The Encoding and Utilization of Magnitudes of Product Attributes: An Investigation Using Numerical and Verbal Information

Madhubalan Viswanathan
Department of Business Administration
University of Illinois

Terry L. Childers
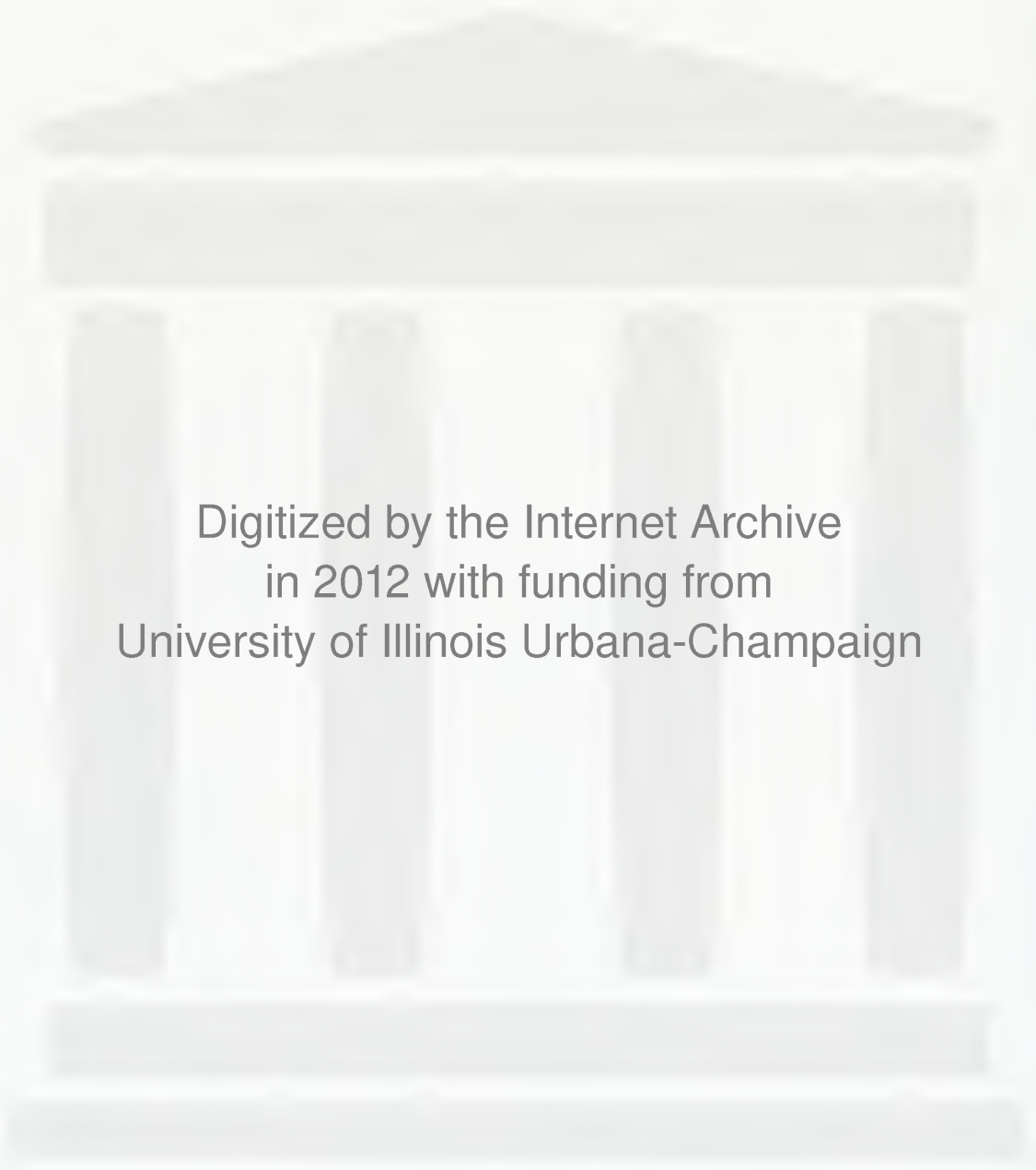*Department of Marketing*
Carlson School of Management

BEBR

The Encoding and Utilization of Magnitudes of Product Attributes:
An Investigation Using Numerical and Verbal Information

Madhubalan Viswanathan

Terry L. Childers

The Encoding and Utilization of Magnitudes of Product Attributes:

An Investigation Using Numerical and Verbal Information

ABSTRACT

This study focuses on the encoding and use of magnitudes of products on specific attributes. Using the attribute, gas mileage, as an example, magnitudes refer to quantifiers such as '32' m.p.g. or 'high' mileage. A conceptualization is proposed which suggests the encoding of magnitudes in two distinct forms; in a relatively coarse-grained (i.e., imprecise) and evaluative form in some situations to conserve processing capacity and facilitate decision-making, and as presented without modification in some situations. The conceptualization is tested and supported across three experiments using hypotheses about the encoding and use of numerical and verbal magnitudes under learning versus choice tasks.

INTRODUCTION

A consumer looking for a new car hears an advertisement on the radio for a car that notes its base price of $14,750. At a dealership, s/he finds that the price for the model on the floor is $16,875. S/he enquires about the difference with the radio ad price and is told that this model has several options including an economy package that boosts the car's gas mileage to an expected 32 miles per gallon. The consumer recalls reading about the first car (i.e., the car mentioned in the radio ad) in Consumer Reports. While s/he remembers that the car was 'high' in gas mileage, she does not recall the precise mileage. S/he could not make a comparison but noting from past experience that most cars don't get much better than 25 m.p.g., s/he judges the second car as high on mileage and thus comparable to the first car. Information such as '32' m.p.g., 'high' mileage, and $'14,750', are referred to in this paper as magnitude information. A magnitude is defined for the purposes of this study as the quantification of a product characteristic or attribute on some continuum. For example, price is an attribute of a car that can vary along a dollar continuum, with a specific car's price of $14,250 being the value or magnitude that quantifies this product characteristic. Examples of magnitudes include information presented on packages using numerical quantifiers (e.g., 19 grams of fat) and verbal quantifiers (e.g., 'low' fat content). Such information represents the magnitude on an attribute (such as fat content) that is possessed by a product.

Marketing information about products is often conveyed by providing information about their magnitudes on specific attributes. Such information is the basic input to consumer decision making that is utilized to make higher-level decisions. Past consumer research has examined ways in which consumers combine attribute information to make a brand decision and ways in which consumer memory is organized around brands and attributes. However, there is little knowledge on how consumers encode magnitude information on product attributes nor how they use such information in decision making (i.e., at a lower level than past research which has typically been at the level of a brand in relation to its attributes). An understanding of the encoding and use of magnitudes could provide a basis for explaining several important phenomena in consumer research such as memory-based judgments that involve the utilization of magnitude information.

The purpose of this study is to understand how consumers encode and use magnitude information by studying two modes of conveying such information, namely numerical and verbal

magnitudes. These two modes were chosen since they differ on certain properties which will be argued to be of importance in understanding the encoding and use of magnitudes expressed in any mode. In the above example, the consumer was able to retrieve from memory the precise price of the first car, but was only able to remember its approximate gas mileage. Magnitude information is sometimes available in a fine-grained (precise) form in memory, but is more often only available in a more coarse-grained (i.e., relatively imprecise) form[1]. However, although coarse-grained information such as the verbal magnitude, 'high', lacks the fine-grainedness or precision of numerical information, the verbal magnitude may convey additional information of an evaluative nature that may facilitate its use in decision making. Our automobile consumer, although unable to recall the exact mileage, was able to infer that the first car's mileage was 'high' and thus, perhaps this information carries a positive connotation that is lacking in its numerical counterpart. These instances are exemplary of the circumstances that characterize consumer purchase behavior. Magnitude information regarding product attributes is often presented in different modes which vary in terms of their *fine-grainedness*, and *evaluativeness* and may also be encoded in ways which differ on these two properties. The objective of the present study is to study the encoding, and use of magnitudes by examining numerical and verbal magnitudes.[2] Towards that end, literature from relevant disciplines is reviewed and used to derive a conceptual framework. Using this conceptual framework, hypotheses are generated and tested across three experiments.

REVIEW OF RELEVANT RESEARCH

Relevant research from several disciplines is presented and is organized as follows. First, is a discussion of research which relates to the precision or fine-grainedness with which magnitudes may be encoded in memory and used in decision making. Second, past research is examined in order to address issues relating to the evaluativeness with which magnitudes can be encoded by consumers and used in decision making. Hence, the review will bring out the significance of considering magnitudes of product attributes in terms of (i) fine-grainedness, and (ii) evaluativeness.

Fine-grainedness of Magnitudes

Research in cognitive psychology and related disciplines has studied how people make

comparative judgments based on magnitude information. Comparative judgment tasks require subjects to compare stimuli along some dimension and make a judgment about the magnitudes of stimuli along that dimension. A robust finding in this area of research, referred to as the symbolic distance effect (cf. Banks 1977), is that judgments are easier (as evidenced by lower response times and/or higher accuracies) when the magnitudes of a pair of stimuli on some dimension (e.g., sets of numbers) are farther apart (2 vs. 59) than when they are closer together (7 vs. 9) (Moyer and Landauer 1967).

Several models developed to explain comparative judgments and the symbolic distance effect differentiate between the fine-grained versus coarse-grained nature of magnitudes used in making comparative judgments. Some researchers (cf. Moyer and Landauer 1967) suggest that size comparisons of stimuli are made using analog representations of magnitudes. Hence, when comparing numbers, representations of their magnitudes are argued to have as many discriminable positions as there are numbers to be compared (i.e., if numbers from 1 to 100 are compared, then representations of a 100 discriminable positions are used). Others (cf. Banks 1977) suggest that comparisons are made using coarse-grained magnitudes (such as 'large' and 'small'). In contrast, Holyoak and Mah (1982) suggest that *both* coarse-grained and fine-grained components are used in making comparative judgments. They argue that magnitudes in long-term memory are of a fine-grained form. However, coarse-grained magnitudes in working memory are argued to be used when making comparative judgments with fine-grained magnitudes being retrieved iteratively from long term memory if the need for a finer discrimination arises. In summarizing this area of research, Holyoak and Walker (1976) assert that these various models of comparative judgments assume the *availability* of fine-grained representations of magnitudes in memory, but differ in terms of the nature of magnitudes *used* in the process of making comparative judgments. This stream of research points to fine-grainedness as being an important property of magnitude information.

Related to this research in terms of the fine-grainedness versus coarse-grainedness of representations of magnitudes, is a study in consumer research (Johnson and Fornell 1987) which suggests that the use of features (which involve representation of magnitudes at only two levels, namely the presence and the absence of a feature) should increase with an increase in the concreteness of product attributes. Johnson and Fornell reason that, since features have only two

levels, less processing capacity may be required to represent features than to represent dimensions in memory. Thus, processing capacity can be conserved for large numbers of concrete attributes by using featural representations. Such a rationale based on an increase in processing capacity with an increase in the number of levels used in representations may be extended to magnitudes of dimensions (which are attributes that can taken on a range of magnitudes on a continuum in contrast to features). It is possible that some dimensions may be represented with as few as two levels while other dimensions may be represented with, say, ten levels. The rationale based on processing capacity (Johnson and Fornell 1987) would suggest that magnitudes that are represented in a more coarse-grained form (i.e., with relatively fewer number of levels similar to the codes such as 'large' and 'small' argued to be used in comparative judgments by some researchers (cf. Banks 1977)) would require less processing capacity than magnitudes that are represented in a more fine-grained form (i.e., with a relatively larger number of levels similar to the fine-grained magnitudes argued to be used in comparative judgments by other researchers (Moyer and Landauer 1967)). Therefore, it appears that the fine-grainedness of representations of magnitudes (i.e., the use of many versus few levels of magnitudes to describe a continuum) has implications for processing capacity .

While research reviewed to this point relates to the use of fine-grained versus coarse-grained representations in memory, some research suggests the use of fine-grained versus coarse-grained representations *as a function of task instructions at encoding*. Hinrichs and Novick (1982) studied memory for numbers and concluded that, depending on the emphasis on precise recall in task instructions, numbers may be represented precisely or approximately. They argue that when exact recall is not emphasized, numbers may be encoded in terms of their approximate magnitudes since approximate information is often sufficient in dealing with magnitude information. Further, Johnson and Tversky (1984) have shown that consumers may use different types of representations (such as featural versus dimensional representations) depending on the *task* they perform. Therefore, past research points to the importance of the encoding task in influencing the fine-grainedness of representations of magnitudes in memory.

Evaluativeness of Magnitudes

Magnitudes expressed through different modes can also vary in the degree of evaluativeness with which they convey information. To understand how magnitudes differ in the degree to which

they are evaluative as well as the implications of such differences, it is useful to examine differences between numerical versus verbal magnitudes. Verbal magnitudes may be relatively more evaluative than numerical magnitudes in that, in addition to conveying magnitudes, they make the inference or evaluation of highness or lowness along an attribute (for example, a key difference between '32' miles per gallon and 'high' mileage is that the verbal magnitude makes the inference of *highness* in gas mileage). Even if some reference point (such as '25' miles per gallon) is available for interpreting a numerical magnitude, an additional comparison is required in order to derive its evaluative meaning. Further, due to their evaluative nature, verbal magnitudes may be easier to use in making evaluations during decision making. Past research provides support for these two claims.

Osgood et al. (1957) suggested that evaluativeness forms an important part of the meaning of a verbal expression. Holbrook (1978), in studying the factualness versus evaluativeness of messages, used larger proportions of numerical versus verbal information to operationalize factual versus evaluative messages, respectively. Other researchers have recognized this difference in evaluativeness between verbal and numerical information and examined its implications. Huber (1980) studied the effect of numerical versus verbal attribute information on aspects of decision-making (with numerical labels opertionalized as anchors on a rating scale). The author hypothesized and found that subjects made evaluations more frequently when using verbal information (since such information was similar to evaluative labels) whereas computations of differences and maximum values were performed more frequently on numerical information. Scammon (1977), while studying information load, also found that adjectival rather than percentage descriptors of nutritional information led to more accurate identification of nutritious brands. Consistent with Scammon's study, research on nutritional information disclosure (Venkatesan et al. 1986) suggests that a number derives its meaning in comparison with other numerical information (e.g, an internal reference point). Therefore, several researchers have suggested that verbal magnitudes are relatively more evaluative than numerical magnitudes. Further, past research suggests that verbal magnitudes may be easier to use in making evaluations during the course of decision making than numerical magnitudes due to their evaluativeness.

Synthesis of Past Literature

Drawing this research together, it appears that fine-grainedness and evaluativeness are two

important properties of magnitudes that may have implications for their encoding and use. Specifically, coarse-grained when compared to fine-grained representations of magnitudes may lead to the conservation of processing capacity while evaluative when compared to non-evaluative magnitudes may lead to the facilitation of evaluations during decision making. Additionally, based on past research (Johnson and Tversky 1984; Hinrichs and Novick 1982), it appears that the consequence of the *encoding task* on consumer representations in memory needs to be considered to understand the processing of magnitude information. Overall, there is a need for a better understanding of the fundamental nature of encoding and use of magnitudes in consumer memory. Critical to this understanding is an explanatory framework that allows for the performance of a variety of tasks at different levels of precision and degrees of evaluativeness.

CONCEPTUAL FRAMEWORK

The review of past research brought out two important properties of magnitudes in terms of the degrees of fine-grainedness and evaluativeness. The proposed conceptualization will suggest the encoding of magnitudes in different forms which vary along these two properties. It will be argued that magnitudes may be encoded in a relatively coarse-grained and evaluative form in some situations to conserve processing capacity, and facilitate their use in decision-making. Further, it will be argued that magnitudes may also be encoded without modification in certain situations (i.e., in the mode in which they are presented irrespective of their degree of fine-grainedness and evaluativeness). The encoding of magnitudes in one form versus the other will be related to the processing goals that consumers have at exposure to these magnitudes in terms of an intent to make a decision. The properties of magnitudes in terms of the degree of fine-grainedness and evaluativeness will then be related to differences between numerically and verbally presented magnitudes. Hence, the conceptualization will be tested by deriving hypotheses about the processing of numerical and verbal information under different processing goals.

Encoding of Coarse-grained, Evaluative Magnitudes

This section develops the rationale for suggesting the encoding of magnitudes which are (1) *coarse-grained*, and (2) *evaluative* in nature when such magnitudes are presented during a task involving decision-making. The rationale underlying the encoding of magnitudes in a coarse-grained

form is in terms of the conservation of processing capacity during encoding and when utilizing magnitude information. The encoding of information in a relatively fine-grained form requires a high degree of processing capacity (since such a representation consists of a larger number of levels). Johnson and Fornell (1987) make a similar argument about processing capacity for featural versus dimensional representations. Further, using magnitudes in a relatively fine-grained form in decision making may require a greater degree of effort both in searching for and in performing operations on magnitudes.

This line of reasoning can be illustrated with a perceptual analogy. Consider the use of a 6 inch scale (sensitive to 1 inch) and a 15 cm scale (sensitive to 1 cm). In terms of capacity, more levels are stored on the 15 cm scale. In terms of usage, it is more difficult to search for any given point on the 15 cm scale (since there are 15 scale points). Further, the computation of differences between any pair of scale points may require more effort for the 15 cm scale when compared to the 6 inch scale since finer discriminations will have to be made with the fine-grained scale. The premise here is that finer discriminations require greater effort, a result that has been demonstrated in the comparative judgment literature in terms of the symbolic distance effect (cf. Banks 1977). Therefore, relatively coarse-grained magnitudes may require less processing capacity to encode and to use in decision making than relatively fine-grained magnitudes.

Further, magnitudes are argued to be encoded in an evaluative form to facilitate their use in decision making. A coarse-grained representation with five levels could either be similar to a 1 to 5 point scale or be paramorphic to a five point verbally anchored scale (say, very low, low, medium, high, and very high). While both scales are relatively coarse-grained, the verbal anchors are inherently more evaluative in nature. Researchers have shown that verbal information (due to its evaluative nature) is easier to use in making evaluations while numerical information cannot be understood in isolation and requires a comparison process to derive meaning (cf. Huber 1980). Based on past research, it appears that to the extent that a magnitude in memory is more evaluative in nature, its use in making evaluations during the course of decision making is facilitated. Hence, magnitudes that are relatively more evaluative (similar to verbal labels) are argued to facilitate decision making.

Additionally, the encoding of magnitudes in a similar (i.e, coarse-grained and evaluative)

form makes it possible for magnitudes to be comparable to each other, a property which also facilitates decision making. The availability of magnitudes in a comparable form would facilitate the performance of operations on magnitudes (such as combinations and comparisons) that are required for tasks related to decision making such as judgment and choice. For example, comparisons of brands along an attribute, which have been argued to occur during a choice task (cf. Biehal and Chakravarti 1982), would be facilitated by the availability of magnitudes in a comparable form. Comparable magnitudes would also facilitate the use of brand-based choice strategies by consumers. Further, comparable magnitudes also facilitate the formation of a judgment about a brand since magnitudes along various attributes are in a similar form. Hence, in addition to advantages in terms of conservation of processing capacity, and facilitation of decision making described above, the encoding of coarse-grained, evaluative magnitudes leads to availability of magnitudes in a comparable form which further facilitates the performance of operations during the course of decision making.

Encoding of Magnitudes without Modification

The line of reasoning developed so far is based on the conservation of processing capacity, and facilitation of decision making. Encoding of relatively coarse-grained magnitudes was argued on the basis of the conservation of processing capacity. However, there is a trade-off involved between conserving processing capacity and losing information by categorizing two slightly different magnitudes into a single category. Therefore, it is suggested here that, in a decision making context, consumers *may* also encode magnitudes as presented without modification. Therefore, relatively fine-grained magnitudes may be available for use in decision making as well. The probability of such encoding would be a function of the anticipated use of magnitudes in future decisions which, in turn, may depend on factors such as the perceived importance of a decision or the perceived importance of an attribute in a decision. As an example, consumers may encode the fine-grained (i.e., precise) gas mileage of an automobile (e.g., '32' miles per gallon), but may not encode its fine-grained length or weight. Such magnitudes could then be used during later stages of decision making in a phased or a two-step choice process in order to make a finer discrimination between alternatives when coarse-grained magnitudes are not sufficient. The encoding of magnitudes in the these two forms leads to both the conservation of processing capacity due to the encoding of coarse-grained magnitudes, and

the availability of more fine-grained magnitudes if required.

Further, when consumers are exposed to magnitudes along attributes *without* specific goals relating to decision making (e.g., a learning task), it may not be necessary to encode magnitudes in a coarse-grained and evaluative form. Since the goal when magnitudes are presented is not related to decision-making, effort may not be expended to translate presented magnitudes to a coarse-grained, evaluative form. Consumers may instead encode magnitudes as presented (i.e., without modification) since the goal at encoding is not related to decision making. Such magnitudes which are encoded without modification may be relatively fine-grained and/or non-evaluative as well. Unmodified magnitudes may then be available for use in subsequent tasks, such as decisions about products, that may be made at a future point in time.

Processing of Numerical and Verbal Information under Learning versus Choice Goals

The discussion to this point provides a basis for developing hypotheses for the encoding and use of magnitudes of different modes under different processing goals. The rationale for the encoding of coarse-grained, evaluative magnitudes was based on the facilitation of decision making. The encoding of magnitudes without modification was argued to occur when the goal at exposure to magnitudes is not related to decision making. Therefore, the proposed conceptualization can be tested by comparing different consumer goals such as learning versus choice. To the extent that consumers are exposed to magnitudes in order to make a choice, it is argued that such magnitudes will be encoded in a coarse-grained, evaluative form. Since magnitudes are encoded in a relatively coarse-grained and evaluative form, magnitudes that are relatively fine-grained and/or non-evaluative will be *recoded* or *translated* into a relatively coarse-grained and/or evaluative equivalent before being encoded. This recoding process also facilitates the consumer's choice process by increasing comparability within and across attributes. In contrast, it is proposed that when consumers with the goal of learning are exposed to magnitudes, they are likely to encode these magnitudes without modification (i.e., in the mode in which the magnitudes were presented and, therefore, not necessarily comparable to each other).

Numerical versus verbal magnitudes offer a means of developing testable hypotheses about the nature of encoded magnitudes in terms of their degree of evaluativeness and fine-grainedness. As discussed earlier, verbal magnitudes are maintained to be more evaluative than numerical

magnitudes. Numerical and verbal magnitudes also differ in terms of fine-grainedness. Jaffe-Katz et al. (1989) argued that numerical magnitudes are more precise than verbal magnitudes and found support using comparisons of probability expressions. Studies of verbal probability expressions have found a high variation in the magnitude values assigned to verbal expressions as well as a high degree of overlap (cf. Beyth-Marom 1982). Also, high intra-individual variability (as a function of context) and inter-individual variability in the interpretation of verbal expressions have been cited (cf. Pepper 1981). It appears that verbal magnitudes are typically interpreted as ranges along a relatively coarse-grained continuum. (A magnitude such as 'high' in gas mileage could be interpreted along a continuum that is described few categories. For example, a continuum with three levels could be described by the categories, 'high', 'medium', and 'low'.) Numerical magnitudes, however, may be typically interpreted to represent a point along a relatively fine-grained continuum (e.g., a label such as 32 m.p.g. could be represented as a point along a continuum where mileage is expressed to a degree of resolution of 1 m.p.g.). Therefore, verbal magnitudes encompass more magnitude range or variability and this is captured by a more coarse-grained representation as compared to relatively fine-grained numerical magnitudes.

Given the fine-grained and evaluative nature of numerical magnitudes, it is proposed that such magnitudes have to be *recoded* or *translated* before being encoded during a choice task (a label such as 32 m.p.g. would be recoded into a label such as "high"). Implicit to this translation would be a comparison of an individual magnitude for a brand (e.g., 32 m.p.g.) to a norm (e.g., average product class gas mileage of 22 m.p.g.) in order to derive an evaluative implication (e.g., "high" in gas mileage) for a specific brand. In contrast, verbal information, due to its more coarse-grained nature, is less likely to undergo this recoding process[3].

In the research to be reported, on-line choice and directed learning tasks (cf. Biehal and Chakravarti 1983) are employed to test these assertions for the proposed model. On-line choice tasks are characterized by information from the environment entering working memory and immediately initiating a choice process. Under such conditions, the proposed recoding of numerical magnitudes would be expected to take place at *encoding* and the resulting coarse-grained, evaluative magnitudes would enter into the choice process. The coarse-grained, evaluative magnitudes would be encoded, while the original, numerical magnitudes *may* also be encoded. Verbal magnitudes are

less likely to be recoded and, hence, more likely than numerical magnitudes to be directly encoded. Additionally, under a directed learning task, *both* numerical and verbal magnitudes will be encoded without modification.

As noted, a distinction between encoding in a coarse-grained, evaluative form and encoding without modification is the the extent to which magnitudes of brands are comparable to each other. If magnitudes are encoded during a learning task, but subsequently retrieved from memory for use in a choice task, verbal magnitudes of brands encoded without modification will become more comparable to magnitudes encoded in a coarse-grained, evaluative form. On the other hand, numerical magnitudes will first have to be recoded before their translated magnitude will become comparable to other coarse-grained, evaluative magnitudes. Further, it should be noted that if, during a phased choice process, some brands are eliminated on the basis of one or a few attributes, only some magnitudes for these brands may be encoded in a coarse-grained, evaluative form while other magnitudes may be encoded without modification, if encoded at all in memory. The above discussion of numerical and verbal magnitude in terms of their fine-grainedness, evaluativeness, and comparability form the basis for making the predictions discussed in the next section.

Operational Hypotheses for Memory and Decision Making

Using a setting where both numerical and verbal information are provided with the single goal of either learning information or making a choice, operational hypotheses can be generated for a variety of tasks which follow these processing goals. The tasks chosen for study consist of (1) recognition of magnitudes, (2) brand ratings along attributes, (3) recall of magnitudes, (4) overall judgments of brands, and (5) comparative judgments of brands along attributes. These tasks provide tests of encoding of magnitude information in memory (recognition and recall) and retrieval of magnitudes for decision making (ratings, overall judgments, and comparative judgments).

Recognition.

Recognition tests have been used in the context of tests of models of semantic memory, with faster recognition of test stimuli similar to their representation in memory generally hypothesized (Chang 1986). To the extent that on-line choice leads to the recoding of a numerical magnitude into a coarse-grained form, the magnitude will be more difficult to recognize in its original (numerical) form (with evidence of difficulty being provided by higher response times). Therefore, the

recognition test exploits the nature of the storage of information in an isomorphic or modified form.

> H1: Numerical magnitudes will be recognized faster than verbal magnitudes following learning when compared to choice.

The nature of this prediction does not rule out baseline differences in recognition across information mode or task, but only predicts an interaction between task and mode of presentation.

Brand Ratings.

A brand rating task refers to the rating of a brand along an attribute on a verbally anchored categorical scale. Such a task would involve the retrieval of magnitudes from memory to utilize in rating the brands. In line with the proposed conceptualization, it is suggested that it is easier to retrieve magnitudes to utilize for a rating task following a choice goal when compared to a learning goal. Magnitudes of brands on a single attribute are argued to be encoded following choice such that they are evaluative and coarse-grained (i.e., categorical) in nature. The rating task requires an evaluation of a brand on a single attribute along a categorical scale. Therefore, faster ratings based on both numerical and verbal magnitudes will occur following choice when compared to learning.

> H2: Attribute ratings based upon both numerical and verbal magnitudes will be made faster following choice when compared to learning.

Recall.

Recall tasks have been used to assess the nature of storage of information in long term memory (cf. Biehal and Chakravarti 1982). A cued recall task where a brand name and an attribute label are provided with retrieval of information in "any mode in which it comes to mind" is used here. Given the recoding of a numerical magnitude into a coarse-grained (i.e., categorical) form during a choice processing goal, the retrieval of such information in its original form is less likely when compared to a learning goal. The predictions made are, however, not in terms of absolute levels of retrieval but in terms of relative levels of retrieval across task conditions since different degrees of recoding may occur across the two goal-based conditions. This discussion leads to the following hypothesis.

> H3: Recall of categorical equivalents of numerical information will be more accurate than the recall of numerical information following choice when compared to learning.

Recall-Judgment.

The relationship between memory and judgment is important in this context in that it offers a test of the proposition that magnitudes encoded following choice have characteristics that facilitate decision making. Therefore, it is proposed that, following choice, coarse-grained, evaluative magnitudes will be accessed when making judgments (as evidenced by a stronger relationship between memory and judgments). A stronger relationship is predicted between recall of such magnitudes and a subsequent judgment (operationalized as numerical magnitudes recalled in a categorical form) than the relationship between recall of unmodified magnitudes and a subsequent judgment (opertionalized as numerical magnitudes recalled in a numerical form) . However, after learning, where information is encoded in an unmodified form, such a relationship is not expected.

> H4: A stronger relationship between recall and judgment would be obtained for the categorical equivalent of numerical magnitudes than for unrecoded numerical magnitudes following choice when compared to learning.

Comparative judgments.

Additional hypotheses are based on the comparative judgment task (cf. Banks 1977) using the comparison of pairs of brands along an attribute. Following choice, magnitudes are encoded such that they are more easily compared to other brand magnitudes along a particular attribute. In contrast, following a learning processing goal, magnitudes are encoded without modification in the mode in which they are presented. Therefore, they are not easily comparable to other brand magnitudes along the same attributes. Therefore, the accessibility or ease of retrieval of magnitudes required to make a comparison may differ as a function of the form of encoding, with faster retrieval following choice (as argued for brand ratings).

> H5: Faster comparative judgments will be made following choice when compared to learning.

The symbolic distance effect found in previous research for comparative judgment tasks (cf. Banks 1977) provides the final test of the conceptualization. This effect, using brands, predicts that it will be easier to compare a pair of brands along an attribute when the brands are farther apart on that attribute in terms of magnitude than when they are closer. In terms of the encoding of magnitudes following choice versus learning, the former was argued to contain magnitudes in a comparable form (i.e., categorical labels). Therefore, the symbolic distance effect is more likely to be found in judgments made following choice, since the only variation in this form is in terms of

distances between brands. Following learning, information is encoded without modification and thus is represented in different modes (i.e., numerical and verbal). Hence, comparisons between brands would involve magnitudes using these different modes of information. Therefore, in addition to the distance along an attribute, a pair of brands would also vary in terms of the mode in which their magnitudes are encoded in memory, resulting in a lesser likelihood that the symbolic distance effect may be detected.

> H6: The symbolic distance effect is more likely to be found following choice when compared to learning.

EXPERIMENTS

Overview of Design

The hypotheses described above were tested through three experiments. The methodology for the three experiments involved a processing goal manipulation between subjects with two levels (i.e., directed learning and on-line choice (cf. Biehal and Chakravarti 1983)). A within subjects manipulation of information mode was employed, since all hypotheses are at the level of individual attribute magnitudes for a brand. In all three experiments, the initial choice or learning goal was followed by a distracter task involving pictorial information (to avoid use of verbal or numerical information) to remove the effects of short term memory and provide tests of long term memory. In experiment 1, the distracter task was followed first, by a speeded recognition task and then, a rating task. In experiment 2, the distracter task was followed first, by a recall task and then a judgment task. In experiment 3, the distracter task was followed by the comparative judgment task.

Stimulus Materials

Product Category.

A number of criteria needed to be satisfied by a product category in order to be used in these experiments. First, subjects were required to have sufficient knowledge of the attribute magnitudes of the chosen product category in order to be able to recode magnitudes before encoding. Second, product attributes should be manipulatable in numerical and verbal modes. While several product categories such as televisions, calculators, and automobiles partially fulfill the criteria listed above, calculators appeared to be a suitable product category since students are familiar with this category

and own calculators (Biehal and Chakravarti 1983; authors 1989).

Brands.

The number of brands to be used was an issue of importance in order to avoid ceiling or floor effects for memory and the use of idiosyncratic processing strategies. Research in the past has usually involved the use of two to eight brands (e.g., authors 1989; Huber 1980). Four or five alternatives appeared to be appropriate for the creation of a typical consumer decision making context and this was assessed in pilot tests. Fictitious brand names were used to prevent the use of prior knowledge on existing brands which may not be possessed at comparable levels across subjects.

Attributes.

Authors (1989) present a list of six attributes for calculators for which students have similar perceptions, a necessary condition in order to be able to create a realistic profile of attribute information for brands. The six attributes chosen were "warranty length," "battery life," "weight," "number of arithmetic functions," "display width," and "memory". The number of attributes to be used in the experiments was determined by pilot tests and was eventually trimmed to four.

Pretest.

A pretest was conducted to provide a basis for the manipulation of attribute magnitudes in numerical and verbal forms. Since the focus here is on translation from one mode to the other, it was important to generate a set of equivalent verbal and numerical labels for each of several attributes of a calculator which covered the range of possible magnitudes for each of the attributes. Further, it was important to determine the number of magnitude levels to be used for each attribute. The details of the pretest are provided in the Appendix. On the basis of the pretest, five verbal labels and five numerical labels were chosen for each attribute.

Pilot tests.

The pilot tests were designed to test and calibrate the experimental procedure to prevent ceiling or floor effects for memory. Identical stimuli were planned for the set of experiments to provide a means of discussing the results across several dependent variables. The important dependent variable in these pilot tests was recall of information. Higher levels of recognition are acceptable because the recognition test is speeded and affords a more sensitive dependent variable than accuracy, and since mere guessing could produce a recognition accuracy of 50% (similar

arguments apply for the rating task).

The pilot tests were also important in assessing the effectiveness of instructions in several ways. Several issues were emphasized in the instructions and assessed in these pilot tests. These issues concerned the credibility of the information presentation using a context of Consumer Reports (ratings of 5.3 and 4.5 for numerical and verbal labels, respectively, on a 7 point scale of Not at all believable - Very believable); processing of both modes to comparable levels (Mean rating = 4.8 on a 7 point scale of Only Verbal processing- Only Numerical processing where a 4 suggested equal processing of both modes of information); and adherence to task instructions to process all pieces of information (6.4 on a 7 point scale of Not at all - To a large extent). To encourage processing of all pieces of information, magnitude values were assigned to brands such that each of the four brands was first, second, third, and fourth, respectively in its ranking based on magnitude on each of the four attributes (brand judgments on a 7 point scale of Very bad - Very good ranged from 3.2 to 4.3).

The processing goal manipulation followed previous research (Biehal and Chakravarti 1983) where the learning instructions informed the subjects that they would be tested on memory for the information while the choice instructions stressed that subjects were not required to learn, but to use the information in making a choice (protocols after completion of the pilot tests supported adherence to instructions). Importance ratings of attributes were desired to be comparable in order to facilitate processing of all pieces of information (these ratings ranged from 4.4 to 6.9 on a 7 point scale of Not at all important - Extremely Important). Although, a greater variation in importance ratings was observed than anticipated, manipulations of numerical and verbal magnitudes are within attribute and therefore, this should not alter the predictions. Moderately high ratings of knowledge (5.2 on a 7 point scale of Very low - Very high), familiarity (4.2 on a 7 point scale of Very low - Very high), motivation (4.4 on a 7 point scale of Very low - Very high), and effort (5.2 on a 7 point scale of Very low - Very high) to perform the tasks were obtained in the pilot tests.

On the basis of the pretest and pilot tests, the set of brand-attribute information to be used in the experiments was determined[4]. Four brands with four attributes appeared to be appropriate and, hence, the attribute "weight" was excluded based on its poor performance in the pretest. The middle points for the four attributes were excluded since only four brands were to be employed. Magnitudes were assigned to brands and attributes such that (i) the proportion of numerical versus

verbal information was constant across brands and attributes, (ii) both modes were used to convey an equal number of scale-points along a five point continuum (therefore, the valence of information is not confounded with mode of information) and (iii) no magnitude was repeated for any brand to eliminate differential levels of interference for different pieces of information.

Experimental Procedures

The experiments were conducted using Macintosh computers. The sample consisted of 120 undergraduate students at a Midwestern university. 40 subjects were assigned to each experiment with 20 subjects assigned to each task in each experiment. Subjects were provided with a short exercise on the use of the Macintosh computer, familiarized with the product category and attributes on which information would be presented, provided instructions for either directed learning or on-line choice, and familiarized with the brand names. Subjects then engaged in either directed learning or on-line choice processing. They were exposed to one piece of information at a time (i.e., a brand name, an attribute, and a magnitude) and self-paced their exposure to each piece of information. The sequence of information was brand-based (on the basis of the pilot tests) with the order of attributes within each brand counterbalanced across all subjects. Further, the order of brands was randomized to create four different versions of the basic information sequence (to prevent primacy or recency effects for certain brands). Subjects had the option of exiting or viewing the information again only at the end of a cycle of sixteen pieces of information (to prevent differential exposure between pieces of information). This initial phase was followed by a distracter task for one minute where subjects were required to complete a partial line drawing of an object.

Experiment 1.

In experiment 1, the distracter was followed by the recognition task, and then the rating task. The recognition task consisted of 32 trials, the 16 pieces of information originally shown and 16 fillers. These fillers were false information about each of the four brands along each of the attributes (with an equal number of trials in each mode, the use of magnitudes which balance the valence of information in each mode, and no repetition of magnitudes which appeared in the 'true' trials or other fillers). Each trial consisted of exposure to a screen containing a brand name, an attribute label, and a magnitude. Subjects were required to provide a response (i.e., True or False) by clicking the mouse on the Macintosh computer on the appropriate button on the screen. Such a response mode

does not require the use of numbers or letters and should prevent differential interference/facilitation of numerical or verbal magnitudes in memory. The sequence of trials was randomized across all subjects with the constraint that no successive trials were for the same brand or attribute to prevent differential priming of information across trials. Subjects were instructed to provide as fast a response as possible without compromising on accuracy in order to prevent them from performing the task at different points along the speed-accuracy curve, both within and across task conditions. Each trial was followed by a masked screen for 3 seconds to mark the end of the trial and alert subjects to the beginning of the next trial.

The rating task consisted of 16 trials. Subjects were provided with instructions to rate brands for each of the four attributes. The sequence of trials was attribute-based allowing for the provision of instructions specific to verbal anchors for each attribute followed by four trials involving that particular attribute (with 3 second masks between trials). Subjects were required to use the mouse on a Macintosh computer to "click" on a chosen label on a five point verbally anchored scale (with similar instructions on speed-accuracy as in the recognition task).

Experiment 2.

In experiment 2, the recall task followed the distracter task and consisted of 16 trials (with 3 second masks between trials). Subjects were provided with a brand name and an attribute label and instructed to type in the attribute value they could recall. The sequence of cues was brand based (on the basis of pilot tests) with the order of attributes for each brand being randomized. The recall task was followed by a judgment task where subjects were instructed to judge one brand at a time. They were required to respond to four 5 point scales for each brand relating to the "goodness" (very bad - very good), liking (not at all - very much), quality (very low - very high), and likelihood of purchase (very low - very high) of the brand.

Experiment 3.

In experiment 3, the comparative judgment task followed the choice or learning phase and the distracter task.[5] The comparative judgment task involved pairwise comparisons of brands along an attribute and required subjects to choose the larger (or smaller) of two brands on a particular attribute. By varying the magnitudes of brands along an attribute, ordinal distances (rank orderings of four brands on an attribute) and interval distances (distances between brands along a 5 point scale

developed on the basis of the pretest) between pairs were manipulated. (For example, if brands B1 to B4 have scores of 1, 2, 3, and 5 respectively, then the ordinal distance between B3 and B4 is 1 but the interval distance is 2 and this assignment of scores leads to four levels of interval distance (1 to 4) and three levels of ordinal distance (1 to 3)).

Given four brands, six pair-wise comparisons are possible along each attribute leading to a total of 24 trials for four attributes. Subjects were instructed to choose the larger of two brands for two attributes and the smaller of two brands for the other two attributes. The verbal labels used to instruct subjects were not "larger" and "smaller" but idiosyncratic to the attributes in question (i.e., "lengthier" for warranty length and so on). A pair of brands were placed on the left and right side of the screen, respectively, and subjects clicked the mouse on the left or the right portion of the screen to indicate the larger/smaller of the pair (with similar instructions on speed-accuracy as the recognition task). An attribute-based sequence of trials was used (with 3 second masks between trials) to allow comparisons along one attribute at a time, with instructions specific to the labels used to describe the attribute preceding the trials in order to familiarize subjects with the labels.

RESULTS

Experiment 1

Analysis of Latencies of Recognition.

Data on the recognition task were analyzed by computing the average response times of accurate responses for each subject for numerical and verbal information under each condition. The mean response times were analyzed using an analysis of variance procedure with a 2 (processing goal; between subjects) by 2 (information mode; within subjects) factorial design. A major concern with the use of response latencies is that subjects within and across task conditions may be performing at different levels of speed-accuracy tradeoffs leading to differences in response times. Correlations between response times and accuracy were not significant (directed learning ($r = .13$; p > .05, choice ($r = .14$; p > .05), and for both goals ($r = .18$; p > .05)) suggesting that subjects were performing at comparable levels of speed-accuracy tradeoffs. The overall accuracy of recognition was 79.6%.

The interaction between information mode and processing goal was significant and in the

predicted direction (F(1,38) = 5.99; p < .05), thereby providing support for H1 (see Table 1 and Figure 1). Using learning as a base-line, without the predicted interaction a parallel line would be expected for the choice manipulation. However, there is a deviation such that numerical information takes more time to recognize following choice and therefore involves greater effort.[6]

Insert Figure 1 and Table 1 about here

Analysis of Latencies of Brand Ratings.

Data from the rating task were analyzed by identifying accurate responses for each subject employing two different criteria, strict and lenient scoring. Using the strict criterion, a response was considered accurate if a subject provided the exact scale point that a brand-attribute was associated with (on a five point scale based upon the pretest). However, such a criterion does not allow for individual differences in assessing magnitudes (i.e., a warranty length of 72 months may be "very lengthy" for some individuals but "lengthy" for others). Nor does such a criterion allow for individual differences in the number of levels of magnitudes encoded in memory. It is possible that some individuals may only encode an attribute at three levels in memory, thereby leading to inaccurate recoding according to a strict criterion. Therefore, the accuracy of responses were also identified using a lenient criterion wherein a response was accurate if it was within one scale point on either side of the "strict" response. The analyses were performed using both criteria in order to provide a complete understanding of the data on the basis of similarities or differences in results.

The mean response times for accurate responses (using the lenient criterion) were calculated for each subject for the mode in which information was originally presented. A significant main effect was found for processing goal with slower response times for the learning task (F(1,38) = 9.50; p < .01), providing support for H2. Numerical information was rated faster under choice when compared to learning (F(1,50) = 10.60; p < .01) as was verbal information (F(1,51) = 6.13; p < .05) (see Table 1). The analysis using the strict criterion produced qualitatively similar results.

Experiment 2

Analysis of Recall Accuracy.

Subjects in the recall task were instructed to retrieve information in any form they preferred

leading to four possible combinations; numerical recall of information that was numerical at exposure (referred to as NN where the first letter refers to mode at exposure and the second letter refers to mode at recall), NV, VV, and VN. The number of accurately recalled times for each of these cells was computed for each subject. The accuracy of recall was computed twice using a strict and a lenient criterion, respectively, as was the case with the rating task in experiment 1. A lenient criterion required a recalled item to be within one scale-point on either side of the original item in order to be identified as being accurate (e.g., if battery life for a brand was "long", then recall of this item as "very long" or "neither long nor short" was considered accurate using the lenient criterion). The strict criterion required recall of the exact scale point of the original item. For verbal information, accurate recall was based on the meaning conveyed by a recalled word (for e.g., a label such as 'high' warranty length conveys the same meaning as 'lengthy' warranty length) while for numerical information, recall of the exact digits was required. Using these two criteria, the recall data were examined to identify accurately recalled items and scores were assigned to each subject in terms of the number of accurately recalled items in each condition.

An analysis of variance procedure was applied to the scores (based on the lenient criterion) using a 2 (processing goal) by 2 (mode at exposure) by 2 (mode at recall) factorial design. The contrast involving the difference between NN and NV across goal conditions was significant and in the predicted direction ($F(1,38) = 7.86$; $p < .01$) providing support for H3 (see Table 1 and Figure 2). The levels of recall for the NN condition across processing goals were significantly lower for learning when compared to choice, which is in line with the conceptualization ($F(1,38) = 5.5$; $p < .05$)). An examination of cross modal recall showed that, also in line with the conceptualization, NV was significantly greater than VN following choice when using the lenient criterion ($F(1,38) = 9.95$; $p < .01$). Qualitatively similar results were obtained using the strict criterion.

Insert Figure 2 about here

An alternate explanation for the results could be that subjects in the choice condition recall numerical information in a verbal form due to a guessing strategy as a result of a lack of memory. This explanation is refuted by the similar pattern of results using the strict and lenient criteria since the strict criterion reduces the likelihood that such recall is merely due to guessing. An ANOVA of

the number of inaccurately recalled items using the lenient criterion resulted in a non-significant interaction between processing goal and mode at recall. The premise here was that, if recall of NV after choice was due to guessing, there should be more inaccurate responses of this form (NV) after choice when compared to learning. The error rate was also found to be less for NV after choice when compared to learning (36.6% after choice versus 41.8% after learning).

Analysis of Relationships between Recall and Judgment.

As described in the previous sub-section, the recall task produced four different types of information (i.e., NN, NV, VV, and VN) across the two task conditions. H4 was assessed by examining the relationships between recall of these four types of information for each brand and subsequent judgments made about each brand. The relationship between memory and judgment has been assessed in past research (Lichtenstein and Srull 1985) with correlations between the normatively determined evaluative implications of recalled attribute information and overall ratings of brands. However, in contrast to past research, information under different goal conditions is further divided into four types in this study. The problem in adopting past approaches is the small number of data points for each of the four information types used to compute the correlations. Therefore, a different approach was taken wherein the evaluative implications of recalled information were compared to overall brand ratings. If the evaluative implication and the brand rating were both positive, both neutral, or both negative, this was considered a match between recall and judgment. This approach captures the strength of relationships between memory and judgment by assessing whether each piece of information recalled matched subsequent judgments.

The information recalled by subjects was separately listed for each brand across the four information types. The information was coded on a 1 to 5 point scale determined on the basis of the pretest. The magnitude expressed along the 5 point scale was considered the evaluative implication of the information recalled, since all four attributes possess vector utility properties. Overall brand ratings were also collected on a 5 point scale[7]. A match was defined when the evaluative implication of recalled information was within a point of the subjects' brand rating. Therefore, it was ensured that a match occurred only when the evaluative implication and the rating were both positive (i.e., both scoring 4 or 5), both negative (i.e., 1 or 2), or both close to neutral (i.e., 2, 3, or 4). A match was given a score of +1 while a mismatch was given a score of -1. These scores were summed for

each subject for each information type. Scores were again computed separately for information recalled using both strict and lenient criteria.

An ANOVA was performed on the total scores for each subject using a 2 (processing goal) by 2 (mode of information at exposure) by 2 (mode of information at retrieval) factorial design. Using the data based on the lenient criterion, the difference in matches between NV and NN across goal conditions was significant ($F(1,38) = 14.8$; $p < .001$) and in the predicted direction, providing support for H4 (see Fig. 3 and Table 1). The difference in matches between NV and NN was also significant at choice ($F(1,38) = 14.62$; $p < .001$) and in the predicted direction, suggesting strong support for the hypothesis (H4). Similar results were found using the strict criterion.

Insert Figure 3 about here

In order to compare these results to past findings in memory-judgment research, the matches for information recalled in its original form (i.e., VV + NN) were compared across goal conditions. A significantly stronger relationship was found after learning when compared to choice using the lenient criterion ($F(1,38) = 5.23$; $p < .05$) and a marginally stronger relationship was found using the strict criterion ($F(1,38) = 3.9$; $p < .10$), which is in line with past findings (cf. Lichtenstein and Srull 1985). Marginally faster judgments were made after choice when compared to learning ($t = 1.32$; $p < .10$), which is also in line with past research. These results suggest that the present study used a comparable procedure and that the index based on matches and mismatches is comparable to traditional ways of assessing memory-judgment relationships.

Experiment 3

Analyses of Comparative Judgments

The accuracy of responses and mean response times of correct responses were computed for each subject for each interval distance as well as each ordinal distance. An analysis of variance using a 2 (processing goal) by 4 (interval distance) factorial design was performed on the mean response times. The difference in response time across goal conditions was marginally significant ($F(1,38) = 2.85$; $p < .10$) with faster responses for the choice condition, thereby providing marginal support for H5. Visual inspection of the results suggested a linear, monotonic trend for choice, but not for learning, with smaller response times for larger distances. A linear trend analysis produced a

significant trend for choice (F(1,57) = 11.13; p < .01) and learning (F(1,57) = 5.88; p < .05) thereby providing support for a modified form of H6. Using accuracy of judgments as a dependent variable, linear trends were found to be significant for the choice condition (F(1,57) = 20.05; p < .001) but not for the learning condition, thereby suggesting support for H6. The analyses using ordinal instead of interval distances resulted in similar findings except that the linear trend analysis using response times in the learning condition was not significant, thereby providing support for an unmodified form of H6. The overall pattern of results in terms of the lack of visual evidence of a linear trend following learning and a significant linear trend for only one of the four analyses suggests support for an unmodified form of H6.

Discussion

The results across three experiments using a range of tasks provide different lines of evidence for the proposed conceptualization. The results of the analysis of recognition and recall (experiments 1 and 2) provide evidence for the suggested nature of encoding of magnitudes either without modification or in a coarse-grained, evaluative form. The results of the analysis of overall judgments (experiment 2) provide support for the utilization of coarse-grained, evaluative magnitudes in making judgments following a choice processing goal. Also, results from the analyses of ratings and comparative judgments (experiments 1 and 3) provide support for the encoding of magnitudes on a particular attribute across different brands that are in a coarse-grained, evaluative form, and which facilitates comparisons. The symbolic distance effect (experiment 3) was not detected following the learning condition, which is consistent with the prediction that magnitudes, following a learning task, are encoded without modification in different modes. The proposed recoding of numerical information can be inferred from these results as well.

GENERAL DISCUSSION

This research suggests that consumers encode magnitudes into two distinct forms. Such encoding in dual forms allows for the conservation of processing capacity, and facilitation of decision making, while retaining the ability to encode and utilize more fine-grained information when needed. The results of the three experiments provide interesting insights into the encoding and use of magnitudes in different modes as a function of processing goals. It appears that there are important

qualitative differences between choice and learning oriented goals. Interactions between processing goals and the modes in which magnitudes are presented suggest that processing under choice leads to the encoding of magnitudes which are used in subsequent judgments, the easier accessibility of information required to make a brand rating or a comparison, and the availability of magnitudes in a comparable form. However, magnitudes appear to be encoded without modification following learning. These results also point to differences in the processing of magnitudes presented in numerical versus verbal modes. Numerical magnitudes may have to be translated into a form that is similar to verbal magnitudes before being used in decision making. While such a translation may require a certain level of knowledge and effort, it may result in the encoding of numerical magnitudes in dual forms.

Several intuitive observations about magnitudes along product attributes can be viewed within the framework of the proposed conceptualization. The number of levels of magnitudes used in encoding may vary as a function of factors such as attribute importance. If sugar content in chewing gum is not an important consideration, few levels of magnitudes (e.g., sugar free or sugar added) may be used. If gas mileage for an automobile is considered more important, more levels (e.g., exceptional, pretty good, mediocre, and unacceptable) may be used. The mode in which consumers typically acquire information about different attributes may also influence how they encode information along these attributes. Some attributes, such as price, are presented in the marketing environment predominantly in a fine-grained form thereby leading to a higher probability of dual encoding as well as more levels of coarse-grained encoding. The latter possibility is not in contradiction with encoding in a coarse-grained form, since it should be noted that coarse-grainedness could vary in terms of the number of levels across different attributes and individuals. Encoding may also differ in terms of the strength of linkages between fine-grained, non-evaluative magnitudes and their coarse-grained, evaluative equivalents. Fine-grained magnitudes that have been retrieved for decision making may have stronger linkages with their categorical equivalents as a result of forming stronger paths in memory. Expertise is another factor that may impact storage and experts may have more detailed encoding as well as stronger linkages between the two forms when compared to novices. Alba and Hutchinson (1987) suggest that experts can make more memory-based discriminations along dimensions than novices. Additionally, Park and Lessig (1981) report

directional support for the prediction that moderate and high product familiarity leads to the use of narrower categories (which was argued to suggest more magnitude categories) when compared to low product familiarity.

The proposed conceptualization is also compatible with several observations made in past research. The conceptualization allows for variation in the level of coarse-grainedness of encoding as a function of the nature of stimuli (such as conceptual versus perceptual stimuli (Tversky 1977)) or the type of attribute (such as concrete versus abstract attributes (Johnson and Fornell 1987)). The variation in the nature of the representation used as a function of task (Johnson and Tversky 1984) can be explained on the basis of the availability in memory of coarse-grained as well as fine-grained magnitudes depending upon task requirements. If a magnitude of a more coarse-grained or categorical nature than is readily accessible in its coarse-grained form is required (e.g., featural representations of a continuum), this could be extracted by a subsequent sampling of the available magnitudes. The conceptualization could also be used to generate 'magnitude-level' explanations of several phenomena. For example, in the area of memory-based judgments, explanations have traditionally focussed on the *organization* of global evaluations and attribute information in memory (i.e., whether attribute information is unified with the global evaluation or separated from it in memory (cf. Kardes 1986)). Explanations at the level of the *representation* of magnitude information would be based on the characteristics of encoded magnitudes and their relationship to overall judgments.

Based upon this research, a general model of storage of magnitudes in memory could be proposed which suggests that consumers prioritize and store magnitudes in two distinctly different forms referred to as primary and secondary forms. Such dual representation models of memory have been suggested in several areas of psychology (cf. Carlston (1980) in the impression formation literature) and used in consumer research (cf. Kardes 1986). The primary form may contain relatively coarse-grained, evaluative magnitudes that are well connected with other product information to form the principal store of product information. In the secondary form, magnitudes would be argued to be stored without modification relatively isolated from other product information. The primary form would, therefore, be easier to access for decision making when compared to the secondary form. Such a model would be very different from previous models (cf. Holyoak and Mah

1982) in that magnitudes would be primarily stored in a coarse-grained, evaluative form in long term memory while more fine-grained information may be available in a secondary form.

Limitations

Several limitations of this research need to be clarified. The experiments manipulated processing goals using either learning or choice tasks. Day to day situations where consumers are exposed to information may involve elements of both learning and choice. Such mixed goals could be understood by interpolating the results obtained here, but remain to be tested. Also, the artificiality of the experimental setting in terms of fictitious brand names and information presentation in a structured sequence enhanced control, but may have reduced the ecological validity of the findings. Experimental control was essential given the lack of research in this area, but future efforts should focus on settings more reflective of typical marketing communications.

The basis for assessing correspondence between verbal and numerical labels as well as the number of levels of magnitude to be utilized was a cross-modal magnitude scaling procedure. However, the results of the pretest were used in aggregation rather than at the individual level to create the experimental stimuli. Hence, individual differences may exist in correspondence between verbal and numerical labels. (A label such as a warranty length of 50 months could be considered "lengthy" by some and "extremely lengthy" by others.) Further, the verbal labels used in the experiment may not exactly match the categorical labels used by individuals in terms of either the number of levels used (some individuals may use more or less than 5 levels) or the specific labels used (some individuals may use a slightly different label, such as, "very lengthy" rather than "extremely lengthy"). These possibilities were accounted for at the data analysis stage by using two criteria in separate analyses of recall and rating which led to similar results (with the two criteria providing estimates based on lower and upper bounds of allowance for variation across individuals). The interpretation of results for the recognition task could be affected by the mismatch between categorical labels, since it could be argued that the recognition of a verbal label may also involve comparison with a slightly different magnitude label, though of the same mode. This explanation can be countered if it is assumed that a translation is easier between verbal forms than from numerical to a verbal form, possibly through assimilation by magnitude categories of magnitudes in their vicinity (e.g., assimilation into the category "extremely lengthy" of the magnitude "very lengthy").

Research Implications

Several interesting research questions need to be examined in light of the conceptualization presented here. One line of research should focus on the development of knowledge of attribute magnitudes. Another line of research should focus on the factors that impact the properties of encoded magnitudes. A third line of research should focus on the processing of magnitudes of different modes. While this study employed a high knowledge setting, it raises issues for future research relating to the processing of attribute information as a function of different levels of product knowledge. Consumers with low levels of knowledge about a product category may use very few magnitude categories or may not be able to recode numerical information, but only encode it in an unmodified form in memory. Hence, even in a decision making context, such consumers may make comparisons between numerical magnitudes without being able to interpret the magnitude conveyed in a meaningful manner. This research also raises the possibility that the encoding of magnitudes of different attributes could vary as a function of factors such as the mode in which consumers usually learn information about an attribute. Visual/pictorial means of learning about products may result in more fine-grained encoding of magnitudes than verbal means. Further, the type of attribute could be related to the nature of encoding (a taxonomic classification like subjective versus objective attributes may be related to the ordinal versus interval nature magnitudes used or the number of magnitude categories used by consumers).

In terms of information in different modes, an interesting result is the better recognition of numerical magnitudes when compared to verbal magnitudes following learning. Perhaps numerical information, by representing a unique point along a continuum, is encoded into memory in a more distinctive fashion (due to the low likelihood of overlap with other magnitudes). Research is needed to understand base-line differences between numerical and verbal magnitudes in terms of processing and memory and explanations for these empirical results need to be generated and tested. This research suggests that the precision or fine-grainedness associated with the conveyance of magnitudes is an important dimension of various types of information which could provide a basis for theorizing about the processing of information in different modes. Pictorial and graphical information could be viewed as conveying magnitudes in a relatively fine-grained fashion. Therefore, several generalizations drawn about fine-grained magnitudes may be extended to pictorial

as well as graphical information. Another avenue of research relates to individual differences in "need for precision" of magnitude information, which may influence important moderators, such as processing effort, ability to discriminate, and the use of decision heuristics.

Methodological implications include the use of the comparative judgment task from cognitive psychology to study the encoding of attribute magnitudes. While similarity judgments have been used in marketing at a brand level to assess "distances" between products, comparative judgments at the attribute level could be used to assess "distances" at the attribute level. This task appears to be face-valid since consumers often compare brands along single attributes. The task also allows for the use of a sensitive dependent variable like response time. Cross-modal magnitude estimation appears to be a feasible approach for the calibration of magnitude labels as well as the identification of the number of levels of encoded magnitudes through the identification of how magnitudes cluster.

Conclusions

The issues raised in this study could be viewed in several ways. First, that magnitude information along dimensions is encoded in memory in distinct ways to conserve processing capacity and facilitate its utilization in decision making. Second, that external information of various modes is likely to undergo distinct processes involving different levels of effort in order to be encoded and used in different ways in decision making. Third, that processing goals have a powerful effect on the encoding and use of magnitudes, not just in a quantitative sense of leading to different degrees of memory, but in a qualitative sense by leading to very different types of encoding which have differential levels of accessibility for subsequent use in decision making. Overall, this study brings out the importance of understanding issues at the level of magnitudes of product attributes as an input to the development of knowledge on consumer memory and decision making.

APPENDIX

The starting point for the pretest was a set of numerical and verbal labels which covered the range of magnitudes for each of six attributes of a calculator. The range of numerical labels, and the verbal anchors to be used for each attribute (such as "lengthy" for warranty length) were chosen on the basis of previous work by authors (1989). Thirteen verbal labels were chosen for each attribute by attaching a range of descriptors (such as "extremely") from previous research (Wildt and Mazis 1978) to labels specific to each attribute (such as "lengthy") . (Wildt and Mazis (1978) provide a list of 50 adverb-adjective combinations that were rated by subjects using a 21 point scale labeled at the end points as "the best (worst) thing I can say about a product".) The chosen labels using the example of warranty length were extremely/very/moderately/fairly lengthy, extremely/very/moderately/fairly brief, lengthy, brief, medium, average, and neither brief nor lengthy. Thirteen numerical labels (such as "72 months" for warranty length) were chosen to cover the range of possible magnitudes for each attribute.

It was necessary for the purpose of stimulus design to determine the number of levels of magnitude to use for each attribute and to develop sets of verbal and numerical labels for each attribute which were equivalent to each other in terms of the magnitude conveyed. The use of category scaling is inappropriate here since it poses an arbitrary number of scale points on subjects. Hence, a magnitude scaling procedure was employed (Lodge 1981) which involved a free elicitation of estimates of magnitudes conveyed by various labels for each attribute. Subjects completed the task of estimating magnitudes conveyed by labels using two different forms of response (by producing numbers or drawing lines such that the size of the number or the length of the line was proportional to their impression of a magnitude). The use of two forms of response provides a means of validating the procedure since past research (cf. Lodge 1981) specifies the relationship between these two forms of response (this relationship is referred to as the characteristic ratio while the relationship observed in practice is referred to as the observed ratio).

Subjects in the pretest were first provided with practice trials in line production and number estimation. These trials, referred to as a calibration procedure, served as practice trials and were used to estimate the observed ratio between the two response modes. According to Lodge (1981), if the characteristic ratio is outside the 95% confidence interval of the observed ratio, a correction factor

is required for scale values of labels. Subjects were then required to estimate the magnitudes represented by a set of 13 verbal and 13 numerical labels for an attribute. Subjects performed the task by producing numbers for each label using the number '50' as an average value. This was followed by a repetition of the procedure where subjects drew lines using a line of length '50 mm' as the average value. For example, subjects judged the magnitudes of warranty lengths represented by labels such as 'extremely lengthy' and '72 months' by producing a number (drawing a line) such that its size (length) reflected how much larger or smaller the label was in comparison with the average warranty length (represented by the number '50' and the line of length '50 mm'). Subjects filled out a scale pertaining to the extent to which the anchors used (e.g., 'lengthy' for warranty length) matched the way they thought about a particular attribute (mean rating across six attributes was 4.7 on a 7 point scale of Not at all - To a large extent). The procedure was then repeated for another attribute. Finally, subjects filled out a scale measuring the credibility of the labels. A context of accurate labels used in Consumer Reports was provided to subjects to control for potential differences in credibility across mode (mean ratings of credibility were 4.5 and 3.7 respectively, for numerical and verbal labels on a 7 point scale of Not at all believable - Very believable). Seventeen subjects were assigned to each set of two attributes for a total of fifty-one subjects. Data on sixteen subjects was excluded due to non-adherence to instructions.

One of the six attributes was deleted (i.e., memory storage), since this attribute was not amenable to manipulation over a range of values. The rest of the data was analyzed in accordance with recommended procedures (Lodge 1981) by computing the geometric means for each label for each attribute. This analysis was performed separately for data from numerical estimates and line lengths. The set of geometric means for each attribute for numerical estimates and line lengths were regressed against each other to assess the observed ratio. The 95% confidence interval for the ratio was found to include the characteristic ratio for the two response modes (which is '1') for three attributes (except "weight" and "warranty length") providing support for the appropriate usage of the two response modes. The 95% confidence intervals for the regression coefficients for the attributes and the correlation coefficients between means of the two response modes were as follows: warranty length (1.01-1.07; $r = 0.99$), battery life (0.89-1.03; $r = 0.97$), arithmetic functions (0.98-1.10; $r = 0.98$), weight (0.84-0.96; $r = 0.97$), and display width (0.90-1.04; $r = 0.97$). The interval

excluded the characteristic ratio for warranty length (marginally) and weight. Lodge (1981) suggests the use of a less stringent criterion, such as correlation, for social judgments to check for linear dependence between the two modes. This criterion was satisfied for all attributes. Given the potential confusion of using numbers to estimate numerical labels, this analysis was repeated for data using only the numerical labels as an additional check. Similar ranges for regressions on data from only the numerical labels led to the following confidence intervals and correlation coefficients: warranty length (0.99-1.04; $r = 0.99$), battery life (1-1.1; $r = 0.99$), arithmetic functions (1.13-1.21; $r = 0.99$), weight (0.81-0.95; $r = 0.96$), and display width (0.82-0.92; $r = .98$) suggesting adherence to instructions in estimating numerical labels using numerical estimation.

The geometric means for verbal labels for each attribute were them examined to identify a clustering of verbal labels that may suggest the number of levels of magnitude to use for an attribute. The labels appeared to divide into five categories for all the attributes suggesting the use of five levels of magnitudes. The five categories were represented by the following sets of labels (using the example of warranty length): extremely/very lengthy, fairly/moderately/lengthy, medium/average/neither brief nor lengthy, fairly/moderately/brief, and extremely/very brief. The mean for each category was significantly different from the means for other categories. In addition, the mean magnitudes of labels from any one category had different rank orderings for different attributes, but the ordering for labels from different categories was consistent across attributes. A subject level analysis of rank orderings of labels resulted in only about 3% of all data points not in agreement with the suggested categories of labels (e.g., associating "high" with a magnitude of "200" versus associating "very high" with a lower magnitude of, say, "190"). Using the example of warranty length, the chosen set of verbal labels consisted of the following: "extremely lengthy," "lengthy," "neither brief nor lengthy," "brief," and "extremely brief." The corresponding numerical labels were chosen by plotting numerical labels against their geometric means and identifying points equivalent to the chosen verbal labels.

Lodge, Milton (1981), Magnitude Scaling: Quantitative Measurement of Opinions, Beverly Hills, CA: Sage.

Wildt, Albert R. and Michael B. Mazis (1978), "Determinants of Scale Response: Label Versus Position," Journal of Marketing Research, Vol. 15 (May), 261-267.

REFERENCES

Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," Journal of Consumer Research, 13 (4), 411-454.

Banks, William P. (1977), "Encoding and Processing of Symbolic Information in Comparative Judgments," in The Psychology of Learning and Motivation, ed. G.H. Bower, New York: Academic Press, Vol. 11, 101-159.

Beyth-Marom, Ruth (1982), "How Probable is Probable? A Numerical Translation of Verbal Probability Expressions," Journal of Forecasting, Vol. 1, 257-269.

Biehal, Gabriel and Dipankar Chakravarti (1982), "Information Presentation Format and Task Goals as Determinants of Consumers Memory Retrieval and Choice Processes," Journal of Consumer Research, 8, 431-41.

Biehal, Gabriel and Dipankar Chakravarti (1983), "Information Accessibility as a Moderator of Consumer Choice," Journal of Consumer Research, 10 (June), 1-14.

Carlston, Donal E. (1980), "The Recall and the Use of Traits and Events in Social Inference Processes," Journal of Experimental Social Psychology, 16 (July), 303-328.

Chang, Tien Ming (1986), "Semantic Memory: Facts and Models," Psychological Bulletin, 99 (2), 199-220.

Hinrichs, James V. and Laura R. Novick (1982), "Memory for Numbers: Nominal Vs. Magnitude Information," Memory and Cognition, 10 (5), 479-486.

Holbrook, Morris B. (1978), "Beyond Attitude Structure: Toward the Informational Determinants of Attitude," Journal of Marketing Research, 15 (November), 545-556.

Holyoak, Keith J. and Wesley A. Mah (1982), "Cognitive Reference Points in Judgments of Symbolic Magnitude," Cognitive Psychology, 14, 328-352.

Holyoak, Keith J. and J.H. Walker (1976), "Subjective Magnitude Information in Semantic Orderings," Journal of Verbal Learning and Verbal Behavior, 15, 287-299.

Huber, Oswald (1980), "The Influence of Some Task Variables on Cognitive Operations in an Information-Processing Decision Model," Acta Psychologica, 45, 187-196.

Jaffe-Katz, Amanda, David V. Budescu, and Thomas S. Wallsten (1989), "Timed Magnitude Comparisons of Numerical and Non-numerical Expressions of Uncertainty," Memory and

Cognition, 17(3), 249-264.

Johnson, Eric J. and Amos Tversky (1984), "Representations of Perceptions of Risk," Journal of Experimental Psychology: General, 113 (1), 55-70.

Johnson, Michael D. and Claes Fornell (1987), "The Nature and Methodological Implications of the Cognitive Representation of Products," Journal of Consumer Research, 14 (September), 214-228.

Kardes, Frank R. (1986), "Effects of Initial Product Judgments on Subsequent Memory-Based Judgments," Journal of Consumer Research, 13 (June), 1-11.

Lichtenstein, Meryl and Thomas K. Srull (1985), "Conceptual and Methodological Issues in Examining the Relationship Between Consumer Memory and Judgment," in Psychological Processes and Advertising Effects: Theory, Research, and Application, eds. Linda F. Alwitt and Andrew A. Mitchell, 113-28.

Moyer, Robert S. and T.K. Landauer (1967), "The Time Required for Judgments of Numerical Inequality," Nature, London, 215, 1519-1520.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum (1957), The Measurement of Meaning, Urbana, IL: University of Illinois Press.

Park, C. Whan and V. Parker Lessig (1981), "Familiarity and its Impact on Consumer Decision Biases and Heuristics," Journal of Consumer Research, 8 (September), 223-230.

Pepper, Susan (1981), "Problems in the Quantification of Frequency Expressions," in New Directions for Methodology of Social and Behavioral Science (9): Problems with Language Imprecision, ed. D. Fiske, San Francisco: Jossey Bass.

Scammon, Debra L. (1977), "'Information Load' and Consumers," Journal of Consumer Research, 4 (December), 148-155.

Tversky, Amos (1977), "Features of Similarity," Psychological Review, 84 (4), 327-352.

Venkatesan, M., Wade Lancaster and Kenneth W. Kendall (1986), "An Empirical Study of Alternate Formats for Nutritional Information Disclosure in Advertising," Journal of Public Policy and Marketing, 5, 29-43.

FOOTNOTES

[1] The terms "coarse-grained" and "fine-grained" refer to how finely distinguished the values on a continuum are from other possible values and are used here interchangeably with the terms "imprecise" and "precise", respectively. A scale that is sensitive to 1 cm is more fine-grained than a scale that is sensitive to 1 inch, since a 1 cm interval is a finer increment than a 1 inch interval. Restated in terms of the number of levels of magnitude used to describe a continuum, if relatively few levels are used (such as the use of 'high', 'medium', and 'low' to describe gas mileage among automobiles), the magnitudes are referred to as being coarse-grained or imprecise and vice versa. These terms are used in a relative sense and do not convey any absolute level of 'grainedness'.

[2] The focus of this research is on numerical magnitudes (such as '32' m.p.g.) and verbal magnitudes (such as 'high' mileage) used to convey attribute information and not on all forms of information involving words or numbers. The latter might include telephone numbers or street numbers. Their purpose is to classify rather than convey degrees or amounts. Further, since the focus here is on magnitudes, the type of attributes of relevance are dimensions, not features. Dimensions of products can have varying magnitudes whereas features are dichotomous in nature.

[3] Although fine-grained verbal schemes are conceivable, the typical use and interpretation of verbal magnitudes, when compared to numerical magnitudes, is relatively more coarse-grained. However, recoding may occur when external information presented verbally is more fine-grained than its internal representation. For example, if the attribute, 'gas mileage' is encoded in the coarse-grained form as "high" and "low," external information such as "very low" will be recoded prior to encoding.

[4] The brand names along with the chosen values along attributes warranty length, battery life, number of arithmetic functions, and display width, respectively, were as follows: (i) 'Baron' - Extremely brief, 40 hours, Extremely high, and 12 digits, (ii) 'Colony' - 5 months, Long, 3 functions, and Extremely wide, (iii) 'Profile' - Lengthy, 380 hours, Low, and 3 digits, and (iv) 'Angle' - 72 months, Extremely short, 38 functions, and Narrow.

[5] This experiment differed from the first two experiments in terms of allowing information

search using a matrix presentation. Proportions of numerical versus verbal information were varied across attributes and some brand magnitudes were replaced by middle points of the 5 point continuum of labels in order to assess issues outside the scope of this study. None of these differences are expected to affect the predictions. Analyses presented here refer to data pooled across all attributes which consisted of equal proportions of both types of information.

[6]   Another line of support for the proposed conceptualization is suggested by an analysis of mean response times for accurate responses to fillers. These fillers were manipulated such that their distance from the actual value was either low (i.e., 2 scale points away) or high (4 scale points away). A 2 (distance; low versus high) by 2 (mode) factorial ANOVA was performed on the data for each goal condition. No significant effects of distance were found following learning in line with the conceptualization that magnitudes are not in a comparable form. A significantly lower response time was found for fillers that were at a high distance when compared to a low distance for verbal fillers following choice (F (1,19) = 6.94, $p < .05$; means = 8.41 and 7.37, respectively, for low and high distances) but not for numerical fillers (F (1,19) = .27, $p > .05$). This is also in line with the conceptualization in that the distance effect is obtained following choice since verbal magnitudes are encoded such that they are comparable to other attribute magnitudes. However, this was not the case for numerical information which may be encoded in a relatively isolated form with only its recoded equivalent being encoded in a comparable form.

[7]   Subjects were required to fill out four items (5 point scales) relating to brand rating, liking, quality rating, and intention to purchase. Conceptually, responses to the first item should be analyzed since responses to subsequent items may have been influenced by the previous judgment. Due to the potentially low reliability of a single item scale of brand rating, the following analysis was performed to support its use. Coefficient alphas were computed for the four item measure of brand evaluation for each brand (Mean alpha = .88; Mean item-to-total correlation for brand rating item = .81). The high reliability of the 4 item measure in combination with the high correlation of the brand rating item with the overall measure justifies its use here. It is also important for the analyses that the mean response for the brand rating scale be in the vicinity of the neutral point (i.e., 3 on a 5 point scale) for determining matches (Mean brand rating = 3.1; 95% confidence interval = 2.9 to 3.2).

TABLE 1

RESULTS OF EXPERIMENTAL TASKS

DEPENDENT VARIABLES	ON-LINE CHOICE GOAL		DIRECTED LEARNING GOAL	
	Numerical information at exposure	Verbal information at exposure	Numerical information at exposure	Verbal information at exposure
Recognition speed (in secs.)	7.05	7.61	7.07	8.74
Rating speed (in secs.)	6.33	6.74	8.88	9.51
Recall (numerical at retrieval)	2.90	0.65	4.35	0.45
Recall (verbal at retrieval)	2.60	3.75	1.60	4.00
Recall-judgment (numerical at retrieval)	-0.10	-0.05	0.65	-0.30
Recall-judgment (verbal at retrieval)	1.20	0.45	0.10	0.80
Encoding time in secs. (Expts. 1 & 2)	12.16	13.21	32.66	36.02
Speed of Comparative Judgments (in secs.)	6.28		7.53	

FIGURE 1

RESULTS OF RECOGNITION TASK

Recognition speed (in seconds)

Information Mode

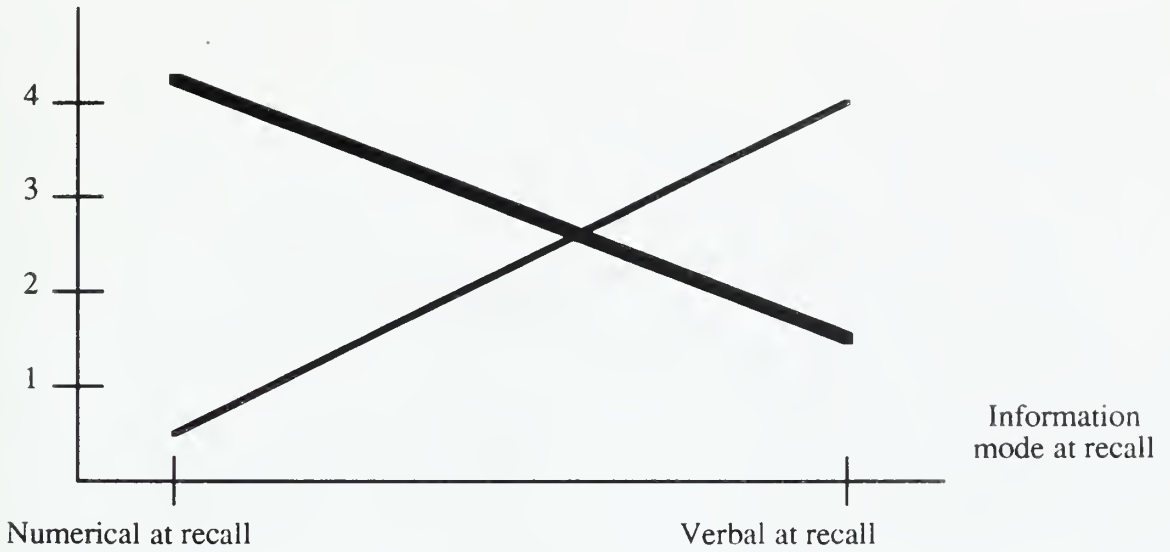Numerical information

Verbal information
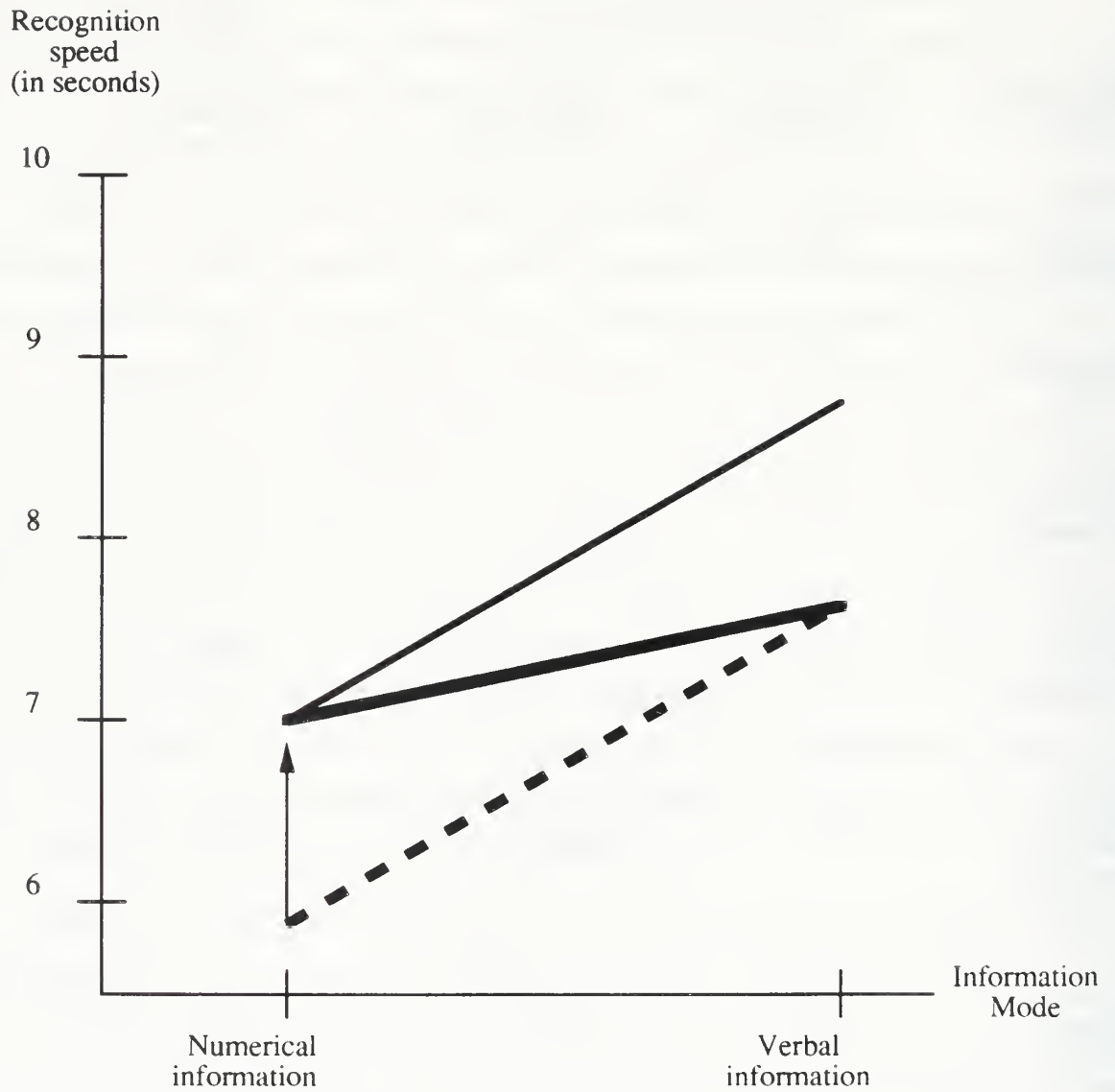
——————— Choice processing goal

——————— Learning processing goal

- - - - - Expected result for choice without an interaction

FIGURE 2

RESULTS OF RECALL TASK

FIGURE 3

RESULTS OF MATCHES BETWEEN RECALL AND JUDGMENT

Matches after learning

Matches after choice

Information mode at recall

Numerical at recall
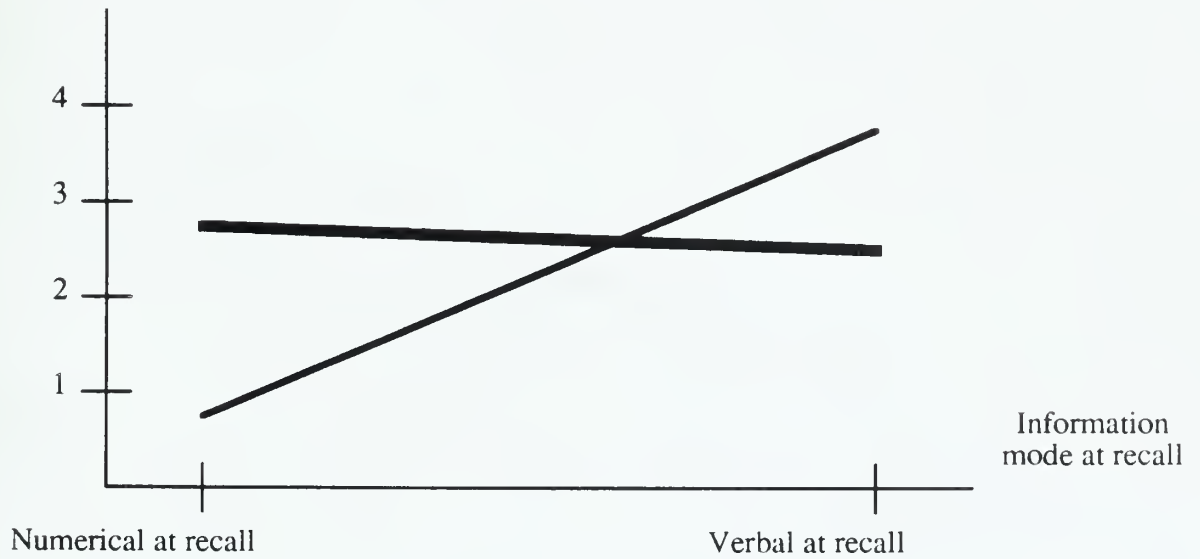
Verbal at recall
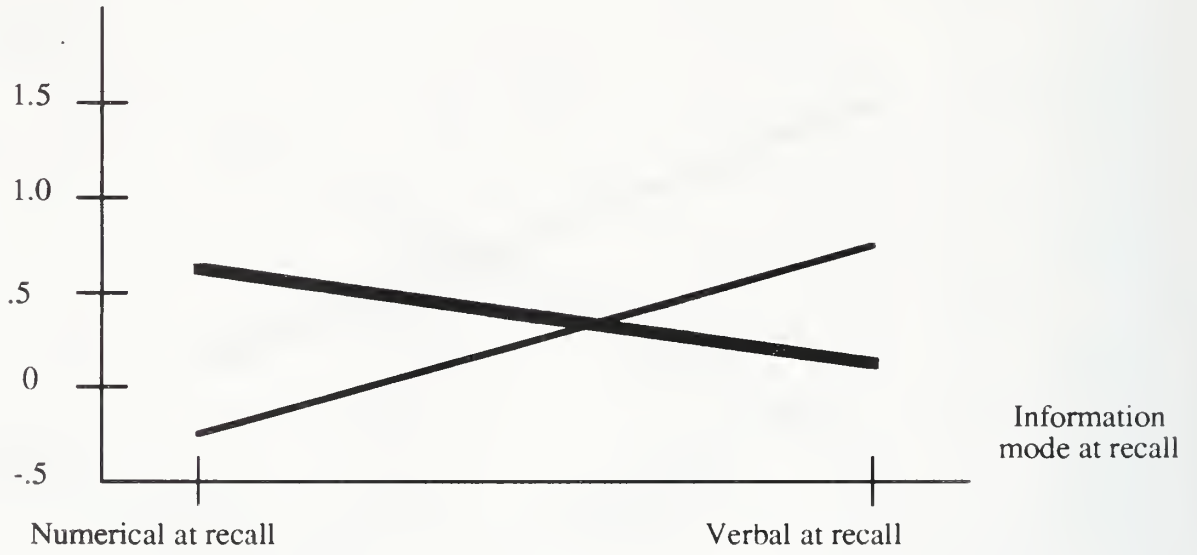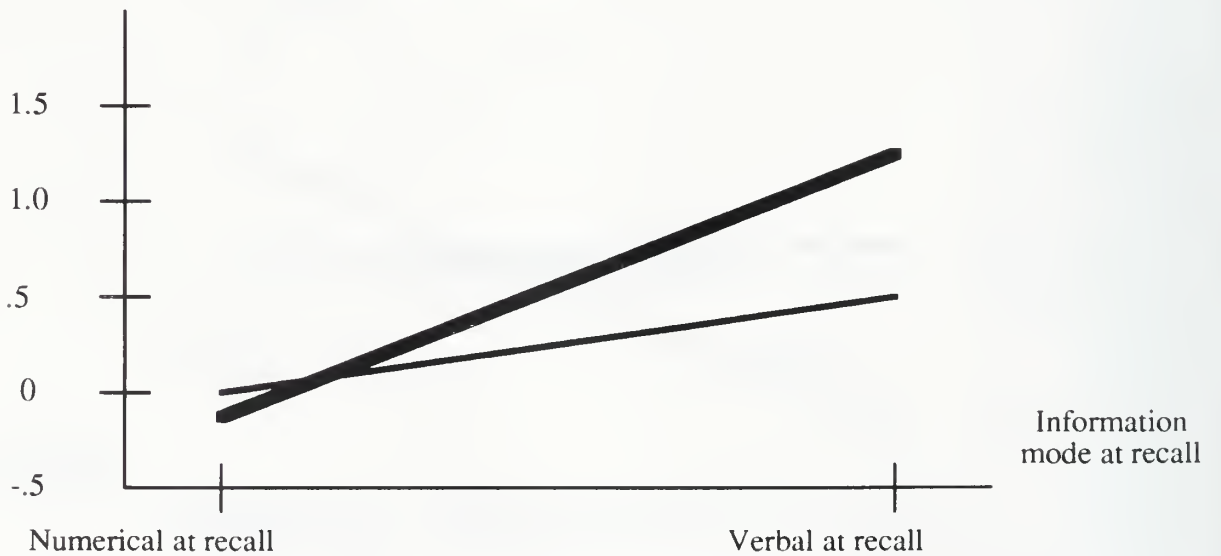
Numerical at exposure

Verbal at exposure