JOURNAL OF CONSUMER PSYCHOLOGY, *4*(1), 61–83

The Effects of Accessibility of Standards and Decision Framing on Product Evaluations

Yaacov Schul and Yoav Ganzach
Hebrew University

In this study, we explore the cognitive process by which people evaluate consumer products. We examine how people's information about good and bad product classes influences the evaluation of product exemplars. Subjects in three experiments learned what constitutes a good alloy (the good standard) and a bad alloy (the bad standard). Then they were exposed to a series of exemplars whose features partially overlapped with the good and the bad standards. In the first two experiments, features associated with the good standard had a greater impact on judgment than features associated with the bad standard. This positive/negative asymmetry was stronger when the decision was framed positively (e.g., how good is this alloy) than when it was framed negatively (e.g., how bad is this product). Also, the asymmetry was stronger when the standards had to be accessed from memory than when they were visually available at the time subjects evaluated the exemplars. In the third experiment, the addition of features that signaled the absence of good characteristics had more influence on product evaluations than the addition of features that signaled the absence of bad characteristics. Implications of these findings for current models of evaluative judgment are discussed.

In this study, we explore the cognitive process by which people evaluate consumer products. We examine how the accessibility of the information people have about good and bad product classes influences the evaluation of product exemplars. To illustrate the issue, imagine you are told that a particular car runs fast and rusts fast. To evaluate the car, you must decide whether "runs fast" adds to or detracts from the quality of the car. To do so, you may activate from memory what you know about good as well as bad cars and then compare this information to the features of the given car. "Runs fast" will

Requests for reprints should be sent to Yaacov Schul, Department of Psychology, Hebrew University–Mount Scopus, Jerusalem, Israel. E-mail: mschul@pluto.mscc.huji.ac.il.

make your evaluation of the car more positive to the extent that it is part of the representation of good cars and more negative to the extent that it is part of the representation of bad cars. Let us term the representation of what one knows about the good product class (e.g., what constitutes a good car) the *good standard* and what one knows about the bad product class the *bad standard.* Standards make it possible to distinguish product classes and allow consumers to evaluate specific brands (Alba & Hutchinson, 1987; Sujan, 1985). Note that our working assumption is that standards consist of networks of features. Such an assumption has proven useful in recent models of learning (Kruschke, 1992), judgments (Smith & Zarate, 1992), and typicality (Loken & Ward, 1990). In this article, we explore whether and how accessibility of standards influences product evaluation.

Standards might differ in accessibility. Let us consider a case in which it is more difficult for a person to access the negative standard than the positive standard (e.g., when the person thinks about cars, she or he spontaneously thinks primarily about the features of her or his ideal car). In this case, there will be a higher likelihood that "runs fast" will be identified as a positive feature than that "rusts fasts" will be identified as a negative feature. As a result, "runs fast" should have more impact on judgments than "rusts fast." In the extreme case, if the bad standard cannot be accessed at all, but the good standard is readily available, consumers will not detect "rusts fast" as a negative feature, and it will, therefore, have little impact on their evaluation of cars. On the other hand, the feature "runs fast" is likely to be recognized as a good feature and have an impact on judgment.

The first issue we study in this article, therefore, involves the accessibility of good and bad standards (cf. Higgins, 1990). At the most general level, we hypothesize that the impact of features associated with a more accessible standard will be greater than the impact of features associated with a less accessible standard. In particular, the research discussed later suggests that positive standards are more accessible than negative standards. Consequently, we hypothesize that, other things being equal, positive features have more impact on product evaluations than negative features.

Matlin and Stang (1978) reviewed the literature on the way people deal with pleasant and unpleasant information. They suggested the *Pollyanna principle,* namely, that in general pleasantness predominates over unpleasantness, and in particular positive information has priority in memory over negative information. We think that this principle also applies to the good and bad standards associated with different product classes. In addition to the general principles argued for by Matlin and Stang (1978), the asymmetry in accessibility makes sense in the context of consumers' typical goal-directed activity. As Meyer (1987) suggested, consumers often have more use for features that characterize good products, products they may actually buy, than for features of bad products. As a result, people may access positive standards more easily than negative standards out of procedural efficiency (Smith & Branscomb, 1987). Indeed, learning the positive features of an unfamiliar product is faster and more accurate than learning its negative features (e.g., Meyer, 1987).

At a first glance the hypothesis that positive features have more impact than negative features may appear inconsistent with the well-known *negativity bias.* To wit, it has been documented that under well-specified conditions (see Ganzach, in press; Skowronski & Carlston, 1989, for reviews) negative features dominate positive features in determining judgments. However, the process that leads to the evaluative response consists of several cognitive phases (e.g., Anderson, 1981; Burnstein & Schul, 1982; Tourangeau & Rasinski, 1988; Wyer & Srull, 1989), so that negativity bias in one phase does not preclude positivity bias in another. Most of the past research on negativity bias was concerned with a relatively late phase of the decision-making process and focused on the different decision rules people utilize in integrating information. Our research focused on a relatively early phase of the process by exploring the nature of the information recruited for the decision. To recapitulate, our discussion so far suggests that (a) the impact of a feature depends on the accessibility of the standard to which it pertains, and (b) positive standards are more accessible than negative standards. This suggests our first hypothesis:

> (H1): Other things being equal, features that pertain to a positive standard have more impact on product evaluations than features that pertain to a negative standard.[1]

The higher accessibility of positive than negative standards can be influenced by compatibility with the decision frame. The positive standard is more likely to be triggered when the decision is framed positively (e.g., a consumer is asked to evaluate how good the given car is) than when the decision is framed negatively (e.g., how bad the car is). This is due either to facilitation at the level of representation (Bargh, Chaiken, Govender, & Pratto, 1992; Gillund & Shiffrin, 1984) or to strategies of information search and utilization (cf. Lehman, Krosnick, West, & Li, 1992; Skov & Sherman, 1986). In particular, this research suggests that people use confirmatory search so that information fitting the focal hypothesis is more likely to be accessed than information that fits alternative hypotheses.

Recently, Shafir (1993), as well as research in our own laboratory (Ganzach & Schul, in press) demonstrated that people's judgments and choices can be dramatically influenced by changing the decision frame, even when the given

[1]Research on opinion and attitude change posits numerous other factors that can affect the impact of a feature. In particular, it suggests that this effect depends most importantly on its diagnosticity (Skowronski & Carlston, 1989), as well as on the characteristics of the source, order position, and the way it is presented (see Eagly & Chaiken, 1993, for a recent review).

information remains constant. This occurs because frame-compatible information has more impact on judgments than frame-incompatible features. Application of these findings to our situation is straightforward: If our analysis is correct, features associated with the positive standard should have more impact when judgments are made under a positive frame rather than a negative frame. Conversely, features associated with the negative standard should have more impact when decisions are framed negatively than when they are framed positively. Therefore, our second hypothesis is:

> (H2): The impact of features will be higher when they are associated with a frame-compatible rather than a frame-incompatible standard.

Note that in positive framing, H1 and H2 work in the same direction. Both claim that the impact of positive features on evaluations will be higher than that of negative features. However, in negative framing, H1 and H2 make opposite predictions: According to H1, positive features should have more impact than negative features, but according to H2, negative features may have more impact than positive features. These opposite tendencies under negative framing may lead to either greater impact of positive features (over negative ones), greater impact of negative features (over positive ones), or equal impact of positive and negative features. Although we cannot specify a priori whether positive or negative features will have more impact under negative framing, the previous discussion suggests that valence of framing and valence of features should interact—that is, the difference between the impact of positive and negative features should be more pronounced when the decision is made under a positive frame rather than a negative frame.

It is difficult to test these hypotheses with familiar products because people differ greatly in their knowledge bases and their goals. Consequently, they may possess different standards for the same product. For example, although a teenager may consider "runs fast" to be a positive feature for a car, a parent may find this same feature negative. To reduce interpersonal variability, the present study followed Meyer's (1987) method and provided subjects with product information in a domain that was completely novel to them. Subjects were first taught the characteristics of good and of bad alloys. We refer to these alloys as the good and the bad standards. Because we were able to manipulate the characteristics of good and bad standards, we were able to completely control the positivity and negativity of the product's features and ensure that these features were equivalent in all respects except for their manipulated valence. After learning, subjects were presented with test alloys whose features partially overlapped those of the good and the bad standards learned earlier (see Method section). Subjects were asked to evaluate the test alloys, and their judgments allowed us to test the aforementioned hypotheses. Although the utilization of a novel domain of knowledge introduces artificiality to the experimental situation, it has the benefit of reducing the influence of extra-experimental factors that can modify the impact of a feature.

EXPERIMENT 1

Overview

Experiment 1 consisted of five phases. In the first phase, subjects were presented with four features of a good alloy (the good standard) and four features of a bad alloy (the bad standard). Subjects were told that they would be tested on their knowledge of these characteristics later in the experiment. During the second phase, subjects responded to five general-knowledge questions. This served to separate the learning from the testing phase. In the third phase, subjects were tested on their knowledge of the good and the bad standards. If they made an error, they repeated the sequence of the three phases. Subjects proceeded to the next phase only after they identified the characteristics of the good and the bad standards with no error. In the fourth phase, subjects were presented with 64 test alloys, which we termed *exemplars.* After viewing each exemplar, subjects rated its quality on a 9-point response scale. Finally, after completing all 64 judgments, subjects were tested again for their knowledge of the good and the bad standards.

Stimuli

Exemplar alloys consisted of either three or four features. Features were selected from a pool of 14, which are listed in Table 1. There were three types of features: Four features were good (belonging to the good standard), four features were bad (belonging to the bad standard), and six features were neutral (belonging to neither of the standards—that is, subjects did not encounter these features during the learning phase). Between subjects, each feature served equally often as good, bad, and neutral. That is, the same feature was good for one subject, bad for a second subject, and neutral for a third subject. More details about the counterbalancing procedure employed in the experiment appear in the Appendix.

Because we wanted to assess the impact of the good and bad features independently, it was important to compare them to features that were not part of the standards. For this purpose, the two good features (e.g., "Vendor: Netherlands" and "Temperature of crystallization: exactly at freezing point" in counterbalancing Condition 1 in Table 1) had two corresponding neutral features (e.g., "Vendor: Belgium" and "Temperature of crystallization: below freezing point"). That is, subjects in counterbalancing Condition 1 learned that good alloys are manufactured in the Netherlands, but nothing (good or bad)

TABLE 1
The 14 Features Employed in the Experiments

		Counterbalancing Condition					
Dimension	*Feature*	*1*	*2*	*3*	*4*	*5*	*6*
Cooling method							
	Cooling by water	G	B	N	B	G	N
	Cooling by air	B	G	B	N	N	G
Process pressure							
	20 atmospheres	G	B	N	B	G	N
	50 atmospheres	B	G	B	N	N	G
Vendor							
	Netherlands	G	N	G	B	N	B
	Belgium	N	G	B	G	B	N
Temperature at crystallization							
	Exactly at freezing point	G	N	G	B	N	B
	Below freezing point	N	G	B	G	B	N
Heating method							
	Continuous	N	B	G	N	G	B
	Intermittent	B	N	N	G	B	G
Additives							
	Zinc	N	B	G	N	G	B
	Carbon	B	N	N	G	B	G
Type of furnace							
	Ceramic	N	N	N	N	N	N
Quenching method							
	Fast	N	N	N	N	N	N

Note. Each column shows the valence of the features in a particular counterbalancing condition (G = good feature, B = bad feature, N = neutral feature).

was stated about alloys manufactured in Belgium. We refer to these four features as the *GN set* to denote the correspondence between the two good features (G) and the two neutral features (N).

Similarly, two bad features (e.g., "Heating method: continuous" and "Additive: Zinc" in counterbalancing Condition 1 in Table 1) had two corresponding neutral features (e.g., "Heating method: intermittent" and "Additive: Carbon"). We refer to these four features as the *BN set* to denote the correspondence between the two bad features (B) and the two neutral features (N). Our major analyses examined the impact of features from the GN and BN sets.

Sixty-four exemplar alloys were constructed from the 14 features. Each exemplar had either three or four features that were selected randomly for each subject according to a pattern with the following constraints: (a) The number of features within each type (good, bad, and neutral) ranged between zero and two. (b) There were 16 exemplars that had two good features and 16 exemplars that had two bad features. (c) Thirty-two of the exemplars had one good feature and 32 of them had one bad feature. The Appendix contains the details of the procedure for constructing the exemplar alloys.

Procedure

Subjects were run in groups ranging in size from 2 to 6. Each subject was seated in front of a personal computer that controlled the instructions and stimulus presentation as well as response collection. There was virtually no intervention of the experimenter during the experimental session. Subjects were first presented with the good or the bad standard (depending on the experimental condition) and were allowed as much time as they wanted to learn it. Then they signaled to the program to shift to the other standard, and again, they had as much time as they wanted to learn it. Next, subjects had to answer five forced-choice general knowledge questions. Shortly after completing these questions they were presented with the 14 features from Table 1 and were asked to select the four features that characterize the good (bad) standard. Once they responded, their choices were erased from the screen, and they were again presented with the list of 14 features. This time they were asked to select the four features of the bad (good) standard. The order of the standards in the test corresponded to the order of presentation of the standards during learning. If a subject was completely accurate in the test, he or she proceeded to the judgment phase. Otherwise, the sequence of learning/general knowledge/testing was repeated.

At the onset of the judgment phase, subjects were told that they would be asked to rate the quality of 64 exemplar alloys. Subjects in the positive-frame condition were asked to evaluate how good the exemplar alloy was on a scale ranging from *a little* (1) to *very* (9). Subjects in the negative-frame condition were to rate how bad the exemplar alloy was on the same 9-point scale. The 64 exemplars were presented in four blocks of 16 trials. Between blocks, subjects were presented with the characteristics of the good and the bad standards.

Following the 64 judgments, subjects were retested on their memory for the characteristics of the good and the bad standards.

Design

The design involved one theoretically important between-subjects factor, namely, the decision frame. It was manipulated by the formulation of the evaluation question (good vs. bad). In addition, several counterbalancing procedures were employed. First, half the subjects were presented with the good standard before the bad standard, whereas the remaining subjects were presented with the standards in the opposite order. Second, features were rotated between sets, so that each feature was equally often a good, a bad or a neutral feature. Finally, each subject saw the 64 exemplar alloys in a different random order, and the features of each exemplar were ordered randomly, differently for each subject.

Subjects

Forty-eight students at the Hebrew University participated in the experiment. Subjects were assigned randomly to experimental conditions.

RESULTS

Learning

On the average, subjects took about two (actually, 2.04) learning replications to achieve a completely accurate knowledge of the standards. Twenty-four subjects (out of the 48) did not make any errors, and the maximal number of replications was 6. For each subject, we computed the number of errors made during the identification of the positive standard and the number of errors in the identification of the negative standard. These were subjected to a two-way mixed model analysis of variance (ANOVA) in which the order of learning (good/bad vs. bad/good) was the between-subjects factor and the valence of the feature (good vs. bad) was the within-subjects factor. The analysis revealed that although subjects were more accurate in the first-learned (and first-tested) standard (mean number of errors = .75) than in the second-learned one (mean number of errors = 1.08), F (1, 46) = 6.47, $p < .05$, there was no overall difference in the ease of learning of good and bad standards. That is, the number of identification errors was similar for good and bad standards, $F < 1$. This makes it unlikely that the judgment effects (to be discussed later) are the consequence of differential learning. This last finding does not parallel the findings of Meyer (1987). However, because learning in Meyer's study involved the discovery of prediction rules for positive and negative information and because our learning procedure involves rote learning, the findings in the two studies cannot be directly compared.

Not surprising, following the extensive learning procedure and the four presentations of the standards between blocks, subjects made very few errors in identifying the standards following the 64 judgments. On the average, subjects made 0.15 errors, and the number of errors did not vary as a function of valence of standards, the order of learning, or the scale used to make judgments.

Exemplar Evaluations

Comparison between responses to the three-feature and the four-feature exemplars revealed that the addition of a feature from the N set had little effect on judgments. Also, the order in which the good and the bad standards were learned made no difference. Hence, we shall not discuss these factors further. Prior to the analysis reported later, judgments on the bad scales were transformed (subtracted from 10) so that high judgment responses indicated more positive evaluations in both the good and the bad scale conditions.

Our main interest focuses on comparing the influence of features from GN and BN sets. If a positive standard is more accessible than a negative standard, then judgments should be influenced more by variations in the GN set (i.e., replacing a good feature by a neutral feature) than by variation in the BN set. To compare these effects, we employed an effect-coding procedure (Cohen & Cohen, 1975). Specifically, we created a variable (labeled *GN-dichotomy*) that assigned the value of + 1 to exemplars in which the GN feature (i.e., the feature belonging to the GN set) was good and the value of − 1 to exemplars in which the GN feature was neutral. Similarly, a second variable (labeled *BN-dichotomy*) coded exemplars as +1 when the BN feature was neutral and as −1 when the BN feature was bad (i.e., when it was part of the bad standard). For each subject, we computed two correlations, one between the GN-dichotomy and judgment and the other between the BN-dichotomy and judgment. These correlations indicated the extent to which judgments were increased by the good feature relative to its corresponding neutral feature (the GN effect) and the extent to which judgments were decreased by the bad feature relative to its corresponding neutral feature (the BN effect).[2]

Before analyzing these correlations, we transformed them to *z*-scores according to the formula suggested by Fisher. This transformation is recommended because the sampling distribution of nonzero correlations is skewed. The means of the transformed correlations are presented in Table 2. The transformed correlations were analyzed in a two-way mixed-model ANOVA in which the type of scale (good vs. bad) served as a between-subjects factor and the valence of the effect (GN vs. BN) served as a within-subjects factor.[3] The analysis revealed that, in support of H1, variation among the GN features

[2]Because of the way we constructed our stimuli, the variables GN-dichotomy and BN-dichotomy are completely orthogonal.

[3]In addition to the analyses of the GN-effect and the BN-effect, we performed an ANOVA on the raw judgments in each of the three experiments. In these analyses, the BN and the GN dichotomies were employed as within-subject factors. Analyses of the raw evaluations revealed patterns of findings similar to the analyses on the GN-effect and BN-effect reported in the text. However, because the relative impact of the GN and BN factors cannot be assessed from an ANOVA on the raw evaluations, this type of analysis cannot be utilized to test H1 and H2.

TABLE 2
Mean Transformed Correlation of GN and BN Dichotomies with Judgment (Experiment 1)

	Effect	
Frame	*GN*	*BN*
Good	.42	.16
Bad	.32	.22

had significantly more impact on judgments (mean $z = .37$) than variation among BN features (mean $z = .19$), $F(1, 46) = 33.10, p < .01$. According to H2, the GN and BN effects should depend on the frame elicited by the response scale. The findings are consistent with H2, as indicated by a significant valence-by-frame interaction, $F(1, 46) = 6.24, p < .02$. Inspection of Table 2 reveals that responses on the bad scale were less influenced than responses on the good scale by variation among the GN features, but were more influenced by variation among the BN features.

DISCUSSION

The findings in Experiment 1 were consistent with the hypothesis that features relevant to the positive standard had more impact than those relevant to the negative standard (H1). Moreover, this positive/negative asymmetry was more pronounced in positive than in negative framing (H2). The higher impact of GN than BN features cannot be attributed to the learning procedure. There was virtually no difference between learning of positive and negative standards. The difference also cannot be attributed to the nature of the features themselves (e.g., positive features are shorter, less ambiguous, etc.), because the same features served as positive features for some subjects and negative features for other subjects. We believe that the findings of Experiment 1 are consistent with the hypothesis that the greater impact of GN features occurred because positive standards were more accessible than negative standards, thus making the features associated with them more impactful.

Experiment 2 took this idea a step further. If the asymmetry in impact depends on the accessibility of the standards, it is hypothesized (H3) that the positive/negative asymmetry will be maximal when subjects must access the standards from memory. It will be minimal, however, when the standards are highly available to the subjects at the time the judgments are made.

EXPERIMENT 2

Experiment 2 manipulated the availability of the standards at the time of judgments directly. The low-availability condition was identical to Experiment 1; namely, subjects learned the standards during the learning phase and then had to access them from memory at the time they made their judgments. In the high-availability condition, subjects did not learn the standards. Rather, the two standards were presented on the computer screen together with the exemplar alloys, so that subjects could consult them at the time they made their judgments. The experiment also included an intermediate condition. In this condition, subjects learned the standards, as did subjects in the low-availability condition, but the standards were available on the screen during judgment, as in the high-availability condition. We reasoned that having already learned the standards, subjects might be less likely to consult them on the screen and therefore the positive/negative asymmetry would be more pronounced in the intermediate condition than in the high-availability condition.

Experiments 1 and 2 differed in another way. The impact of framing was studied in Experiment 1 by having subjects respond on one of two different unipolar scales. That is, subjects in Experiment 1 evaluated either how good or how bad the exemplar was. It could be argued that because a good scale is not the direct opposite of a bad scale, ratings on the two scales cannot be directly compared. Although we do not think that this argument should affect our interpretation of the findings, Experiment 2 utilized a subtler way to frame decisions, one not vulnerable to this argument. Experiment 2 employed bipolar response scales that varied only in the positioning of their labels. Half the subjects responded on a scale that ranged from *bad* (1) to *good* (9). The other half responded on a scale that ranged from *good* (1) to *bad* (9). We reasoned that because our subject population habitually scans information from right to left (Hebrew is read from right to left) and because the right-side label was always associated with the highest response alternative, the right-side anchor would focus attention on its standard more than the left-side anchor. Therefore, we hypothesized that the positive/negative asymmetry would be more pronounced on the good/bad scale[4] than on the bad/good scale.

METHOD

The stimuli and procedure were identical to those in Experiment 1, except for the response scales and the availability manipulation. The low-availability condition was identical in procedure to Experiment 1. Subjects in the high-

[4]The good/bad scale refers to the scale in which the right end point was labeled with *good* and the left endpoint labeled with *bad*.

availability condition did not undergo the learning phase, general-knowledge questions, and testing phase. They were, however, presented with the good and the bad standards on the screen throughout the judgment phase. The positioning of the standards of the screen (top-left and top-right) corresponded to the labels (good and bad) of the response scale. Subjects in the intermediate condition learned the standards the same way as subjects in the low-availability condition, but were presented with the standards on the screen during judgment, like subjects in the high-availability condition.

Subjects

Seventy-two subjects participated in the experiment. They were randomly assigned to conditions, 12 subjects to each of the six conditions (availability [3] × response scale [2]).

RESULTS AND DISCUSSION

Learning

The analysis was based on subjects from the low-availability and the intermediate conditions, because subjects in the high-availability conditions did not undergo the learning phase. As in Experiment 1, subjects took about two (actually, 1.78) learning replications to achieve a completely accurate identification of the standards. However, this experiment differed from Experiment 1 in that we could not reject the null hypothesis that there were no differences in the number of errors between standards presented and tested first and second. As in Experiment 1, there were virtually no errors in the identification of the standards following the judgments.

Exemplar Evaluations

We analyzed the effect of positive and negative features on judgments using the same procedure employed in Experiment 1. For each subject, we computed the correlation between judgments and GN-dichotomy and the correlation between judgment and BN-dichotomy. These correlations were transformed to z-scores and subjected to a three-way mixed-model ANOVA in which the availability condition (low, intermediate, high) and scale (good/bad vs. bad/good) served as between-subjects factors, and the valence of the effect (GN vs. BN) served as the repeated-measure factor. The means of the transformed correlations are presented in Table 3.

All three hypotheses were supported by the results. First, the effect of GN (mean $z = .47$) on judgment was significantly stronger than the effect of BN

TABLE 3
Mean Transformed Correlations of GN and BN Dichotomies With Judgments (Experiment 2)

	Low Availability		*Intermediate*		*High Availability*	
Frame	*GN*	*BN*	*GN*	*BN*	*GN*	*BN*
Good/bad	.55	.13	.54	.28	.40	.28
Bad/good	.39	.26	.48	.30	.47	.39

(mean z = .27), $F(1, 66) = 73.37, p < .01$. Thus, this experiment also supports the hypothesis that features associated with the good standard have more impact on judgment than those associated with the bad standard (H1). Second, consistent with H2, the asymmetry varied as a function of the scale, as indicated by a significant valence by scale interaction, $F(1, 66) = 9.06, p < .01$. Third, consistent with H3, the positive/negative asymmetry varied as a function of the availability condition, as indicated by the Valence × Availability interaction, $F(2, 66) = 4.88, p < .01$. Finally, the triple interaction was significant, $F(2, 66) = 3.14, p < .05$. This indicated that the asymmetry between the impact of GN and BN features varied according to the condition-by-scale interaction. We discuss each of these effects in turn.

GN features had more impact on judgment than BN features. Inspection of Table 3 revealed that this occurred in each of the availability conditions. However, the superiority of GN over BN varied as a function of the availability conditions. If one averages the entries in the two rows in Table 3 and compares the impact of GN and BN effects, it becomes apparent that the positive/negative asymmetry was least pronounced in the high-availability condition (mean difference between GN and BN = .10), larger in the intermediate condition (mean difference = .22), and largest in the low-availability condition (mean difference = .28). Pairwise comparisons indicate that, as expected by H3, subjects in the high-availability condition showed significantly more positive/negative asymmetry than subjects in the low-availability condition, $F(1, 66) = 9.32, p < .01$. The intermediate condition elicited significantly larger positive/negative asymmetry than the high-availability condition, $F(1, 66) = 4.42, p < .05$, and was not significantly different from the low-availability condition, $F(1, 66) = .90$).

Averaging the entries in Table 3 over the three availability conditions revealed that, as in Experiment 1, the asymmetry was more pronounced when subjects responded on the good/bad scale (mean difference between GN and BN = .27) than when they responded on the bad/good scale (mean difference = .14). We believe that in the former case the scale acted as a positive decision frame, whereas in the latter case it acted as a negative frame. The three-way interaction suggests that the effect of availability on the positive/negative asymmetry was more pronounced under the good/bad scale than under the bad/good scale. This pattern is consistent with the idea that in low-availability condition the BN features were particularly unlikely to trigger the negative standard when the framing was positive. The GN features, in contrast, were more likely to trigger the positive standard under a negative frame because the positive standard had relatively high accessibility. This asymmetry, however, was reduced when both standards become more available, leading to a three-way interaction.

The results of Experiment 2 support the three hypotheses, namely: (a) Features associated with positive standards have more impact than features associated with negative standards; (b) The positive/negative asymmetry is magnified when judgments are cast within a positive frame; and (c) The asymmetry is magnified when standards are retrieved from memory. These findings are consistent with our general hypothesis that positive standards are more accessible than negative standards.

EXPERIMENT 3

In the first two experiments, we explored the impact of GN features relative to that of BN features. Specifically, we compared two differences: (a) Difference between judgments of exemplars including good features and judgments of exemplars including neutral features from the GN set (i.e., the GN effect); and (b) Difference between judgments of exemplars including neutral features from the BN set and judgments of exemplars including bad features (i.e., the BN effect). There are, thus, two possible sources for the higher impact of GN features relative to BN features: One possibility is that it may reflect a difference between the impact of good features and the impact of bad features (i.e., good features have more impact than bad features). The other possibility is that it may reflect a difference in the impact of the two neutral features: Neutral features from the GN sets (denoted as N_G features) may have more impact than neutral features from the BN sets (denoted as N_B features).

In particular, if good standards tend to be accessed spontaneously more often and earlier than bad standards, people may be more likely to realize that a neutral feature from the GN set signals the absence of a good feature. As a result, it is possible that although a neutral feature from the BN set does not influence the valence of judgment (i.e., it is perceived to be "truly neutral"), a neutral feature from the GN set makes judgments somewhat negative because it signals the lack of a good feature. The purpose of Experiment 3, therefore, was to disentangle the role of positive, neutral, and negative features in the GN and BN effects observed in the previous two experiments.

METHOD

Subjects

Forty-eight students at the Hebrew University participated in the experiment. Subjects were assigned randomly to one of the two framing conditions.

Procedure

Subjects learned the good and the bad standards the same way as subjects in Experiment 1. Then, once they could identify the standards perfectly, they were presented with 64 exemplars. They evaluated these exemplars on the same unipolar scales as subjects in Experiment 1. That is, half of the subjects evaluated how good each alloy was while the rest evaluated how bad it was.

Stimuli

Subjects evaluated 64 exemplars in the experiment. Fifty-two exemplars were part of the experimental design. The rest were filler exemplars. The fillers served to equate the number of judgments in the different experiments and to mask the structure of the experimental exemplars. The 52 experimental exemplars contained either one, two, or three features that were good, bad, or neutral. As in previous experiments, the frequency of exemplars with one, two, or no positive features was identical to that of exemplars with one, two, or no bad features. The last section in the Appendix contains more details about the construction of the experimental exemplars.

RESULTS AND DISCUSSION

Learning

There was no overall difference in the learning of features associated with the good and the bad standards, $F < 1$. As in Experiment 1, features of the standard that was learned first were learned with fewer errors ($M = .42$) than features of the standard that was learned second ($M = .93$), $F(1, 46) = 8.99$, $p < .01$. There were very few errors in identifying the features following the judgments ($M = .11$).

Exemplar Evaluations

Only judgments involving the experimental exemplars were analyzed. Judgments given under the bad response scale were transformed, so that higher judgments indicated more positive evaluations for all subjects. The analyses reported later compare and contrast the impact of good and bad features as well as the impact of their corresponding neutral features. To do so we created for each subject four *effect variables.*

The first effect reflects the increase in positivity of product evaluation caused by the addition of a good feature. To compute this effect, we assigned a value of +1 to exemplars in which the GN feature was G and the value of 0 to exemplars in which the GN feature was missing. We labeled this effect the *G-M effect* to denote that it compares judgments of exemplars in which the GN feature was good (G) to judgments of exemplars in which the GN feature was missing (M). In other words, this effect estimates the impact of the addition of a good feature.

The second effect reflects the impact of the addition of a neutral-GN feature (denoted by N_G). To create the N_G–M effect we assigned the value −1 to exemplars in which the GN feature was N (i.e., not part of the good standard) and the value of 0 to exemplars in which the GN feature was missing. Note that a positive correlation between the G–M index and the judgment response indicates that good features make the judgment more favorable. A positive correlation between the N_G–M index and the judgment response indicates that the neutral feature from the GN set makes the judgment more negative (i.e., absence of a feature from the good standard reduces the quality of the product).[5]

Analogously, we created the third and the fourth effects, the B–M and the N_B–M effects. The B–M effect reflects the decrease in positivity of judgment caused by the addition of a bad (denoted by B) feature. To create the B–M effect, we assigned the value of −1 to exemplars in which the BN feature was B and the value of 0 to exemplars in which the BN feature was missing. The N_B–M effect reflects the impact of the addition of a neutral-BN feature (denoted by N_B). To create the N_B–M effect, we assigned the value of +1 to exemplars in which the BN feature was neutral (i.e., not a part of the bad standard) and a value of 0 to exemplars in which the BN feature was missing. A positive correlation between the B–M index and the judgment response indicates that the addition of a bad feature reduces the evaluation of a product. A positive correlation between the N_B–M and the judgment response indicates that the absence of a negative feature makes the product appear more positive.

For each subject, we transformed the four correlations to *z*-scores. The means of the transformed correlations are presented separately for the two frame conditions in Table 4. The four transformed correlations were analyzed in a two-way mixed-design ANOVA in which decision frame (positive vs. negative) served as the between-subjects factor and the four correlations as the within-subject factor. We explored three main questions: First, we compared

[5]The choice of sign is arbitrary. We use this convention to facilitate comparisons between the correlations.

TABLE 4
Mean Transformed Correlations of G–M, N_G–M, B–M, and N_B–M With Judgment (Experiment 3)

	Effect			
Frame	*G–M*	*N_G–M*	*B–M*	*N_B–M*
Good	.42	.11	.34	.00
Bad	.31	.10	.37	.04

the impact of the addition of good features to the addition of bad features. Second, we compared the influence of the addition of neutral features from GN sets to that of neutral features from BN sets. Third, we compared the effect of features that were part of a standard (i.e., the G and B features) to the effect of features not part of a standard (N_G and N_B).

Inspection of Table 4 reveals that the impact of adding a G feature was similar to that of adding a B feature (.37 vs. .36 respectively, $F < 1$). In accordance with H2, however, the interaction with the frame condition was significant, $F(1, 46) = 5.12, p < .05$. Adding G features had more influence on judgments than adding B features (.42 vs. .34) when subjects evaluated how good the exemplar was. Adding G features had less influence than adding B features (.31 vs. .37) when subjects evaluated how bad the product was. Thus, as in the previous experiments, the positive/negative asymmetry was reduced (and actually reversed) under a negative-decision frame.

The analysis also suggests that the effect of adding N_G features was significantly greater ($M = .10$) than that of adding N_B features ($M = .02$), $F(1, 46) = 6.01, p < .05$. In other words, although adding N features from the GN set lowered product evaluation, $t(46) = 3.45, p < .01$, adding N features from the BN set had a small and nonsignificant effect on the evaluation, $t(46) = .78, p > .4$). This pattern did not interact with the framing condition, $F < 1$. This result is interesting because the good or the bad features are likely to trigger the standards compatible with them. Therefore, the spontaneous accessibility of a standard is best assessed by exploring the impact of N_G and N_B features. Indeed, N_G features had more impact than N_B features in both frame conditions. We believe that N_G features signaled to subjects absence of a good feature and therefore, lowered judgments. N_B features, in contrast, were less likely to remind subjects of a bad standard. Therefore, such features did not signal the absence of a bad feature. As a consequence, exemplars containing N_B features were evaluated similarly to exemplars with missing bad features.

An alternative explanation for the greater impact of N_G than N_B features is that the effect of missing information from a good standard is different than the effect of missing information from a bad standard.

According to this explanation, missing information functions analogously to neutral features by reminding individuals of the absence of the good or the bad feature. This alternative suggests, therefore, that evaluation of exemplars characterized by missing information on a good dimension are lowered, whereas evaluation of exemplars characterized by missing information on a negative dimension are increased.[6] Although this possibility might be theoretically important, and although our data cannot reject it, we think that it should not change our interpretation of the findings. First, the effect of missing information should be small in magnitude, even compared with that of neutral features, because neutral features are more likely to activate the relevant dimension than missing information (see Houston, Sherman, & Bacon, 1989; Jaccard & Wood, 1988; Simmons & Lynch, 1991). Second, the analysis may actually underestimate the effect of neutral features, especially in the case of N_G features, because missing information is assumed to act similarly to presence of the corresponding neutral feature (i.e., to make judgments more negative). Therefore, the $N_G - M$ (and possibly $N_B - M$) difference might be smaller than the actual impact of N_G (and N_B). Notwithstanding, because of the theoretical importance of issue of missing information (see Dick, Chakravarti, & Biehal, 1990; Simmons & Lynch, 1991), more research is needed on the relations among missing, neutral, and good/bad features. In particular, the previous analysis suggests that missing information need not always lead to discounting. It may lead to a more positive product evaluation if it signals the absence of a negative feature. This can serve as another test to the suggestion of Simmons and Lynch (1991) that the effect of missing information is not mediated by inferences about its value.

Finally, the analysis revealed that adding features that were part of one of the standards (i.e., G and B features) had more impact on judgments than adding features that were not part of a standard (i.e., N_G and N_B features), $F(1, 46) = 92.57$, $p < .01$. As can be seen from Table 4, the average impact of the former was about .36, whereas the average impact of the latter was about .06. This pattern did not interact with the nature of the response scale or with the valence of the standard ($p > .1$ for both interactions). This tallies nicely with research on hypothesis testing. The research suggests that people (as well as animals) have difficulties in using nonevidence or nonoccurrences as cues for making judgments and solving problems (Einhorn & Hogarth, 1978; Wason, 1968). Accordingly, it was expected and found that the impact of features that

[6]Comparison of the three-feature and the four-feature exemplars employed in Experiment I can yield an estimate for the effect of a missing neutral feature. The effect of missing neutral information was fairly weak (mean correlation = .027, $p > .1$), indicating that three-feature exemplars were evaluated slightly more positively than four-feature exemplars. This suggests that the absence of a feature by itself was not associated with negative evaluation.

were part of a standard (occurrences) was significantly larger than that of features not part of a standard (nonoccurrences).

GENERAL DISCUSSION

Subjects in three studies learned what constitutes a good alloy (the good standard) and a bad alloy (the bad standard). Then they were exposed to series of exemplars whose features partially overlapped those of the good and the bad standards. In the first two studies, it was found that features associated with a good standard had greater impact on judgment than features associated with a bad standard. This positive/negative asymmetry was stronger when the decision was framed positively (e.g., how good this alloy is) than when it was framed negatively (e.g., how bad this alloy is). Also, the asymmetry was stronger when the standards had to be accessed from memory than when they were visually available at the time subjects evaluated the exemplars. The third study suggested that addition of a good feature has an impact of the same strength (but in the opposite direction) as the addition of a bad feature. However, the addition of features that signaled the absence of good (N_G) had more influence on product evaluations than that of features signaling the absence of bad (N_B). The former made the judgments less positive, whereas the latter had a negligible impact on judgments.

This pattern of findings is consistent with the idea that at the time individuals evaluate products they compare the features of the given product to standards they have. This comparison drives, in part, the impact the features have on judgments. Positive features had a larger impact on product evaluation because the positive standard is generally more accessible. Such a difference in accessibility is quite adaptive in a typical buying situation, when consumers have to determine the best among many good alternative products. In such situations, positive cues are more informative than negative cues. Consumers can operate optimally in such cases if they can match each alternative to the ideal product—the good standard in our study.

Although the results of this article suggest that good standards are more accessible than bad standards, there may be situations in which the reverse is true. For example, as one's commitment to the evaluation increase, the negative alternatives may become more salient. Ganzach (in press) has recently shown that choices involve more commitment than judgments. This reasoning suggests that in situations in which consumers perceive the decision problem as a choice of the least bad alternative among several bad products, the accessibility of the bad standard will be higher. As a result, the impact of negative features in these situations is likely to increase. The effect of our decision-frame manipulation is consistent with this interpretation. Generally, therefore, the difference in impact of positive relative to negative information depends on factors that makes these types of information more available in memory.

In some respects, our results are similar to the findings in the literature on memory-based versus stimulus-based judgments (see the review in Dick, Chakravarti, & Biehal, 1990; Hastie & Park, 1986). The accessibility effect, as indicated by the asymmetry of positive and negative attributes, was maximal when the standards had to be retrieved from memory (similar to memory-based judgments) and minimal when the standards were visually available (similar to stimulus-based judgments). However, note that all our exemplar judgments were stimulus-based according to the usual definition. That is, subjects had the exemplar information available to them continuously during the time they made their judgment. Thus, our findings generalize the memory-based versus stimulus-based distinction by suggesting that judgment models must include consideration of the availability of all information relevant to the judgment, including information about the standards, and cannot be limited to the stimulus information.

These considerations make it difficult to predict the relative impact of positive and negative features in memory-based evaluations. As an example, consider what happens when a consumer needs to evaluate a service company. He or she retrieves episodes associated with service from memory and compare these episodes to the appropriate standards of service. In this case, the impact of positive and negative episodes may depend on their ease of retrieval. Negative information has more impact on evaluations of service when it is easier to retrieve the negative episodes (Schul & Schiff, 1993).

Our experiments employed unfamiliar products in order to increase the homogeneity of the knowledge base. Can we generalize our findings to evaluations of familiar products? One important difference between familiar and unfamiliar products concerns the involvement of subjects with the decision. Decisions concerning familiar products are likely to be associated with a higher degree of involvement. As a consequence, the decision criteria for familiar and unfamiliar products may differ. Ganzach (1993) found that the impact of negative information (as compared to positive information) increases under high involvement. This may occur because high involvement makes the negative standard more salient and hence increases the impact of information compatible with it. We are currently exploring this question.

ACKNOWLEDGMENTS

Preparation of this article was supported, in part, by a grant from the United States–Israel Science Foundation (BSF) to Yaacov Schul and by funds from the K-Mart Center for International Retailing and Marketing to Yoav Ganzach.

We thank Asher Cohen and David Mazursky for commenting on earlier versions of this article.

REFERENCES

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research, 13,* 411–454.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62,* 893–912.

Burnstein, E., & Schul, Y. (1982). The informational basis of social judgments: Operations in forming an impression of another person. *Journal of Experimental Social Psychology, 18,* 217–234.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dick, A., Chakravarti, D., & Biehal, G. (1990). Memory-based inferences during consumer choice. *Journal of Consumer Research, 17,* 82–93.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* New York: Harcourt Brace.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgments: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Ganzach, Y. (1993). Goals as determinants of nonlinear noncompensatory judgment strategies: Leniency verses strictness. *Organizational Behavior and Human Decision Processes, 56,* 422–440.

Ganzach, Y. (in press). Attribute scatter and decision outcome: Judgment versus choice. *Organizational Behavior and Human Decision Processes.*

Ganzach, Y., & Schul, Y. (in press). The influence of quantity of information and goal framing on decision. *Acta Psychologica.*

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93,* 258–268.

Higgins, E. T. (1990). Personality, social psychology, and the person-situation relations: Standards and knowledge activation as a common language. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 301–338). New York: Guilford.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison on preferences. *Journal of Experimental Social Psychology, 25,* 121–141.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Jaccard, J., & Wood, G. (1988). The effects of incomplete information on the formation of attitudes toward behavioral alternatives. *Journal of Personality and Social Psychology, 54,* 580–591.

Lehman, D. R., Krosnick, J. A., West, R. L., & Li, F. (1992). The focus of judgment effect: A question wording effect due to hypothesis confirmation bias. *Personality and Social Psychology Bulletin, 18,* 690–699.

Loken, B. L., & Ward, J. (1990). Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research, 17,* 111–126.

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle.* Cambridge, MA: Schenkman.

Meyer, R. J. (1987). The learning of multi-attribute judgment policies. *Journal of Consumer Research, 14,* 155–173.

Schul, Y., & Schiff, M. (1993). Measuring satisfaction with organizations: Predictions from information accessibility. *Public Opinion Quarterly, 57,* 536–551.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition, 21,* 546–556.

Simmons, C. J., & Lynch J. G. (1991). Inference effects without inference making? Effects of missing information on discounting and use of presented information. *Journal of Consumer Research, 17,* 477–491.

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Review, 105,* 131–142.

Smith, E. R., & Branscomb, N. R. (1987). Procedurally mediated social inferences: The case of category accessibility. *Journal of Experimental Social Psychology, 23,* 361–382.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3–21.

Sujan, M. (1985). Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of Consumer Research, 12,* 31–46.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103,* 299–314.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273–281.

Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Accepted by Eric Johnson.

APPENDIX

Stimulus Construction

Experiments 1 and 2. The 14 features were divided into four mutually exclusive sets. The GB set included two good features and two bad features, the GN set included two good features and two neutral features, the BN set included two bad features and two neutral features, and the N set included only two neutral features.

Features in each of the four sets belonged to two dimensions (the arrangement of features within the dimensions is also presented in Table 1). For the GB, GN, and BN sets, each of the dimensions was associated with features that had different valences. Thus, for example, if the features of the GB set were G1, G2, B1, and B2 (where G1 and G2 designate good features and B1 and B2 designate bad features), then the features G1 and B1 were associated with one dimension (e.g., cooling method), whereas G2 and B2 were associated with a second dimension (e.g., process pressure). Similarly, if the features of the GN set were G3, G4, N1, and N2 (where G3 and G4 designate good features and N1 and N2 designate neutral features), then features G3 and N1 were associated with a third dimension, whereas features G4 and

N2 were associated with a fourth dimension. The four features in the BN set belonged, similarly, to the fifth and six dimensions. The N set differed from the first three sets in that each of the two dimensions was associated with only one feature.

This organization of features into four sets allowed us to build 64 exemplar alloys. Each exemplar was defined by a value on each of the four sets (GB, GN, BN, and N). Thus, an exemplar had either a good or a bad feature from the GB set, either a good or a neutral feature from the GN set, either a bad or a neutral feature from the BN set, and either a neutral feature from the N set (which resulted in a four-feature description) or no feature from the N set (which resulted in a three-feature description). The 64 exemplars could thus be viewed as four replications of a 2 (GB: Good vs. Bad) × 2 (GN: Good vs. Neutral) × 2 (BN: Bad vs. Neutral) × 2 (N: Neutral vs. None) design.

Experiment 3. The 52 experimental exemplars consisted of two replications of a 3 (GB: Good, Bad, or Missing) × 3 (GN: Good, Neutral, or Missing) × 3 (BN: Bad, Neutral, or Missing) within-subject design. For obvious reasons we did not include an exemplar in which all three dimensions had a missing feature. Consequently, each of the two replications consisted of 26 exemplars.