants were told that there were two models of cordless phones available in the store and shown the information about these models as presented above. They were then asked, ''What is the most you are willing to pay for a unit of each model?''

In each of the separate evaluation conditions, participants were given the same reference information as in the joint evaluation condition, were told that there was only one model available, and were presented with the information about either model. They were then asked, ''What is the most you are willing to pay for a unit of this model?''

Respondents were 142 unpaid M.B.A. students from managerial decision-making and organizational behavior classes at a midwestern university. They randomly received one of the six versions of the questionnaire and completed it individually in class.

Results of Study 1

The results are summarized in Table 2. We analyzed the data for the two stimulus options separately. Recall that our main hypothesis predicted an interaction between options-reference relation and evaluation mode. To test this hypothesis, two 2 (options-reference relation; better than reference vs. worse than reference) × 2 (evaluation mode; joint vs. separate) ANOVAs were performed, one for model A and one for model B. Confirming Hypothesis 1, there were significant interactions between options-reference relation and evaluation mode for both model A ($F(1, 99) = 6.92$, $p < .05$) and model B ($F(1, 88) = 4.05$, $p < .05$).

To test for Hypotheses 1a and 1b, we conducted four planned comparisons between the two evaluation modes, one for each model in the better-than-reference condition, and one for each model in the worse-than-reference condition. All of the results were in the predicted direction. In the better-than-reference condition, both model A and model B were valued more highly in separate evaluation

TABLE 3

MEAN WILLINGNESS-TO-PAY VALUES ($)
FOR EACH BRAND IN STUDY 2

Options-reference relation and evaluation mode	Brand A	Brand B
Better than reference:		
Joint	1.61	1.99
Separate	2.31	3.31
Worse than reference:		
Joint	1.42	1.77
Separate	.94	1.65

NOTE.—N = 158.

than in joint evaluation ($t = 1.72$, $p < .05$[5] for model A and $t = 1.90$, $p < .05$ for model B). Conversely, in the worse-than-reference condition, both model A and model B were valued more favorably in joint evaluation than in separate evaluation, although the difference for model B was not statistically significant ($t = 2.00$, $p < .05$, and $t < 1$, n.s., for models A and B, respectively).

Method and Results of Study 2

Study 2 was a replication of Study 1 using a different product. The stimulus options were two brands of light bulbs.

	Life expectancy	*Light output*
Brand A	1,000 hours	900 lumens
Brand B	1,500 hours	800 lumens

In the better-than-reference condition, participants were told, ''the life expectancy of most 75-watt light bulbs is 500 hours, and the light output is 700 lumens.'' In the worse-than-reference condition, they were told, ''the life expectancy of most 75-watt light bulbs is 2,500 hours, and the light output is 1,000 lumens.''

As in study 1, there was a significant interaction between options-reference relation and evaluation mode for both brand A ($F(1, 112) = 4.77$, $p < .05$) and brand B ($F(1, 101) = 3.93$, $p = .05$; see Table 3). In the better-than-reference condition, both brands were valued more in separate evaluation than in joint evaluation ($t = 1.72$, $p < .05$ for brand A and $t = 2.56$, $p < .05$ for brand B). In the worse-than-reference condition, both brands were valued more in joint evaluation than in separate evaluation ($t = -1.98$, $p < .05$ for brand A and $t < 1$, n.s. for brand B).

Discussion of Studies 1 and 2

The results of these studies provided support for our predictions, highlighting the interdependence between evaluation mode and the relationship between the stimulus options and their reference. If the focal options are better than their reference (e.g., a friend's phone), the focal options will receive more favorable valuations when evaluated separately than when evaluated jointly. If the focal options are worse than their reference, these options will enjoy more favorable judgments when evaluated jointly.

These findings suggest that one can manipulate the direction of the evaluation mode effect by setting different reference points, as long as the evaluator does not already have a clear reference in mind. If one sets the reference lower than the focal options, then the focal options will be valued more favorably in separate evaluation than in joint evaluation; if one sets the reference higher than the focal options, then they will look more attractive in joint evaluation than in separate evaluation.

Studies 3 and 4

In studies 3 and 4 we sought to extend the results of studies 1 and 2 by using a different method to manipulate the relationship between the stimulus options and the reference. In studies 1 and 2, the reference was explicitly given, and the options-reference relation was manipulated by varying the R. In studies 3 and 4, no explicit reference was given. Instead, we assumed, on the basis of Kahneman and Miller's (1986) norm theory, that in separate evaluation, participants would evoke some internal norm of the given product category, for example, what they believe to be the average product in that category, and use that as their reference. The options-reference relation was then manipulated by varying the values of the stimulus options rather than by varying the reference.

Method of Study 3

The products to be evaluated in this study were CD changers. As in studies 1 and 2, the questionnaire versions represented a 2 (options-reference relation) × 3 (evaluation mode) design. In all six conditions, participants were asked to imagine that they were shopping for a CD changer and planned to spend between $150 and $300. In the better-than-reference condition, the focal CD changers were:

	CD capacity	*Length of warranty*
Model A	10 CDs	40 months
Model B	15 CDs	20 months

In the worse-than-reference condition, the focal CD changers were:

	CD capacity	*Length of warranty*
Model A	2 CDs	4 months
Model B	3 CDs	2 months

[5]All t-tests are one-tailed.

We expected that most people would have some idea about what the average CD changer was like and would use that information as their reference. We developed our stimuli so that the two focal options would dominate the average in the better-than-reference condition. In the worse-than-reference condition, the focal options were designed so that they would be dominated by the average CD changer.

As in studies 1 and 2, respondents were exposed to and indicated their willingness to pay for either both models (joint evaluation) or one of the models (separate evaluation).

To check that participants indeed had some idea what most CD changers are like and that our options-reference manipulation was effective, we asked participants the following questions after they had completed the questionnaire: (1) ''How many CDs do you think most CD-changers can hold?'' and (2) ''How long do you think the warranty coverage is for most CD-changers?''

Respondents were 157 students recruited in the dining halls of a large Midwestern university. They randomly received one of the six versions of the questionnaire and completed it individually. Upon completion, each participant received a candy bar as compensation.

Results of Study 3

Manipulation Check. On average, respondents had the following impressions of ''most CD changers'': that they would have a capacity of 7.5 CDs and that the length of the warranty would be 13 months. Clearly, ''most CD changers,'' which presumably served as a reference, lay between the focal CD changers in the better-than-reference condition and those in the worse-than-reference condition. In other words, in the better-than-reference condition, the focal CD changers were superior to this reference; in the worse-than-reference condition, the focal CD changers were inferior to this reference. These results assured that our options-reference-relation manipulation was effective.[6]

[6]A closer examination of the data revealed that respondents in the better-than-reference condition had very different impressions of the typical CD changer from those in the worse-than-reference condition. The means are summarized below:

Condition	*CD capacity*	*Length of warranty*
Better than reference	10 CDs	17 months
Worse than reference	5 CDs	9 months
Mean	7.5 CDs	13 months

This difference was probably due to an anchoring-and-adjustment process. Because participants indicated their impressions of the typical CD changer after having evaluated the focal model(s), their impressions of the norm may well have been influenced by the values of the stimuli, which were obviously different between the two options-reference-relation conditions. However, it is important to note that despite the difference in impressions, the norm evoked by respondents in either the better-than-reference or the worse-than-reference condition lay between the focal options in those two conditions. Thus, the options-reference-relation manipulation was effective.

TABLE 4

MEAN WILLINGNESS-TO-PAY VALUES ($) FOR EACH MODEL IN STUDY 3

Options-reference relation and evaluation mode	Model A	Model B
Better than reference:		
Joint	220.42	216.39
Separate	259.09	258.75
Worse than reference:		
Joint	169.60	187.20
Separate	134.74	160.00

NOTE.—N = 157.

Willingness-to-Pay Values. The results are summarized in Table 4. We subjected the data to the same set of analyses as we did in studies 1 and 2. Again, in support of Hypothesis 1, the interactions between the options-reference relation and evaluation mode were significant for both model A ($F(1, 109) = 9.84$, $p < .05$) and model B ($F(1, 102) = 5.75$, $p < .05$).

The four planned subsequent analyses revealed patterns of results similar to those in studies 1 and 2 and lent further support to Hypotheses 1a and 1b. In the better-than-reference condition, both CD changers received more favorable valuations when evaluated separately than when evaluated jointly ($t = 2.52$, $p < .01$ and $t = 1.99$, $p < .05$, for models A and B, respectively). By contrast, in the worse-than-reference condition, both CD changers received more favorable valuations when judged jointly than when judged separately ($t = 2.09$, $p < .05$ and $t = 1.85$, $p < .05$, respectively).

Method and Results of Study 4

Study 4 was a replication of study 3 with a different product. The stimulus options were two dot-matrix computer printers. In the better-than-reference condition, they were:

	Speed	*Number of fonts*
Model A	2 seconds per page	Up to 200 fonts
Model B	4 seconds per page	Up to 300 fonts

In the worse-than-reference condition, they were:

	Speed	*Number of fonts*
Model A	60 seconds per page	Up to 2 fonts
Model B	120 seconds per page	Up to 3 fonts

To ensure that our manipulation was effective, as in study 3, we asked subjects their impression of most printers. As in study 3, subjects reported that the ''average printer'' (speed = 41 seconds/page, number of fonts = 59) was better than the stimulus printers in the worse-than-reference condition and better than the stimulus printers in the worse-than-average condition.

TABLE 5

MEAN WILLINGNESS-TO-PAY VALUES ($)
FOR EACH MODEL IN STUDY 4

Options-reference relation and evaluation mode	Model A	Model B
Better than reference:		
Joint	172.86	163.03
Separate	202.76	251.43
Worse than reference:		
Joint	124.38	145.21
Separate	97.16	77.19

NOTE.—N = 155.

TABLE 6

PERCENTAGE OF RESPONDENTS CHOOSING
EACH MODEL IN STUDY 5

Options-reference relation and evaluation mode	Model A (%)	Model B (%)
Better than reference:		
Joint	43	30
Separate	63	58
Worse than reference:		
Joint	27	17
Separate	15	7

NOTE.—N = 232.

The results replicated those in study 3 (see Table 5). There was a significant interaction between the options-reference relation and evaluation mode for both model A ($F(1, 108) = 4.06$, $p < .05$) and model B ($F(1, 99) = 7.21$, $p < .05$). In the better-than-reference condition, both models were valued more in separate evaluation than in joint evaluation ($t = 1.46$, $p < .10$ for model A and $t = 1.97$, $p < .05$ for model B). In the worse-than-reference condition, both models were valued more in joint evaluation than in separate evaluation ($t = 1.94$, $p < .05$ for model A and $t = 4.45$, $p < .001$ for model B).

Discussion of Studies 3 and 4

The findings of these studies suggest that the effects predicted in Hypotheses 1a and 1b can be obtained not only by varying the reference, but also by varying the stimulus options. Unlike studies 1 and 2, in which the reference was explicitly specified, studies 3 and 4 did not provide any external reference; rather, the reference was one's implicit impression of the typical member of the given product category. The results of these studies suggest that, as long as we know whether the focal options are better or worse than people's impression of the typical member of the given product category, we can accurately predict which evaluation mode leads to more favorable judgments. If the focal options are better than the norm, separate evaluation will result in more favorable judgments; if the focal options are worse than the norm, joint evaluation will yield more favorable judgments.

Studies 5 and 6

So far we have only tested our hypotheses with judgment (willingness to pay) as the dependent variable. Studies 5 and 6 extend the previous studies by using a different dependent variable, whether to buy the given product(s) or not (i.e., purchase choice).

Method of Study 5

Study 5 involved the same stimulus options as study 1, except that a constant price was attached to the two options:

	Maximum operative distance	*Battery life per recharge*	*Current price*
Model A	150 feet	2 days	$100
Model B	60 feet	10 days	$100

We added the price information because, in order to make the purchase choice decision, price information was necessary.

The questionnaire versions for this study represented the same 2 (options-reference relation) × 3 (evaluation mode) design as study 1. The only difference was in the final question. In the joint evaluation condition, participants were asked,

> What would you do now? Circle one option:
> (A) Buy Model A now;
> (B) Buy Model B now;
> (C) Buy neither now.

In the separate evaluation condition, they were asked,

> What would you do now? Circle one option:
> (A) Buy this model now;
> (B) Do not buy it now.

Respondents were 232 unpaid students recruited in the dining halls of two large midwestern universities. They randomly received one of the six versions of the questionnaire and completed it individually. Upon completion, each participant received a candy bar as compensation.

Results of Study 5

The results are summarized in Table 6. Recall that the main proposition of this research (Hypothesis 1) predicts an interaction between options-reference relation and evaluation mode. Translated in terms of purchase choice, this hypothesis implies that whether more people would purchase a given option when it is presented alone or when it is paired with an alternative option depends on the relationship between these options and their reference. To test this hypothesis, we analyzed the choice data separately for model A and for model B using the categorical data modeling procedure. Both sets of analyses revealed significant interactions between options-reference relation

TABLE 7

PERCENTAGE OF RESPONDENTS CHOOSING EACH MODEL IN STUDY 6

Options-reference relation and evaluation mode	Model A (%)	Model B (%)
Better than reference:		
Joint	62	34
Separate	71	44
Worse than reference:		
Joint	22	50
Separate	3	11

NOTE.—N = 203.

and evaluation mode in the predicted direction ($\chi^2 = 4.44$, $p < .05$ for model A, and $\chi^2 = 5.75$, $p < .05$ for model B).

To test for Hypotheses 1a and 1b, we further analyzed the data for the better-than-reference and the worse-than-reference conditions separately. In the better-than-reference condition, the percentage of respondents choosing a given option was higher in separate evaluation than in joint evaluation, and this was true for both options ($\chi^2 = 3.29$, $p < .1$ for model A, and $\chi^2 = 5.47$, $p < .05$ for model B). Conversely, in the worse-than-reference condition, the percentage of respondents choosing a given option was higher in joint evaluation than in separate evaluation, and, again, this tendency was true for both options, although the effects were not statistically significant ($\chi^2 = 1.31$ and 1.39 for models A and B, respectively).

Method and Results of Study 6

Study 6 used the same stimuli as study 3. As in Study 5, a constant price was added to the original stimulus options:

	CD capacity	*Length of warranty*	*Today's price*
Model A	10 CDs	40 months	$225
Model B	15 CDs	20 months	$225

We analyzed the data in the same manner as we did for the study 5 data (see Table 7). In support of Hypothesis 1, the interactions between options-reference relation and evaluation mode were significant for both model A ($\chi^2 = 4.47$, $p < .05$) and model B ($\chi^2 = 9.78$, $p < .005$). Subsequent analyses reveal that in the better-than-reference condition, more respondents chose to purchase a given model in the separate evaluation than in the joint evaluation condition, although the differences were not significant ($\chi^2 = .63$ and .60, respectively). In the worse-than-reference condition, more respondents chose to buy a given model in the joint evaluation than in the separate evaluation condition, and these effects were significant for both model A ($\chi^2 = 4.49$, $p < .05$) and model B ($\chi^2 = 10.53$, $p < .005$).

To increase the power of our analyses, we pooled the data of studies 5 and 6 together and reanalyzed the combined data with product category (cordless phone vs. CD changer) as an additional independent variable. As before, there were significant interactions between options-reference relation and evaluation mode for both option A ($\chi^2 = 8.39$, $p < .01$) and option B ($\chi^2 = 15.09$, $p < .001$). In the better-than-reference condition, more respondents chose to buy each option in the separate evaluation than in the joint evaluation condition ($\chi^2 = 3.19$, $p < .1$ for option A, and $\chi^2 = 4.71$, $p < .05$ for option B). In the worse-than-reference condition, more people chose to buy these options in the joint evaluation than in the separate evaluation condition ($\chi^2 = 5.36$, $p < .05$ for option A, and $\chi^2 = 10.41$, $p < .005$ for option B).[7]

Discussion of Studies 5 and 6

These results show that our hypotheses concerning the interdependence of evaluation mode and option-reference relation apply not only to willingness-to-pay judgments, but also to purchase choices. If the focal products are attractive (relative to their reference), more people will buy each of them if they are presented alone than if they are presented together. However, if the focal products are unattractive (relative to their reference), more people will buy each of them if they are presented together than if they are presented alone.

GENERAL DISCUSSION

A series of studies reported herein tapped into different product categories, used different types of reference information, and involved different dependent variables. A highly consistent pattern of results emerged: When the focal products are better than a reference, which could be one's impression of the norm of the given product category or a specific example, they will look more attractive and will be more likely to be purchased when presented separately than when presented jointly. If the stimulus products are worse than the reference, they will be perceived more favorably and will be more likely to be chosen when presented together than when presented separately.

The choice results in the worse-than-reference condition are particularly noteworthy. These results violate a fundamental normative assumption of consumer behavior called the ''regularity'' assumption, which postulates that the market share of a given product cannot be increased by including an alternative product. In contrast to this assumption, our results show that more people would opt to purchase a product when there is an alternative product

[7]Product category did not have significant interaction effects with any other variables, except for an interaction between options-reference relation and product category for option A ($\chi^2 = 4.82$, $p < .05$). This effect does not seem to have any theoretical significance.

(joint evaluation) than when there are no alternatives (separate evaluation). The well-documented attraction effect produces a similar violation (Huber et al. 1982; Simonson 1989).

It should be mentioned that in this research we have tested our hypotheses only in highly simplified situations. We have assumed that there are only two stimulus options, that they vary on only two attributes, and that they share a common reference in separate evaluation. What will happen in situations involving more options, more attributes, or multiple references? It is beyond the scope of this article to exhaust all of these possibilities. Nevertheless, we believe that the basic assumptions underlying our hypotheses will hold in these more complex scenarios, but the exact prediction will vary from situation to situation. For example, in situations involving more than two options, as long as all of the options are better (worse) than the reference information, then separate evaluation of these options will result in more (less) favorable outcomes than joint evaluation.

It should also be mentioned that in this research we focused only on two extreme scenarios of the relationship between the stimulus options and the reference, namely, the better-than-reference and the worse-than-reference conditions. What will happen in other scenarios, for example, when one option is better than the reference on one attribute but the other option is better than the reference on another attribute, or when both options are better than the reference on one attribute but worse than the reference on another? Again, we believe that the basic assumptions underlying our hypotheses are applicable to those situations as well, even though the final predictions may be impossible to make without further information. Let us consider one such scenario, for example, where A is better than the reference on attribute 1 and B is better than the reference on attribute 2. Based on Equations 1–4, each of the stimulus options will be perceived positively on one attribute and negatively on the other in both joint and separate evaluations. Consequently, without further information about the relative weight of the two attributes and the exact distance between the attribute values and the reference point, no systematic predictions can be made as to whether the options would be more attractive under separate evaluation or joint evaluation mode.

Finally, a number of factors are likely to moderate the effect of evaluation mode on product attractiveness. Given that evaluation mode affects product attractiveness through the substitution of a natural reference by the more salient alternative option, any factors affecting the likelihood of this substitution should moderate the evaluation-mode effect. Factors can affect the likelihood of substitution in two different ways, either by making the natural reference highly accessible or by making the alternative option harder to use as a reference. For instance, it is likely that experience or familiarity with the product category will moderate the differential impact of joint evaluation. For experienced decision makers, the salience of the alternative option in the joint evaluation mode may not be sufficient for it to replace an internal reference point since for them this internal reference is likely to be highly accessible.[8] Another factor that could moderate the evaluation-mode effect is whether the alternatives are difficult to compare, such as when the two products being compared are not from the same product category. In such cases, the alternative option in joint evaluation will not provide an easy-to-use reference. As a result, the consumer might just rely on a more natural reference to evaluate the options even when both options are presented together.

The studies reported in this article corroborate existing research in marketing and behavioral decision making by demonstrating that preferences are not necessarily well defined or stable (see Slovic [1995] for a review). This research also adds to the literature (Puto 1987; Thaler 1985; Tversky and Kahneman 1991) showing that people use reference points as the basis for judging the value of alternatives. More important, it proposes and demonstrates that a natural reference point can be replaced when a more salient one becomes available, as is the case under a joint evaluation mode. As a result, the evaluation of a given alternative can differ as a function of the reference point that is being used.

Specifically, the present research complements prior research on joint-separate evaluation preference reversals (see, e.g., Hsee 1996; Hsee et al. 1996; Nowlis and Simonson 1997) by showing that not only does the rank order of the stimulus options change between joint and separate evaluation, the absolute attractiveness of these options also changes. It is important to note that the evaluation mode effects studied in the current research are orthogonal to, and can coexist with, the preference-reversal effects investigated in the prior research. For instance, it is possible that two options, A and B, are each valued more highly in separate evaluation than in joint evaluation (an evaluation mode effect), but, at the same time, option A is valued more highly than option B in separate evaluation and option B is valued more highly than option A in joint evaluation (a preference-reversal effect). It is also important to note that what drives the evaluation-mode effect observed in the current research is different from what causes the preference-reversal phenomenon. The evaluation-mode effect depends on the relationship between the stimulus options and their reference point, and not on attribute evaluability, attribute-task compatibility, or any other factors responsible for preference reversals.

The current research has important practical implications. As mentioned at the beginning of the article, it can help retailers decide whether to carry multiple models or brands of a certain product category or to limit themselves to one model or brand. Our research would suggest that if the two brands of coffeemakers that are available to a

[8]Thaler (1985) makes a similar argument for the role of the suggested retail price as a potential reference price, that is, that it will be more influential for unfamiliar products or inexperienced consumers.

retailer, Braun and Krups, are better than the average brand, then the retailer would be better off carrying only one of them. On the other hand, if the two brands are of lower quality than the average brand, then each brand will look more attractive if they are offered together than if they are offered independently. So, this suggests (and to some extent, it is consistent with current practices, though we do not claim that this reasoning is the primary motivation) the following strategy for a car dealership. Car dealers who sell luxury cars would be better off carrying only one make, say, Lexus. Conversely, for car dealers who sell lower-end cars, carrying multiple makes would be a better strategy.

The effect of evaluation mode on judgment and choice is obviously not only limited to product display but could impact consumer decision making in a large number of activities. For example, in product concept testing, consumers are often provided with descriptions of product concepts that they have to evaluate. Marketing managers should be aware that responses could be biased as a function of whether consumers evaluate one or more concepts at the time and whether these new concepts are improvements (which presumably they are) over average existing products. If these concepts are improvements over the average existing products, evaluating multiple concepts simultaneously will lead to less positive responses than evaluating them independently.

Another domain in which evaluation mode is likely to have an impact is the placement of advertisements in a media vehicle or in a catalog. In fact, there is some empirical evidence consistent with such effects. In a recent study, Leclerc and Little (1996) used scanner data to look at the effectiveness of coupons in free-standing inserts (FSI). They found that when coupons for brands of the same product category were presented on the same page, they were significantly less effective than when placed on different pages of the FSI. In other words, the fact that the two brands were evaluated jointly seemed to have made them less attractive than if they had been evaluated separately (note that this effect is also consistent with a conflict notion explanation).

In sum, managers should be aware of the significant and highly predictable effect of evaluation mode on consumer judgment and choice. Good products will be hurt in comparison and should be presented separately. Lesser products will be enhanced through comparison and should be presented together.

[*Received October 1996. Revised October 1997. Robert E. Burnkrant served as editor, and Joel Huber served as associate editor for this article.*]

REFERENCES

Bazerman, Max H., George F. Loewenstein, and Sally B. White (1992), ''Reversals of Preference in Allocation Decisions: Judging an Alternative versus Choosing among Alternatives,'' *Administrative Science Quarterly,* 37 (June), 220–240.

Brickman, Philip J. and Donald T. Campbell (1971), ''Hedonic Relativism and Planning the Good Society,'' in *Adaptation-Level Theory: A Symposium,* ed. Mortimer H. Appley, New York: Academic Press.

Helson, Harry (1964), *Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior,* New York: Harper & Row.

Hsee, Christopher K. (1996), ''The Evaluability Hypothesis: An Explanation of Preference Reversals between Joint and Separate Evaluations of Alternatives,'' *Organizational Behavior and Human Decision Processes,* 67 (September), 247–257.

———, Sally Blount, George F. Loewenstein, and Max H. Bazerman (1996), ''Preference Reversals between Joint and Separate Evaluation: A Theoretical Analysis,'' working paper, Graduate School of Business, University of Chicago, 60637.

Huber, Joel, John W. Payne, and Christopher Puto (1982), ''Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypotheses,'' *Journal of Consumer Research,* 9 (June), 90–98.

Johnson, Eric J., John W. Payne, and James R. Bettman (1988), ''Information Displays and Preference Reversals,'' *Organizational Behavior and Human Decision Processes,* 42 (August), 1–21.

Kahneman, Daniel and Dale T. Miller (1986), ''Norm Theory: Comparing Reality to Its Alternatives,'' *Psychological Review,* 93 (April), 136–153.

——— and Amos Tversky (1979), ''Prospect Theory: An Analysis of Decision under Risk,'' *Econometrica,* 47 (March), 263–291.

Leclerc, France and John D. C. Little (1996), ''Reason-Based Decisions in Consumer Promotions,'' working paper, Graduate School of Business, University of Chicago, 60637.

Linville, Patricia W. and Gregory W. Fischer (1991), ''Preferences for Separating and Combining Events,'' *Journal of Personality and Social Psychology,* 60 (January), 5–24.

Nowlis, Stephen M. and Itamar Simonson (1997), ''Attribute-Task Compatibility as a Determinant of Consumer Preference Reversals,'' *Journal of Marketing Research,* 34 (May), 205–219.

Puto, Christopher P. (1987), ''The Framing of Buying Decisions,'' *Journal of Consumer Research,* 14 (December), 301–315.

Russo, J. Edward, Victoria H. Medvec, and Margaret G. Meloy (1996), ''The Distortion of Information during Decisions,'' *Organizational Behavior and Human Decision Processes,* 66 (April), 102–110.

Shafir, Eldar B., Daniel N. Osherson, and Edward E. Smith (1993), ''The Advantage Model: A Comparative Theory of Evaluation and Choice under Risk,'' *Organizational Behavior and Human Decision Processes,* 55 (August), 325–378.

———, Itamar Simonson, and Amos Tversky (1993), ''Reason-Based Choice,'' *Cognition,* 49 (October–November), 11–36.

Simonson, Itamar (1989), ''Choice Based on Reasons: The Case of Attraction and Compromise Effects,'' *Journal of Consumer Research,* 16 (September), 158–174.

Slovic, Paul (1995), ''The Construction of Preference: 102d Annual Convention of the American Psychological Associ-

ation Distinguished Scientific Contribution Award Address,'' *American Psychologist,* 50 (May), 364–371.

Thaler, Richard H. (1985), ''Mental Accounting and Consumer Choice,'' *Marketing Science,* 4 (Summer), 199–214.

——— and Eric J. Johnson (1990), ''Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choices,'' *Management Science,* 36 (June), 643–660.

Tversky, Amos (1969), ''Intransitivity of Preferences,'' *Psychological Review,* 76, 31–48.

——— and Daniel Kahneman (1973), ''Availability: A Heuristic for Judging Frequency and Probability,'' *Cognitive Psychology,* 5 (September), 207–232.

——— and Daniel Kahneman (1974), ''Judgment under Uncertainty: Heuristics and Biases,'' *Science,* 185 (September), 1124–1131.

——— and Daniel Kahneman (1981), ''The Framing of Decisions and the Psychology of Choice,'' *Science,* 211 (January), 453–458.

——— and Daniel Kahneman (1991), ''Loss Aversion in Riskless Choice: A Reference-Dependent Model,'' *Quarterly Journal of Economics,* 106 (November), 1039–1063.

——— and Eldar Shafir (1992), ''Choice under Conflict: The Dynamics of Deferred Decision,'' *Psychological Science,* 3 (November), 358–361.