

Bringing Meaning to Numbers: The Impact of Evaluative Categories on Decisions

Ellen Peters and Nathan F. Dieckmann
Decision Research, Eugene, Oregon and University of Oregon

Daniel Västfjäll
Decision Research, Eugene, Oregon and Goteborg University

C. K. Mertz
Decision Research, Eugene, Oregon

Paul Slovic
Decision Research, Eugene, Oregon and University of Oregon

Judith H. Hibbard
University of Oregon

Decision makers are often quite poor at using numeric information in decisions. The results of 4 experiments demonstrate that a manipulation of evaluative meaning (i.e., the extent to which an attribute can be mapped onto a good/bad scale; this manipulation is accomplished through the addition of visual boundary lines and evaluative labels to a graphical format) has a robust influence in health judgments and choices and across diverse adult populations. The manipulation resulted in greater use of numeric quality-of-care information in judgments and less reliance on an irrelevant affective state among the less numerate. Recall results for provided quality-of-care numbers suggested that the manipulation did not influence depth of number processing with the exception of cost information that was not remembered as well. Results of a reaction-time paradigm revealed that feelings were more accessible than thoughts in the presence of the manipulation, suggesting that the effect may be due, at least in part, to an affective mechanism. Numeric information is often provided in decisions, but may not be usable by consumers without assistance from information providers. Implications for consumer decision making and the functions of affect are discussed.

Keywords: decision making, health, information presentation, affect, numeracy

Having access to and understanding information are first steps to making good decisions about health care, financial investments, and other important matters. However, decision makers also must be able to evaluate and use provided information. A major theme emerging from judgment and decision-making research is that decision makers often do not know their own “true” values for provided information (e.g., the importance of the quality of a health plan vs. the importance of its cost). Instead, they construct their values and preferences “on the spot” when asked to form a

particular judgment or to make a specific decision (Lichtenstein & Slovic, 2006). The emerging view is that, in many situations, preference measurement is best considered as architecture (building a set of values) rather than as archaeology (uncovering existing values; Gregory, Lichtenstein, & Slovic, 1993; Payne, Bettman, & Schkade, 1999). Preference construction may be particularly likely when numeric information is involved due to the difficulties individuals face in processing numeric information and the extent of innumeracy that exists (Dehaene, 1997; Paulos, 1988).

In the present paper, we examine judgments and choices in health. In particular, we focus on the use of performance-quality information in decisions about health-insurance plans and hospitals. Such information has received much attention in the public-health sector. The purpose of providing this information is to ensure that plans and hospitals compete on quality rather than only on cost and to give consumers more control over the care they receive. This kind of information is unfamiliar to consumers though, and early efforts to inform them met with little success (e.g., Chernew & Scanlon, 1998). Numeric quality information may be particularly difficult in health contexts because some of the most important indicators (e.g., survival rates) have very small but substantively important differences across options (e.g., 93% vs. 96% survival) at the same time as cost differences (e.g., monthly premiums) are more salient and easier to evaluate. The goals of the present paper are to examine the effects of one information-presentation

Ellen Peters, Decision Research, Eugene, Oregon, and Psychology Department and Department of Planning and Public Policy and Management, University of Oregon; Nathan F. Dieckmann and Paul Slovic, Decision Research, Eugene, Oregon, and Psychology Department, University of Oregon; Daniel Västfjäll, Decision Research, Eugene, Oregon, and Psychology Department, University of Oregon; C. K. Mertz, Decision Research, Eugene, Oregon; Judith H. Hibbard, Department of Planning Public Policy and Management, University of Oregon.

The studies were conducted with financial assistance from the Robert Wood Johnson Foundation (Grant 035643), the National Science Foundation (Grants SBR 9876587, SES 0111941, SES 0112158, SES 0339204, and SES 0517770), and Blue Cross Blue Shield Association (BCBSA). Many thanks to BCBSA and Mark Spranca for assistance in choosing stimuli.

Correspondence concerning this article should be addressed to Ellen Peters, Decision Research, 1201 Oak Street, Suite 200, Eugene, OR 97401. E-mail: empeters@decisionresearch.org

technique—the use of evaluative categories—in both judgment and choice across diverse adult populations and to test plausible mechanisms underlying the effect.

Converging lines of evidence have demonstrated that the weight accorded a stimulus attribute in decisions depends on how easily the value of that attribute can be mapped on to an evaluative (good/bad) scale. For example, research on “evaluability” shows that in separate evaluation (i.e., a single option is shown; Hsee, Blount, Loewenstein, & Bazerman, 1999), attributes that decision makers can evaluate by drawing on prior experience carry greater weight (e.g., a torn cover on a dictionary). In joint evaluation (i.e., multiple options are presented simultaneously), less familiar, often numeric, attributes can be compared and evaluated across options, thus receiving greater weight in decisions (e.g., the number of entries in a dictionary).

Information within a single option can also be made easier to evaluate through a number-comparison process, in a manner similar to that of joint evaluation. Bateman, Dent, Peters, Slovic, and Starmer (2007) observed that the attractiveness of playing a simple gamble (7/36 to win \$9.00; otherwise win nothing) is greatly enhanced by introducing a small loss (7/36 win \$9.00; otherwise lose \$0.05). Subjects responding to the “no-loss” gamble reported lacking precise feelings for how good \$9.00 is (they rated their feelings for the \$9.00 as about neutral), and they gave it little weight in their judgment of the first gamble. In the second gamble, evaluators appeared to compare the \$9.00 with the small loss, making the \$9.00 “come alive with feeling” (their feelings about the \$9.00 were highly positive), and the \$9.00 became weighted in the judgment, thus increasing the gamble’s attractiveness. The number-comparison process appeared to result in a change in affect (good and bad feelings) about the gamble, and this affect was then used as information to guide judgments about the gamble’s attractiveness. Slovic, Finucane, Peters, and MacGregor (2002) argued that affect provides meaning to data, thereby allowing it to be used in decisions and that, without affect, information lacks meaning and is not used (see also Damasio, 1994). Thus, evaluability appears linked with affect (Hsee et al., 1999).

Numeric information in general may be more difficult to evaluate because it is abstract, its meaning changes dramatically from one context to another (9 °F vs. 9 billion dollars vs. 9% wrong on an exam), and numbers are often used to convey small, unfamiliar differences. Numbers are provided in important decisions, however, because they are intended (and assumed) to be useful. In Study 1, we examined whether subjects were able to integrate multiple-numeric quality indicators, including survival rates, in judgments of a hospital without any evaluative aid.

Hypothesis 1: Consumers will not always use provided numeric information. This lack of use will particularly occur when a number is provided without a comparison context, when small, absolute differences exist between options, and/or when the domain is unfamiliar.

Of course, some individuals may be better than others at deriving evaluative meaning from numeric information. Numeracy is assessed through a math test that identifies how well individuals understand numbers and perform number calculations. Highly numerate individuals may also go beyond comprehension and be more likely than those who are less numerate to compare and

contrast numbers in such a way that they draw evaluative meaning from numeric comparisons and, as a result, use numeric information more. This does not necessarily mean they will always make better decisions, however. In the gambling example above, for example, it is only the highly numerate who appear to compare numbers, draw meaning from the comparisons, and show the effect of the five-cent loss. We measure numeracy in Study 1.

Hypothesis 1A: Highly numerate consumers will be able to use numeric information more in decisions compared to the less numerate.

Judgments and decisions are often influenced by nonnumeric sources of information. In Hsee’s (1996) research on evaluability, for example, experienced attributes such as a torn book cover are thought easier to evaluate relative to numeric stimuli such as the number of dictionary entries. In other research, vivid information has been shown to carry more weight than pallid, statistical information (Nisbett & Ross, 1980). When numeric information lacks evaluative meaning, nonnumeric sources of information may receive greater weight as if they are being used as substitute information in a heuristic process (Kahneman, 2003). The less numerate, in particular, may have less access to the evaluative meaning of numeric information and may substitute other sources of information in judgments and choice. In Study 1, we examined the extent to which subjects substituted an irrelevant source of information—naturally occurring mood states—into their judgments.

Hypothesis 2: Other nonnumeric sources of information will be substituted to provide evaluative meaning when numeric information has not otherwise acquired evaluative meaning.

By what means might numeric information acquire meaning? The examples above include number comparisons and transformations (see also Peters, Dieckmann, Dixon, Hibbard, & Mertz, 2007; Peters, Västfjäll, et al., 2006). In patient decisions, presenting statistical information in pictographs (argued as a more experiential, easier to evaluate form of information presentation) limited the influence of anecdotes (Fagerlin, Wang, & Ubel, 2005). Category labels may also assist decision makers by making options easier to evaluate; their power has long been recognized by social psychologists (e.g., Allport, 1954). Zikmund-Fisher, Fagerlin, Keeton, and Ubel (2007) harnessed this power by providing labels such as “normal” or “positive” to prenatal screening test results and found that they influenced risk perceptions and behavioral intentions more than numbers alone. Although information providers are sometimes reluctant to provide labels that categorize and provide an evaluation of the meaning of numeric information (whether due to a desire to allow consumers the freedom to assess the information, a concern about possible litigation, or companies that like to receive positive but not negative scores), it may be that the absence of such labels means that numeric information will not be used.

Hypothesis 3: Helping consumers to evaluate the meaning of provided numbers will increase the use of numeric information in judgments and will influence how numeric information is interpreted and weighted in choices compared to when no help is provided.

Hypothesis 3A: Such help will also result in less use of other information.

In the second part of Study 1, we examined the integration of numeric information and an irrelevant source of information—mood states—in the presence of *evaluative categories*. Evaluative categories are a graphical information display that categorizes options and provides an evaluative label such as “good” or “fair” for each option (see Figure 1 for an example). In Studies 2 and 3, we examined the influence of evaluative categories on health-insurance choices. In Study 2, we also explicitly tested the effect of visual boundary lines that categorize health plans into groups but do not provide the explicit evaluative meaning that may be necessary to influence choice.

In subsequent studies, we tested possible mechanisms for the effects of evaluative categories. Evaluative categories, for example, could simply substitute for provided numeric information, congruent with Hypothesis 2. Decision makers given evaluative categories may skim over detailed numeric information and simply remember whether an option is better on an attribute based on the evaluative categories only. A version of this possibility, tested in Study 3, is that decision makers given evaluative categories will process numeric information more superficially and not remember it as well as those not given evaluative categories (who must compare and contrast numeric information across options to determine which option is better).

Alternatively, evaluative categories could add, not just meaning, but affective meaning to numeric information (Zikmund-Fisher et al., 2007). Affect has been proposed as a key component in the construction of values and preferences (Kahneman, Schkade, &

Sunstein, 1998; Loewenstein, Weber, Hsee, & Welch, 2001; Slovic et al., 2002). Damasio (1994), for example, argued that a lifetime of learning leads decision options and attributes to become “marked” by positive and negative feelings linked directly or indirectly to somatic or bodily states. By consulting and being guided by these feelings, Damasio claimed that we make better and more efficient decisions. Without these feelings (similar to being without evaluability), information in a decision may lack meaning and not be used in the resulting choice.

Slovic and associates (Slovic, Finucane, Peters, & MacGregor, 2004) proposed the affect heuristic in which affect can be learned from experience or constructed from an unfamiliar context and act in a manner similar to that described by Damasio (1994). Slovic et al. argued that in the process of making a judgment or decision, people consult or refer to an “affect pool” containing all the positive and negative tags consciously or unconsciously associated with the representation of the decision problem. An underlying idea, and a justification of the use of the term *heuristic*, is that individuals consult this “affect pool” for information when making a judgment, rather than reviewing all available information. Evaluative categories may influence this affect pool consistent with other links between categorical processes and affective mechanisms (Alexander, Brewer, & Herrmann, 1999). Evaluative categories thus could add meaning and evaluability to numeric information similar to the affective meaning drawn from number comparisons in the earlier gambling example.

Affect from evaluative categories, if present, could then influence information processing and decisions in multiple ways (Peters, 2006). It could act as information when decision makers

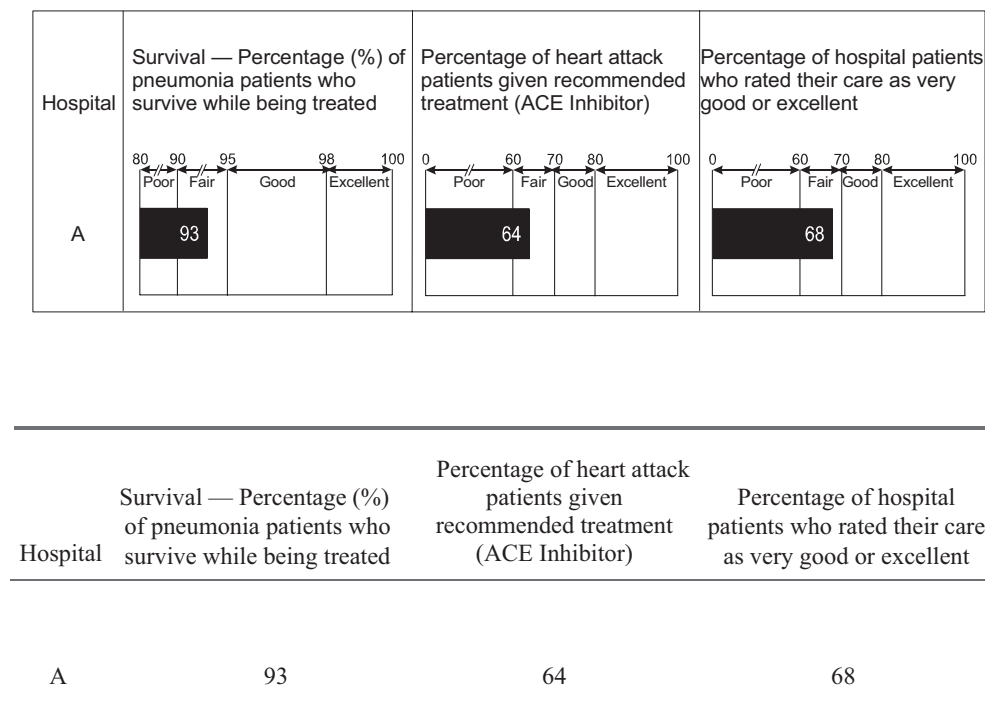


Figure 1. Study 1: Hospital information is provided in an evaluative-categories format (top) or with numbers only (bottom).

are aware of and use that affect to guide judgments as suggested by the affect heuristic (see also Schwarz & Clore, 2003). This function of affect is consistent with the effects of labels in Zikmund-Fisher et al. (2007).

Affect can also serve other functions in decision-making processes, including affect as a common currency. Neural and behavioral evidence suggests that information integration may require complex information to be translated into a common affective currency (Cabanac, 1992; Montague & Berns, 2002; Peters, 2006). Thus, decision makers can compare and integrate good and bad feelings rather than attempting to make sense out of a multitude of conflicting logical reasons. Affect instead may act as a more direct motivator of behaviors. Classical theories of emotion include, as the core of an emotion, a readiness to act and the prompting of plans (Frijda, 1986). Although affect is a much milder experience compared to a full-blown emotion state, recent research has demonstrated that decision makers tend to automatically classify stimuli around us as good or bad in a manner linked to behavioral tendencies (Chen & Bargh, 1999). Affect as a motivator may promote particular choices directly without influencing perceptions of or memory for the underlying attribute information. Finally, affect could act as a spotlight, directing attention toward some information and away from other information. These functions of affect are different from the functions of attitudes as attitudes come from past experiences that an individual has had with the attitude object and then serve as a summary evaluation of the object that is drawn from memory (Fazio, 2000). In the present studies, subjects draw meaning (and perhaps affective meaning) from the stimuli and structure of a novel task.

In Study 4, we used a reaction time paradigm to investigate more directly whether an affective or cognitive component of evaluation was more consistent with the evaluative-categories effect. We also briefly discuss whether the results are more consistent with cognitive evaluation or one of the functions of affect.

A diverse set of subjects was recruited across four studies (nonstudent adults ranging from 18 to 64 years old in Study 1, older adults aged 65 to 99 years in Study 2, and college students in Studies 3 and 4). This approach allows for an examination of the generalizability of any effect.

Study 1

In health and other decision-making contexts, it is often necessary to integrate multiple pieces of information to judge an option. In the present study, we examined whether subjects could integrate three different indicators of hospital quality when provided with only numerical information (as is usually the case) versus when provided with those same numbers in an evaluative-categories format. Evaluations like these are important, for example, when deciding whether to go to a convenient hospital or to one that is highly rated. In Study 1, consumers were provided with commonly used numeric indicators of hospital quality and judged the attractiveness of a hospital and their feelings about choosing it. Without additional aid, we expected consumers given numbers only to be relatively unable to use such unfamiliar (albeit important) information in their judgments. Such a finding would be consistent with Hypothesis 1 that consumers will not always use numeric information in judgments.

Individual differences, however, exist in the ability to draw meaning from numbers. Recent research in *numeracy*—defined as the ability to understand basic probability and mathematical concepts (e.g., Lipkus, Samsa, & Rimer, 2001; Woloshin, Schwartz, Black, & Welch, 1999)—suggested that people differ substantially in the ability to process numbers and that, in fact, many people are “innumerate” (Paulos, 1988). Peters, Västfjäll, et al. (2006) recently began to address how numeracy relates to judgments and decisions beyond simple comprehension of numbers. They found that highly numerate decision makers appeared to draw more precise affective meaning from numbers and numerical comparisons and to use this affect to guide them in making decisions; higher numeracy was linked with greater use of numbers and generally (but not always) with better decisions. The less numerate were also influenced by affect, but not by affect toward numeric information. Instead, their judgments were associated with competing, irrelevant affective considerations such as the frame of provided information.

The aim of Study 1 was to examine whether numerical information, an irrelevant source of affect (a naturally occurring mood state), and evaluative categories differentially influenced hospital evaluations among individuals who varied in numeracy. Consistent with Hypothesis 1A, we expected the highly numerate to demonstrate use of numbers in the absence of evaluative categories whereas the less numerate were expected to be relatively insensitive to the numeric information but to rely instead on their mood state, consistent with substituting mood for numeric information in their judgments (i.e., Hypothesis 2). Use of a very important quality indicator (survival rates) was expected to be minimal even for the highly numerate because small absolute differences exist between hospitals (the differences are nonetheless quite important), and consumers have little experience with such indicators.

The presence of evaluative categories was expected to help consumers interpret the meaning of numeric information and thus increase its use. Particularly less numerate consumers were then expected to use provided numeric information and not substitute naturally occurring mood states as a source of information in their judgments (Hypothesis 3 and 3A). Evaluative categories were also expected to improve the use of unfamiliar survival-rate information that consumers recognize as important but may be unable to evaluate low and high values without aid. The addition of evaluative categories thus may shift what information consumers use as they are able to evaluate all of the information and can choose what information is most important. Experts often believe that subjective measures of quality (such as patient ratings of quality of care) are less important than objective measures such as survival rates (Chernew & Scanlon, 1998). It is not clear how consumers will respond when all information is easy to evaluate.

Method

Subjects. Nonstudent adults ($N = 303$; mean age = 37.4 years; age range = 18 to 64 years; 52% women) recruited from the local community were paid \$20 for completing this survey plus unrelated tasks not reported in the present paper. Forty-five percent of the sample had a high school degree or less, 37% had some trade school or college, and 19% had a college degree or more.

Task materials and procedure. On the page prior to this study in a presumably unrelated task, we asked subjects about their mood

("How do you feel right now? I feel good/bad/happy/upset") based on four 5-point scales ranging from 0 (*not at all*) to 4 (*extremely*); an average of the bad and upset mood was subtracted from the average of the good and happy mood to calculate overall mood. Scale reliability was acceptable (Cronbach's $\alpha = .73$).

Subjects were randomly assigned to one of two formats for presenting hospital quality information (numbers only and numbers with evaluative categories shown in the top and bottom, respectively, of Figure 1). In each condition, subjects were shown three quality indicators about a hospital and received one of three levels of each quality indicator in a full factorial design. As a result, subjects in each format were randomized to see one of 27 possible indicator combinations, allowing us to estimate the subjective weight placed on each indicator in subsequent regression analysis. The study was powered to detect medium-sized effects of the indicators at 85% and 86% in the evaluative-categories and numbers-only conditions, respectively.

The quality indicators and numeric levels were chosen (as they were throughout the present paper) to represent information consumers might see. Thus, subjects were given unfamiliar information with numeric differences that were sometimes small but represented meaningful differences on quality indicators judged important by experts. In Study 1, three numeric levels were chosen to reflect low, medium, and high quality hospitals on each indicator, based on a review of hospital-comparison websites and in consultation with domain experts. The three quality indicators and their three levels were: (a) survival rate—percentage (%) of pneumonia patients who survive while being treated (93%, 96%, or 99%), (b) recommended treatment—percentage of heart attack patients given recommended treatment (ACE inhibitor; 64%, 74%, or 84%), and (c) patient rating—percentage of hospital patients who rated their care as very good or excellent (68%, 78%, or 88%). The first two indicators are more objective and arguably more important indicators of the quality of medical care offered by a hospital. Numbers only were provided in one condition because such information is normally provided in this format on hospital websites and elsewhere. In the evaluative-categories condition, the three levels of each indicator fell into the "fair," "good," and "excellent" categories, respectively.

Subjects were asked, "Imagine that you need to choose a hospital. How attractive is this hospital to you?" on a 7-point Attractiveness scale ranging from -3 (*extremely unattractive*), 0 (*neither attractive nor unattractive*), to $+3$ (*extremely attractive*). Next, they were asked, "How good or bad would choosing this hospital make you feel?" on a 7-point Affect scale ranging from -3 (*very bad*) to $+3$ (*very good*). The second item was intended to tap into experienced anticipatory affect toward the hospital. Finally, subjects responded to a 15-item numeracy scale used in Peters, Dieckmann, et al. (2007). Each item was scored as correct (1) or incorrect (0); a sum was calculated for each subject and could range therefore from 0 to 15. This 15-item scale showed significant negative skewness (skewness = $-.68$, $Z = 4.9$, $p < .01$). A square root transformation resulted in data that were not significantly different from normal (skewness = $-.16$, $Z = 1.14$, $p = .25$); results were similar regardless of the scale used. The transformed scale was used in subsequent analyses.

Results and Discussion

Numeracy. Scores averaged 9.3. Scale reliability was acceptable ($\alpha = .83$). The square-root transformation of numeracy was used as a continuous, mean-centered variable in inferential analyses. When it was helpful to identify individuals lower and higher in numeracy, a median split was used with respective scores of 0 to 9 and 10 to 15 on the untransformed scale.

Mood. Mean (median) mood scores were 1.8 (2.0) overall and 1.5 (2.0) and 2.1 (2.0) for those low and high in numeracy, respectively, $t(261) = 3.3$, $p < .01$. The possible range of the scale was -4 to $+4$, indicating that subjects were in mildly positive moods on average.

Mean of attractiveness and affect. The original intent was to examine variables representing attractiveness and affect separately. However, the two variables were highly correlated ($r = .72$, $p < .01$), suggesting the existence of a single evaluative construct. The average of the two scales (each scale was standardized first) was highly reliable (Cronbach's $\alpha = .83$; range = -1 to $+1$) and was used in all subsequent analyses. We modeled these average evaluations as a function of the level seen by the subject for each quality indicator, self-reported mood, numeracy level, and interactions of numeracy with every other variable (see Table 1). Regression analyses were conducted separately for subjects in the numbers-only condition and evaluative-categories condition. Hypotheses 1 to 3 were tested by examining the proportion of variance in judgments explained (R^2) and the significance of quality-indicator effects, mood, and their interactions with numeracy.

The numbers-only condition. Although we expected little use of numeric information in the numbers-only condition based on Hypothesis 1, two of the three indicators (patient rating and recommended treatment) were used significantly by subjects at an average level of numeracy and after controlling for the other variables in the model, $\beta = .17$, $p = .03$, $f = .19$ (patient rating; according to Cohen, 1992, $f = .10$ and $f = .25$; correspond with

Table 1
Study 1: Beta Weights From Two Regression Analyses With Subjects' Evaluations

	Analysis of numbers-only condition ^a		Analysis of evaluative-categories condition ^b	
	β	p	β	p
Survival rate	-.02	.82	.35	<.01
% treated	.15	.05	.24	<.01
Patient rating	.17	.03	.10	.18
Mood	.12	.15	-.14	.08
Numeracy	-.10	.22	-.11	.20
Interactions of numeracy with:				
Survival rate	-.15	.06	.05	.52
Recommended treatment	-.01	.95	-.06	.58
Patient rating	.12	.12	-.05	.51
Mood	-.19	.02	.04	.65
R^2	.15		.24	

Note. The mean of attractiveness and affect ratings after standardizing each separately with the hospital as the dependent variable.

^a $n = 152$. ^b $n = 145$.

small- and medium-sized effects, respectively), $\beta = .15$, $p = .06$, $f = .16$ (recommended treatment; see the numbers-only column of Table 1); overall model, $F(9, 142) = 2.8$, $p < .01$, $R^2 = .15$. This significant use of numeric information by the average subject was somewhat unexpected based on Hypothesis 1 although, as will be seen, the proportion of variance explained by the numeric indicators (and not mood) was only small to moderate. As expected, the survival-rate indicator, although important, was not used; consumer ability to use such indicators with small and familiar differences between options may particularly benefit from an evaluable-categories information display.

We also examined interactions with numeracy. Consistent with Hypothesis 2 (that individuals should be more likely to substitute less relevant affective information for numeric information into their judgments when numeric information had less evaluative meaning), less numerate subjects in more positive moods compared to more negative moods evaluated the hospital more positively, $\beta = -.19$, $p = .02$, $f = .19$ (see Figure 2 and the numbers-only condition of Table 1). Based on Hypothesis 1A, we expected the highly numerate to use the quality indicators more. In fact, the use of survival rate information did demonstrate a tendency to be moderated by numeracy, $\beta = -.15$, $p = .06$, $f = .16$. The direction of the effect was opposite that expected, however, and highly numerate subjects evaluated hospitals with the lowest survival level more positively than hospitals with the highest survival level. Similar to the gambling study with the five-cent loss (Peters, Västfjäll, et al., 2006), the highly numerate showed a greater use of numbers but not a superior use of them. We speculate that, whereas the less numerate simply did not use these unfamiliar numbers, the highly numerate may have searched for meaning in the numbers and used the context to bring to mind a number comparison. The unfamiliarity of the context, however, may have led to an inappropriate use of context. For example, the 93% survival context may have brought to mind lower survival rates (e.g., in the 80% range), causing the 93% survival to appear good whereas the 99% survival context might have brought to mind that some people will perish, causing the 99% survival to appear relatively bad in this between-subjects design. Such admittedly post hoc speculation deserves further study, but is consistent with prior research demonstrating the role of context in bringing to mind number comparisons (Birnbau, 1999).

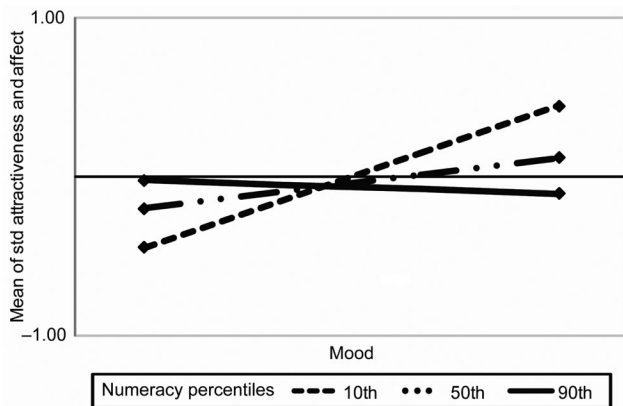


Figure 2. Study 1: Predicted evaluation in the numbers-only condition based on the Mood \times Numeracy interaction. Std = standardized.

To explore the mood-numeracy interaction and examine whether the highly numerate were able to use provided numeric information without the aid of evaluative categories, we conducted regression analyses separately for lower and higher numerate subjects (based on a median split of numeracy). The analysis revealed that ratings of the highly numerate did not depend on mood, $\beta = -.09$, $p = .41$, $f = .10$. The highly numerate demonstrated significant use only of the patient-rating indicator ($\beta = .33$, $p < .01$, $f = .37$; for patient rating). Note that they did not significantly use the survival rate indicator, $\beta = -.16$, $p = .12$, $f = .18$; and $\beta = .16$, $p = .14$, $f = .17$; for survival rate and recommended treatment, respectively; model $F(4, 75) = 3.8$, $p < .01$, $R^2 = .17$. As expected from Hypothesis 1A, the identical model conducted among less numerate subjects explained a more moderate proportion of the variance in their evaluations than that for the highly numerate, for the less numerate: model $F(4, 67) = 2.1$, $p = .09$, $R^2 = .11$. Consistent with the substitution idea of Hypothesis 2, hospital evaluations of the less numerate were only significantly associated with their mood ratings ($\beta = .27$, $p = .02$, $f = .35$); they did not significantly use any of the numeric information ($\beta = .06$, $p = .58$, $f = .08$; $\beta = .10$, $p = .38$, $f = .16$; and $\beta = .12$, $p = .30$, $f = .13$ for the respective indicators of patient ratings, survival rate, and recommended treatment). Overall, Hypothesis 1A that the highly numerate would use numeric indicators more than the less numerate received only partial support, because none of the interactions of indicators with numeracy were conventionally significant in the omnibus analysis. The combined results support Hypothesis 1 that consumers will not always use important numeric information in forming judgments.

The evaluative-categories condition. Analysis of data from the evaluative-categories condition revealed a greater proportion of variance in the evaluations being explained by the three numeric indicators compared to when numbers only were provided (without the biasing effects of mood in the model, the proportion of variance explained in evaluations was greater in the evaluative-categories condition compared to the numbers-only condition, their respective $R^2 = .19$ and $.06$, $Z = -1.93$, $p = .054$). These results were consistent with Hypothesis 3 that the use of evaluative categories will result in a greater use of numeric information in judgment.

With the inclusion of self-reported mood, subjects at the average numeracy level significantly used the two objective indicators—survival rate, $\beta = .35$, $p < .01$, $f = .34$; and recommended treatment, $\beta = .24$, $p < .01$, $f = .24$ (see the evaluative-categories column in Table 1; overall model with mood, $F(9, 136) = 4.9$, $p < .01$, $R^2 = .24$). With the meaning of these arguably more important indicators as easier to evaluate, consumers no longer used the more subjective patient ratings or their naturally occurring mood states, and no interactions emerged with numeracy. Thus, consumers who varied in numeracy were aided by the evaluative-categories manipulation. With evaluative categories, subjects were presumably able to draw meaning from and use the important hospital quality indicators and, with this greater use of numbers, did not rely on irrelevant affect as a source of information, consistent with Hypothesis 3A. Separate analyses of individuals low and high in numeracy revealed little difference in beta weights of predictors or the proportion of variance explained in their ratings ($R^2 = .21$ and $.24$, respectively, for those low and high in numeracy).

These results suggest that consumers are not adept at using information presented in raw numerical form. Unfortunately, these types of numbers (presented simply as percentages in tables) reflect what is often offered in health-care decision environments. Characteristics and states of the individual (e.g., numeracy or mood states) appear to alter the interpretation and weighting of numeric information when no other help is offered. However, when information providers consciously attend to how information is presented, they can help decision makers use important numeric information more in judgments, consistent with Hypothesis 3.

Finally, the mood results from the less numerate in the numbers-only condition provide some preliminary support for our contention that affect is perceived as a relevant source of information when processing numbers (see also Peters, Västfjäll, et al., 2006) although such correlational results must be interpreted with care. The evaluative-categories manipulation may have led to increased accessibility of feelings about the quality indicators, allowing for effective integration of multiple indicators into the judgment and less substitution of mood. Affect as a possible mechanism was tested in Study 4.

Study 2

In Study 2, we were interested in whether evaluative categories might influence choices even when two options were present (joint evaluation) and decision makers could presumably compare numeric data about the options to make them evaluable. We suspected that the comparison process would be an imperfect one and that evaluative categories would nonetheless influence the interpretation and coding of numeric information and thus would influence choices (consistent with Hypothesis 3). Also, because Study 1 did not allow us to isolate the effect of evaluative categories from the effects of a graphical presentation (the graphical evaluative-categories condition was tested against a numbers-only condition in Study 1), we also compared only graphical formats with and without evaluative categories. We included a third graphical format to test the effect on choice of visual boundary lines (without the labels) and examined whether any effect was due to the evaluative labels or simply the use of a bar chart that categorizes the scores (i.e., the visual boundary lines) and provides information about expected ranges of scores. Although Rothbart,

Davis-Stitt, and Hill (1997) found an effect of categorization on similarity judgments, it is not clear whether visual boundary lines could influence the evaluability and weighting of an attribute in choice without an explicit valenced evaluation from the label. We tested older adults in this study because robust age-related declines in deliberative capacity have been associated with greater effects of categorical information (Peters, Hess, Västfjäll, & Auman, 2007) so that tests in this age group were thought likely to maximize the chances of finding an effect of the boundary lines.

In the United States, older adults must make Medicare health-insurance choices that can influence their finances, health-care quality, and thus their quality of life. In the present study, we tested the influence of evaluative categories on such choices. Three conditions were employed: (a) bar charts only (no visual category lines, no labels), (b) visual category-lines only (no labels), and (c) evaluative categories (visual category lines and evaluative labels). Subjects were faced with the five-attribute, six-option choice shown in Figure 3.

Based on Hypothesis 3 (that helping consumers to evaluate the meaning of numeric information would influence how they code and interpret numbers in choice), we predicted that evaluative categories would influence health-plan choices such that the proportion of subjects who preferred the health plans in the better evaluative category when categories were present would be greater when the categories were present versus absent. Although older adults rely more than younger adults on categorical types of information (e.g., stereotypes and schemas; Peters, Hess, et al., 2007), we nonetheless expected that even with older adults, the use of visual category lines alone would not influence choices relative to a bar-chart-only condition because the lines alone do not convey valence. Unlike Study 1 in which evaluative categories provided superior processing in judgments (e.g., the R^2 was greater with evaluative categories than without them), in Study 2, no option was clearly best. Instead, we examined the extent to which an attribute was weighted in choice depending on whether the evaluative-categories manipulation delineated (or “marked”) options into different evaluative categories on that attribute.

Method

Subjects. Older adult subjects ($N = 207$; 66% women; mean age = 76 years; age range = 65 to 99 years) were recruited

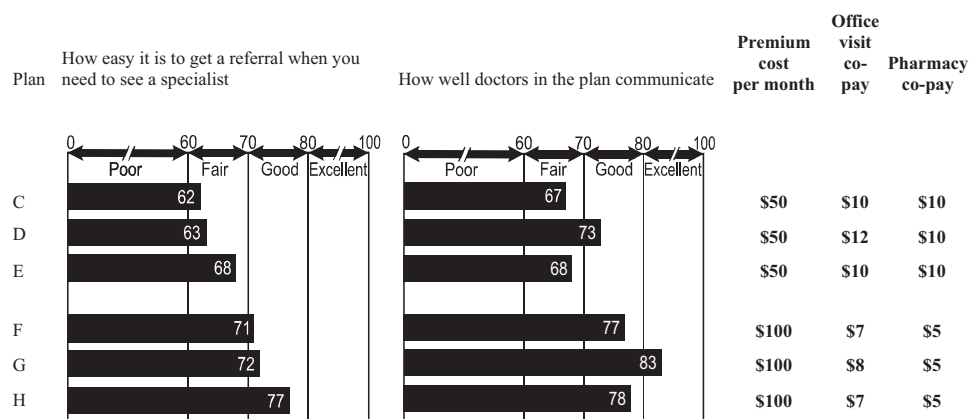


Figure 3. Studies 2 and 4: The evaluative-categories condition in an elaborated choice.

through local senior centers and community advertisements. Each older adult was paid \$20 for participation; the senior center was paid an additional \$5 per subject. Subjects also completed several unrelated decision tasks not reported in the present paper.

Task materials and procedure. Subjects were presented with a choice among six health-care plans that differed on two quality-of-care measures (how well doctors in the plan communicate and how easy it is to get a referral when you need to see a specialist) and three cost measures (monthly premium costs, office visit copay amounts, and prescription copay amounts). Quality-of-care information was presented in bar chart format with the actual score received embedded in the bar chart (see Figure 3). The information for one third of the subjects (the evaluative-categories condition) was supplemented (i.e., the health plans were categorized as *poor*, *fair*, *good*, or *excellent* through the use of category boundaries and labels). Another one third of the subjects received the same information with category boundaries and no labels (the visual category-lines-only condition) whereas the remaining subjects received information without any category or label information (the bar-charts-only condition). As Rothbart et al. (1997) pointed out, the labels and boundaries could be perceived as providing new information that is not redundant with the numerical data. We attempted to minimize their meaningfulness by creating the perception of arbitrary category boundaries. To make the arbitrariness clear, in the first two conditions, subjects were told, "we created ranges you can use to help you think about how well each health plan did. These ranges vary somewhat unpredictably from year-to-year because they are based on the number of plans in our analysis and the scores received by those plans" (see Rothbart et al., 1997).

The three types of cost information shown in Figure 3 were intended to add complexity to the task, similar to actual choices subjects might face, without differentiating the plans much, so that subjects would still focus on the two quality attributes. As a result, the cost information varied little among the plans. There were two levels of monthly premium costs, \$50 and \$100. Within the \$50 plans, the pharmacy copays were always \$10; within the \$100 plans, the pharmacy copays were always \$5. The office visit copays were set up in a similar manner except that the health plan (within each premium cost grouping) in the better evaluative category had slightly higher office visit copays. Thus, the \$50 plans had office visit copays of \$10 except that Plan D in the better evaluative category had a copay of \$12. The \$100 plans had office visit copays of \$7 except that Plan G in the better evaluative category had a copay of \$8.

The attribute information was designed such that Plans D and G were in the highest evaluative categories within their premium cost grouping on both attributes. Health plans C and F were dominated by the other plans. Between plans D and E, no best option existed, but plan D was in a better category when evaluative categories were present. Similarly, a tradeoff existed between plans G and H, but G was in a better category when the categories were present. The specific scores for ease of referral and doctors' communication were counterbalanced across subjects such that, for half of the subjects, the average doctor communication scores were higher (doctor communication averaged 74 across the six plans in Figure 3 compared to an average score of 69 for ease of referral); for the other half, the scores were reversed and average ease of referral scores were higher.

Subjects chose one of the six plans. They then indicated for each of the five attributes how much that attribute had influenced their choice based on a 9-point scale ranging from 0 (*did not influence me at all*) to 8 (*influenced me a great deal*).

Results and Discussion

Consistent with Hypothesis 3, the weight accorded a numeric attribute in choice was influenced by evaluative categories. Fifty-four percent of subjects in the evaluative-categories condition chose Plan D or G compared to only 35% and 36% of subjects in the bar-chart-only and visual-boundary-only conditions, respectively. Choices in the bar-chart-only condition differed from the evaluative-categories condition, $\chi^2(1, n = 129) = 4.7, p = .03, w = .18$; but the bar-chart only and category-boundaries conditions did not differ from each other, $\chi^2(1, n = 133) = .01, p = .91$. The study had 93% power to detect a medium-sized effect between any two conditions at $p < .05$.

Although choices differed between the three conditions, no significant differences existed between them for any of the five attribute importance ratings (p values ranged from .19 to .92; $p = .63$ for doctor communication when Plans D and G were in a better category for doctor communication in the presence of evaluative categories and $p = .40$ for ease of getting a referral when Plans D and G were better on it in the presence of evaluative categories). The study was somewhat underpowered, however (57% power to detect a medium-sized group difference between importance ratings at $p < .05$). We provided more adequate power in Study 3.

Results of the present study were consistent with the importance of the evaluative labels themselves and an influence of evaluative categories on the interpretation and weighting of information that is "marked" with the labels. It is possible however, that the labels were used as affective or cognitive information to simplify the decision and the numeric information was processed more superficially.

Study 3

In Study 3, the evaluative meanings of health-plan data again were manipulated through the presence versus absence of evaluative categories in choices between health-insurance plans that involved tradeoffs between two attributes. We tested its influence on choices among college students and examined again whether the manipulation influenced attribute importance ratings. In addition, we tested whether decision makers relied so much on evaluative categories that they failed to process (or processed only superficially) the more precise numeric information.

Consistent with Hypothesis 3, which concerned the influence of evaluative categories on the interpretation and weighting of numeric information, we expected subjects to more strongly prefer the health plan (in a pair of plans) with an attribute in a better evaluative category when evaluative categories were present versus absent. In addition, if the presence of evaluative categories caused decision makers to rely on the information content of the labels and only superficially process the numeric information, then subjects in the evaluative-categories condition should remember the detailed numeric information less well. At the same time, the evaluative-categories manipulation might increase gist memory for which of two plans is better either due to attending to the labels

and not the numbers or because the labels provided additional cognitive or affective information for simpler heuristic processing.

Method

Subjects. Subjects ($N = 218$; 54% women; mean age = 21.5, age range = 18 to 51 years) responded to an ad in the student newspaper. They worked on the present task plus several other tasks (not reported in the present paper) during a 45- to 60-min session, and were paid \$10 for their participation.

Task materials and procedure. Subjects were asked to compare two health-insurance plans that differed on two attributes—quality of care received and overall member satisfaction. Scores on these attributes could vary between 0 and 100 and were shown on bar charts. For both attributes, the health-insurance plans differed by seven points (subjects were told that anything fewer than three points apart might not indicate real differences). One plan was always better on one attribute and worse on the other so that there was no one “right” answer. Cost information was also included. The high quality, low satisfaction plan always cost \$60; the low quality, high satisfaction plan always cost \$45. Subjects examined the attribute information and then expressed a preference between the plans based on a 7-point scale ranging from 1 (*health-insurance Plan A much more preferred*) to 7 (*health-insurance Plan B much more preferred*). In addition, subjects indicated the relative importance of quality of care, member satisfaction, and cost to the choice they just made by dividing 100 points between the three attributes.

On a final page, subjects were asked to estimate the score for each plan, A and B, on quality of care, overall satisfaction, and cost without looking back at the information provided earlier. We examined the verbatim estimates and recoded them to indicate whether the subject accurately remembered the gist of which plan scored higher on each attribute. For example, based on the information in Figure 4, if the subject estimated that B scored a 70 on member satisfaction and A scored 60, the subject would be coded as accurately remembering the gist that B was better even though the precise numbers were not remembered correctly.

Subjects randomly received one of four versions of the study in a 2 (presence/absence of evaluative categories) \times 2 (two plans

scored 61 and 68 on quality of care and 63 and 56 on overall satisfaction or the reverse) design. See Figure 4 for two of the four versions. Half of the subjects were shown bar charts that included evaluative categories (poor, fair, good, excellent). On one of the attributes, the two plans fell in different evaluative categories (quality of care for half of the subjects—see the top right side of Figure 4—and member satisfaction for the other half); the two plans were in the same evaluative category for the other attribute. For the other half of the subjects, the same bar charts were presented without evaluative-categories information and merely ranged from 0 to 100. Plan A always had an attribute in a better evaluative category when categories were present. More choices of plan A when evaluative categories were present versus absent would support Hypothesis 3.

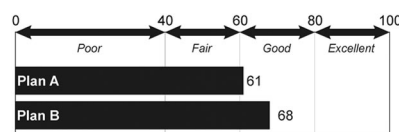
Results and Discussion

Fifty-four percent of subjects chose health-insurance Plan A when evaluative categories were present compared to only 39% who chose it when they were absent: overall analysis of variance model, $F(3, 214) = 8.0, p < .01$; evaluative-categories effect, $F(1, 214) = 5.5, p = .02, f = .16$; counterbalancing effect, $F(1, 214) = 18.1, p < .01, f = .29$. Similar effects were shown in both conditions, when the two plans scored 61/68 on quality of care and 63/56 on overall satisfaction and the reverse; interaction, $F(1, 214) = 0.3, p = .59$. Similar to Study 2, subjects did not rate any attribute as significantly different in importance in the presence versus absence of evaluative categories (p values ranged from .33 to 1.0). The study had adequate power to detect a medium-sized effect in attribute importance differences (96% power at $p < .05$).

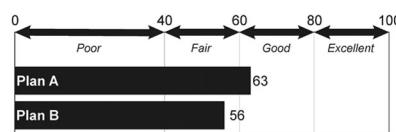
It may be, of course, that the evaluative-categories manipulation helps decision makers to cognitively simplify their choice, and decreases the need to examine and compare the numbers to determine which one is larger. If the presence of evaluative categories causes decision makers to only superficially process the numeric information compared to when evaluative categories were absent, then evaluative-categories subjects should have worse memory for precise numeric information. No significant differences existed, however, in the proportion of subjects who correctly remembered precise numeric information with versus without evaluative cate-

Condition 1: Evaluative categories

Member Satisfaction with HMO

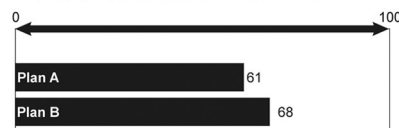


Quality of Care Received



Condition 2: No evaluative categories

Member Satisfaction with HMO



Quality of Care Received

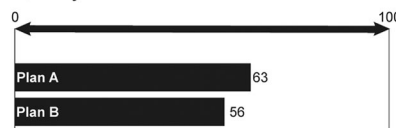


Figure 4. Study 3: The evaluative-categories manipulation. HMO = Health Maintenance Organization.

gories. For example, 22% and 26% in the bar-chart-only and evaluative-categories conditions, respectively, correctly remembered the numeric information about patient satisfaction for Plan A; results of a logistic regression of memory accuracy with the presence versus absence of evaluative categories, the counterbalancing, and their interaction as independent variables revealed no significant effects, evaluative-categories Wald $\chi^2(1, N = 218) = 0.4, p = .51$; the same numbers for Plan B were 14% and 25%, logistic regression: evaluative-categories Wald $\chi^2(1, N = 218) = 2.9, p = .09$. The numerical difference in the latter case was in the direction opposite to what one might expect if evaluative categories were substituted for the numeric information. Similar results were found with respect to memory for the quality-of-care indicator.

Of interest, subjects remembered Plan A's cost information significantly less well when evaluative categories were present (64% and 46% of subjects in the bar-chart-only and evaluative-categories conditions, respectively), remembered Plan A's cost; in a logistic regression of memory accuracy similar to the analyses conducted for the quality indicators, evaluative-categories Wald $\chi^2(1, N = 218) = 6.8, p < .01, f = .18$. Memory accuracy for Plan B's costs did not depend on the presence of evaluative categories (66% and 58% in the bar-chart-only and evaluative-categories conditions, respectively), accurately remembered Plan B's costs; logistic regression: evaluative categories Wald $\chi^2(1, N = 218) = 1.8, p = .18$. These results are consistent with decision makers attending to and overweighing cost information in the bar-chart-only condition as has been reported in early consumer studies on the effects of providing quality information (Chernew & Scanlon, 1998). Cost may usually be easier to evaluate, but decision makers may process it less when important quality information is made easier to evaluate. Adequate power (96% at $p < .05$) existed to detect medium-sized differences between the conditions.

This same cognitive-simplification argument also predicts that decision makers given evaluative categories will be better able to remember the gist of which plan is better on an attribute in certain circumstances. In particular, evaluative-categories subjects may have superior memory for the gist of which plan is better than no-evaluative-categories subjects when the two plans are in separate categories because no number processing is required from the former group (if they rely on the labels for a cognitive or affective evaluation of the difference between the plans). If the two plans were in the same category, however, then the no-evaluative-categories subjects should demonstrate superior memory because the labels are the same for the two plans and the evaluative-categories subjects would process the numbers more superficially. Finally, evaluative-categories subjects should have superior memory for the better plan when the two plans fall in different categories for an attribute than when they are in the same category for similar reasons. Superior memory for the gist of the difference on the quality indicators would be consistent with the evaluative categories adding additional information.

The cognitive-simplification predictions were not supported. First, subjects were not significantly different in their memories of which plan was better with versus without evaluative categories. Specifically, when the two plans were in different categories for member satisfaction and categories were present, 66% of evaluative-categories subjects versus 74% of bar-chart-only subjects remembered which plan was better for member satisfaction

($p = .39$); 68% and 74% of these evaluative-categories and bar-chart-only subjects, respectively, accurately remembered which plan was better for quality of care ($p = .51$). When the two plans were in different categories for quality of care instead, 79% of evaluative-categories subjects and 73% of bar-chart-only subjects remembered the better plan for quality of care ($p = .47$) and 75% and 65% in each group, respectively, remembered which plan was better for member satisfaction ($p = .25$). Summary logistic regressions predicting the accuracy of gist memory for either indicator showed no significant results for condition, the counterbalancing, or their interaction. Although the choice made was a relatively simple one, subjects were asked about their memory immediately after choice, and overall gist memory was better than memory for the precise numbers, gist memory did not differ between the conditions nor did it show a ceiling effect. Adequate power existed to detect medium-sized effects (99% power at $p < .05$).

Because this analysis is based on a between-subjects design, it is possible that random assignment resulted in between-groups memory differences. Therefore, we also examined the cognitive-simplification hypothesis within the evaluative-categories condition. These evaluative-categories subjects were no better at remembering which plan scored better when the plans were in different evaluative categories versus the same evaluative category. For example, when the two plans were in different categories for member satisfaction (and the same category for quality of care), 66% correctly remembered which plan was better for satisfaction and 68% remembered correctly for quality of care ($p = .66$). In the other condition, when the two plans were in different categories for quality of care instead (and the same category for member satisfaction), 79% and 75% correctly remembered which plan was better for quality of care and satisfaction, respectively ($p = .46$). Adequate power existed to detect a medium-sized effect (96% power at $p < .05$), and subjects clearly processed the numeric information at some level given the relatively large proportion of subjects who remembered the gist of the difference on each quality indicator whether the plans were in the same or different evaluative categories for that indicator.

Providing evaluative categories might have made it easy to skim over the numeric information and simply remember whether the outcome was good or bad, but this cognitive-simplification explanation was not supported. Although we cannot rule out all alternative explanations for the lack of effect, it did not appear that the labels encouraged superficial processing of the numbers with the labels used as substitute information in a heuristic fashion. Instead, the results were more consistent with subjects attending to the numbers provided in both conditions but interpreting and weighting the information differently in the two conditions.

The lack of memory effects for the gist of the difference between the options is also inconsistent with evaluative categories providing cognitive or affective information to the choice process. Instead, it may be that the manipulation provided affective meaning to the process that functioned in a different manner. For example, affect may have directly motivated choice similar to the approach and avoidance tendencies discussed by Chen and Bargh (1999). Alternatively, the results demonstrating significantly worse memory for costs in the presence of evaluative categories are congruent with the function of affect as a spotlight. It may be that evaluative categories directed attention away from the cost

information—decreasing negative affect to the more expensive Plan A—rather than the evaluative categories directing attention to the quality attribute—thus increasing positive affect and motivating choice of the better quality Plan A. It is also possible that subjects feel better and more confident about choosing a plan in a better evaluative category so that the labels provide affect as information in a more indirect way that does not influence memory. Whether the mechanism underlying the influence of evaluative categories is based in affect, however, remains an open question as none of the previous studies examined process data for affect.

Study 4

Results of the previous studies have demonstrated a robust, although modest, effect of evaluative categories on choices and judgments. In Study 4, we borrowed a methodology from the attitude literature, based on such evaluations having a tripartite structure, and examined the possible separable effects of affective and cognitive components of evaluation. Testing for such separable mechanisms is difficult, of course, given the subtle nature of affect and the inherent ties between affective and cognitive components of evaluation. We turn to a reaction time experiment and the accessibility of these components for evidence.

If affect is a driver of the evaluative-categories effect, then one might expect affective feelings to come to mind more quickly than cognitive thoughts in the presence of evaluative categories compared to their absence and for affect to be more strongly related to choices than thoughts are. If, on the other hand, the manipulation adds more cognitive information (but information that nonetheless did not help decision makers remember the gist of the information as tested in Study 3), then one might expect thoughts to be accessed more quickly than feelings in the presence of this manipulation. Study 4 provided a test of these opposing hypotheses using a methodology developed originally by Verplanken, Hofstee, and Janssen (1998). They demonstrated that subjects responded to affective evaluations faster than cognitive evaluations in two attitude domains (brand names and countries), suggesting that affect-based evaluations were more accessible in memory than cognitive-based evaluations.

Method

Subjects. Subjects were 83 undergraduate students (61% women; average age = 20.4, age range = 18 to 37 years). The sample size was selected to provide adequate power to detect the hypothesized reaction time effect based on the Verplanken et al. (1998) studies. Data from an additional 11 subjects were deleted due to problems while reaction times were being recorded (subject coughing or sneezing frequently, noisy interruptions).

Task materials and procedure. Subjects were presented with three health-plan choice tasks in Phase 1 of this computer-controlled experiment. Two of the choices were simple ones similar to the two-option, three-attribute choice of Study 2 (one choice included cost; the second did not); the third choice was the complex choice shown in Figure 3. The order of the tasks was randomized for each subject. Half of the subjects viewed each choice in a bar-chart-only format whereas the other half viewed each choice in an evaluative-categories format. Subjects made choices

by pressing the button that corresponded with their preferred health plan. This phase was intended to familiarize subjects with the choice options so that they could more easily report their thoughts and feelings next.

Phase 2 was the experimental portion of the study derived from Verplanken et al. (1998). In this phase subjects were asked to indicate how they felt and how they thought about the health plan highlighted by evaluative categories when they were present (i.e., Plan G in Figure 3 and Plan A in Figure 4). Using blood donation as an example, subjects were told that one may have thoughts as well as feelings about the same object and that the thoughts and feelings may be the same or they may be different. Subjects were then told that they would be asked to respond as quickly and accurately as possible to a list of bipolar statements concerning one health plan in each of the choice scenarios they just saw. For each specified health plan (e.g., Plan G in Figure 3, Plan A in Figure 4), they were asked first about their feelings about the plan and then about their thoughts about that plan; the order was reversed for half of the subjects. The order of the health-plan choice scenarios was randomized for each subject. For each trial, the following appeared on the screen in vertical order: the statement “your feelings about Plan [x]” or “your thoughts about Plan [x],” the information from one of the health-plan choices the subject saw previously in Phase 1 (e.g., Figure 3), and a pair of bipolar adjectives (e.g., good–bad). Subjects pushed the corresponding “1” or “3” key on the number pad to express their feelings (in the feelings block) or their thoughts (in the thoughts block). For each health plan task, subjects responded to eight bipolar adjective pairs randomly presented within the feeling or thought block. The right–left orientation of the adjective pairs was randomized.

As in Verplanken et al. (1998), subjects responded to target and nontarget adjectives in each block. The target adjectives, identical in both blocks, were good–bad, negative–positive, and favorable–unfavorable. Thus, in the feelings block, subjects responded whether their feelings about a particular health plan (e.g., Plan G from Figure 3) were “good” or “bad” by pressing the corresponding button; in the thoughts block, subjects responded whether their thoughts about the plan were “good” or “bad” in similar fashion. Subjects also responded to five nontarget adjective pairs to create the affective or thought context desired. The nontarget adjectives in the affective (feeling) block were: unpleasant–pleasant, uncomfortable–comfortable, nervous–calm, excited–not excited, and wonderful–terrible. The nontarget thought adjectives were: useful–useless, unimportant–important, wise–unwise, expensive–cheap, and valuable–worthless so that a subject would indicate in one case whether his or her thoughts about the plan included “useful” or “useless.” To avoid confounding of response class and adjective, we only analyzed responses to the three target adjectives in each block. In addition to reaction time (RT), the valence (i.e., positive or negative) of each response was recorded.

RTs were recorded from the moment the adjectives appeared on the screen to the moment subjects pushed one of two response buttons. As in the Verplanken et al. (1998) studies, RTs greater than 4 s were treated as missing values; RTs less than 100 ms also were treated as missing values. The resolution of the RTs was 16.7 ms. Mean RTs were calculated from the three target items for both thoughts and feelings for each choice. Those six mean RTs were subjected to a $1/RT$ transformation to correct for skewness (e.g.,

Fazio & Hilden, 2001). Transformed mean RTs are reported in milliseconds in the text and tables.

Results and Discussion

A 3 (Health-Plan Tasks 1, 2, and 3) \times 2 (feelings or thoughts) multivariate analysis of variance (MANOVA) was conducted on the RTs with both factors as repeated measures and the presence versus absence of evaluative categories as the between-subjects factor. In the present study, unlike the Verplanken et al. (1998) original studies, feelings were not accessed more quickly overall than thoughts, $F(1, 76) = 0.5, p = .50$ with 77% power to detect a medium-sized effect at $p < .05$. As shown in Table 2, no significant RT differences existed between feelings and thoughts without evaluative categories (mean RTs = 1,545 and 1,460 ms, respectively, $p = .31$). However, as hypothesized, they accessed feelings about the options faster than thoughts when evaluative categories were present (mean RTs = 1,266 and 1,382 ms, respectively), $F(1, 41) = 4.5, p = .04, f = .31$; $F(1, 76) = 4.9, p = .03, f = .25$; for the interaction of the categories manipulation with feelings versus thoughts. This difference in the accessibility of feelings versus thoughts in the presence compared to the absence of evaluative categories did not differ across the three health-plan scenarios: three-way interaction, $F(2, 75) = 1.5, p = .23$.

To inspect the data for a possible confound with asking for thoughts versus feelings first, this order effect was introduced as a predictor. In addition, response valences of the items were included as predictors to control for a possible confound of thoughts versus feelings and valence of the responses (i.e., good vs. bad responses). The interaction of thoughts versus feelings with the evaluative-categories manipulation remained significant, $F(1, 70) = 4.2, p < .05, f = .25$.

In the present study, affective feelings came to mind more quickly than cognitive thoughts in the presence of evaluative categories despite the fact that the thought and feeling reactions were measured to the identical evaluative target items (e.g., good–bad). This finding mirrors Verplanken et al.'s (1998) original finding that feelings come first when considering brands of consumer goods. The present results indicate that, with evaluative categories, feelings come first. Because information that comes to mind first appears to have a disproportionate impact on choices

(e.g., query theory; Weber et al., 2007), these results point toward the influence of evaluative categories on choice as due, at least in part, to an affective rather than a cognitive mechanism. These results, however, highlight the point that feelings do not inevitably come first (in the absence of the manipulation, feelings and thoughts did not differ significantly in their accessibility). The present study implies that evaluative categories can be used to create a more precise affective feeling toward an option than exists without them.

The evaluative-categories manipulation also influenced the accessibility of feelings more than it influenced the accessibility of thoughts. It could be argued that evaluative categories provided a holistic evaluation of the option that was easier to access. Although such holistic evaluation may not tap exclusively into affect, the present findings are consistent with our argument that the manipulation has its influence more on feelings than thoughts, perhaps by increasing access to feelings in the evaluative-categories condition.

The focus of the present study was on RTs; we examined choices nonetheless to confirm that the direction of the effect was shown in each scenario. In a logistic regression of the first choice, 56% and 46% of evaluative-categories and bar-chart-only subjects, respectively, chose the highlighted plan ($p = .60$); these same numbers for the second choice were 53% and 44% ($p = .45$) and for the third choice were 63% and 43% ($p = .06, \phi = .20$).

Of course, in the present experiment subjects chose first and then gave valence responses so that it may be that we code choices either as an overall evaluation from which it is easier to derive feelings than thoughts or that we code choices directly as feelings. In fact, evaluative-categories subjects who chose the plan highlighted by evaluative categories responded more quickly to their feelings about that plan than their thoughts (mean RTs = 1,219 and 1,347 ms, respectively). However, subjects who did not receive the evaluative-categories manipulation but chose the same plan responded more slowly to their feelings than their thoughts (mean RTs = 1,387 and 1,271 ms, respectively). This pattern was consistent across all three choice tasks and suggests that the act of choosing an option does not have a simple causal effect on the accessibility of thoughts and feelings about it afterward.

Table 2

Study 4: Reaction Times in Milliseconds Toward Evaluations of Health-Plan Options Highlighted by Evaluative Categories in the Presence and the Absence of Evaluative Categories

Health-plan scenarios	Reaction times among subjects given bar charts only		Reaction times among subjects given evaluative categories	
	In the feelings context	In the thoughts context	In the feelings context	In the thoughts context
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
1. Health-plan simple	1,548 (553)	1,413 (602)	1,279 (472)	1,433 (632)
2. Health-plan simple	1,628 (598)	1,543 (644)	1,277 (495)	1,461 (539)
3. Health-plan complex	1,468 (666)	1,461 (513)	1,251 (500)	1,266 (416)
Average	1,545 (396)	1,460 (359)	1,266 (355)	1,382 (356)

Note. Interaction $F(1, 70) = 4.2, p = .04$; after controlling for order and valence.

How is it possible that responses to the same target items for feelings versus thoughts could have such different implications for choice? It may be that our thoughts are more mixed and ambivalent than our feelings (Trafimow & Sheeran, 2004). In fact, although the average interitem correlation for the three feeling target items across the three choices was .67 (r range was .54 to .88), the same average interitem correlation for these target items for thoughts was $-.16$ (r range was $-.25$ to $.02$). The mean valence response for thoughts was neutral ($-.01$ on a scale from -1 to $+1$) whereas the mean response for feelings was positive ($+.61$). Although subjects responded to the identical three evaluative target items in both the feelings and the thoughts context, the speed of responses, their valence, and the correlations of the valenced responses with each other in the two contexts differed markedly.

We speculate that thoughts carry information content that can be mixed and ambivalent whereas feelings code behavioral actions or likelihoods of acting (Chen & Bargh, 1999). The link between feelings and behavioral actions is consistent with the functions of affect as a motivator or as a spotlight and with the memory results of Study 3. It is not consistent with the function of affect (or cognition) as information with the possible exception that the subject may feel better or more confident by choosing a plan in a better category, and this process could indirectly influence choices through the subject's current feeling state. It is also possible that the study design caused inconsistency in thoughts more so than in feelings. Although further replication will help to determine any issues with the study design, we believe at this point that the difference may be inherent in the complexity of thought compared to the simplicity and power of affective feelings.

Finally, an applied contribution of Study 4's reaction time methodology is its potential use as a measure of affect in studies of judgment and decision making. Although interest in affective processes has grown, the limited number of methods to test hypotheses has been detrimental to theory development. We believe that Study 4's methodology may prove fruitful.

General Discussion

In four studies we demonstrated that an evaluative-categories manipulation influenced preferences and the weight accorded to attributes in judgment and choice across diverse samples. In addition, the influence of the manipulation depended on numerical abilities. For subjects low in numeracy, the presence of evaluative categories resulted in less influence of an irrelevant source of affect—naturally occurring mood states—in their judgments. The presence of evaluative categories also resulted in subjects, across numeracy levels, taking into account more numeric information (as measured by the model R^2) and effectively using an important but difficult to evaluate quality indicator, namely survival rates in a hospital. Consistent with research in other domains, how product information is communicated to consumers should be tested for its effectiveness just as the products themselves are tested (Fischhoff, 2009; Hibbard & Peters, 2003; Peters, Hibbard, Slovic, & Dieckmann, 2007).

Moreover, the results were more consistent with an affective mechanism (Peters, 2006; Peters, Lipkus, & Diefenbach, 2006) than they were with other explanations. Categorization through visual boundary lines, for example, did not influence choices in Study 2 compared to the control condition. The lack of memory

differences in Study 3 suggested that evaluative categories did not add cognitive information. Based on these results, we argue that manipulations such as evaluative categories may act as overt markers of affective meaning and guide choices. In a test of affect as a causal mechanism underlying the effect, affective feelings were accessed more quickly than cognitive thoughts in the presence versus absence of evaluative categories in Study 4. Given recent research concerning the greater impact of affect and emotion on older adult decisions compared to those of their younger counterparts (Peters, Hess, et al., 2007), evaluative categories may have a greater influence in older populations. This hypothesis remains to be tested. The influence of evaluative categories does not seem to function as a simple, direct input of affective information, however, given Study 3's memory results. Instead, the affect from the manipulation appears to function to spotlight some information (and not other) or to motivate behaviors. The present set of studies cannot differentiate between these latter functions of affect.

In addition, if we are correct that the influence of evaluative categories is due at least in part to an affective mechanism, then Study 1's results provide some initial evidence that making information more affective can trigger better judgments. It should be noted that, although affective information may be easier to integrate, it does not necessarily lead to better decisions (Shiv, Loewenstein, Bechara, Damasio, & Damasio, 2005). Whether affect is helpful or harmful to decision quality will depend on the situation. In Study 1, we claim it is helpful because subjects integrated both more numeric information as well as more important numeric information (according to experts) into their judgments in the presence of evaluative categories.

The results of Study 4 also support the notion that our thoughts and feelings are somewhat separable; we speculate that in the type of problems studied here, we may be more "expert" with our feelings leading to their greater consistency and association with choice (Shanteau, 1992). The results also suggest that the detrimental impact that analysis of reasons sometimes has on choice satisfaction and quality may occur because our thoughts conflict more with one another than do our feelings (e.g., Wilson et al., 1993). As a result, any thoughts are more likely to be in opposition to one another than are feelings.

Conclusions

Numbers appear to be just that—numbers. The present results suggest that consumers may not use them in decision making until available data are compared and contrasted to determine their affective meaning for choice or until the data acquire meaning through other means such as evaluative categories. These results are consistent with the constructed preferences approach in decision making and imply that information providers cannot present "just the facts." The two methods of data presentation (i.e., the presence and the absence of evaluative categories) in the current studies had different influences on choice. Neither method was clearly superior in the choice studies because no option was clearly best. In Study 1, however, the presence of evaluative categories led to a greater integration of important attributes in judgments and less influence of irrelevant mood states. In addition, Study 3's memory results suggest that health-plan costs, which some experts believe are currently over weighted, will be weighted less if quality

attributes are made easier to evaluate. Whatever method of presentation information providers choose will influence how decision makers find meaning in numbers and construct choices in unfamiliar domains.

These results imply that decision makers need help in interpreting not only what the numbers are but what they mean, and this meaning is tied to affect at least in the present context. Information providers would need to take responsibility for and make decisions about the meaning of the numbers they provide. This process is likely to be difficult and rife with ethical and political concerns (e.g., what scores are fair vs. good). However, making numerical information easier to evaluate may increase use of information and better assist decision makers in making high quality choices in health, financial, and other domains (Hibbard & Peters, 2003). This approach is consistent with the important notion of "choice architecture" proposed by Thaler and Sunstein (2008). Without assistance, today's world of instantaneous information could lead to a greater influence of irrelevant sources of affect and emotion, especially for the less numerate among us.

These findings have implications for the success of current health-care policy aimed at controlling costs and improving quality. This policy rests, to a large extent, on consumers' and patients' ability to make informed choices. Information about treatments, providers, health plans, nursing homes, and hospitals is being made available to a degree never before possible in an effort to help patients and consumers increase control over their health-care experiences and health outcomes. At the same time, health and financial choices can be difficult and choices involving numbers may be undermined by a lack of numeric ability. The presentation of simple numbers (a common phenomenon in our world of instantaneous data and informed choice) appears to be much more complex than might be thought at first glance.

References

- Alexander, M. G., Brewer, M. B., & Herrmann, R. K. (1999). Images and affect: A functional analysis of out-group stereotypes. *Journal of Personality and Social Psychology*, 77, 78–93.
- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison Wesley.
- Bateman, I. J., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavior Decision Making*, 20, 365–380. doi: 10.1002/bdm.558
- Birnbaum, M. H. (1999). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods*, 4, 243–249.
- Cabanac, M. (1992). Pleasure: The common currency. *Journal of Theoretical Biology*, 155, 173–200.
- Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality & Social Psychology Bulletin*, 25, 215–224. doi: 10.1177/0146167299025002007
- Chernew, M., & Scanlon, D. P. (1998). Health plan report cards and insurance choice. *Inquiry*, 35, 9–22. doi: 10.1016/S0167-6296(02)00075-9
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Avon.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.
- Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making*, 25, 398–405. doi: 10.1177/0272989X05278931
- Fazio, R. H. (2000). Accessible attitudes as tools for object appraisal: Their costs and benefits. In G. R. Maio & J. M. Olson (Eds.), *Why we evaluate: Functions of attitudes* (pp. 1–36). Mahwah, NJ: Erlbaum.
- Fazio, R. H., & Hilden, L. E. (2001). Emotional reactions to a seemingly prejudiced response: The role of automatically activated racial attitudes and motivation to control prejudiced reactions. *Personality & Social Psychology Bulletin*, 27, 538–549. doi: 10.1177/0146167201275003
- Fischhoff, B. (2009). Risk perception and communication. In R. Detels, R. Beaglehole, M. A. Lansang, & M. Gulliford (Eds.), *Oxford textbook of public health* (5th ed.; pp. 940–952). Oxford, England: Oxford University Press.
- Frijda, N. H. (1986). *The emotions: Studies in emotion and social interaction*. New York, NY: Cambridge University Press.
- Gregory, R., Lichtenstein, S., & Slovic, P. (1993). Valuing environmental resources: A constructive approach. *Journal of Risk and Uncertainty*, 7, 177–197. doi: 10.1007/BFO1065813
- Hibbard, J. H., & Peters, E. (2003). Supporting informed consumer health care decisions: Data presentation approaches that facilitate the use of information in choice. *Annual Review of Public Health*, 24, 413–433. doi: 10.1146/annurev.publhealth.24.100901.141005
- Hsee, C. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior & Human Decision Processes*, 67, 247–257. doi: 10.1006/obhd.1996.0077
- Hsee, C. K., Blount, S., Loewenstein, G. F., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125, 576–590.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720. doi: 10.1037/0003-066X.58.9.697
- Kahneman, D., Schkade, D., & Sunstein, C. R. (1998). Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty*, 16, 49–86.
- Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. New York, NY: Cambridge University Press.
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21, 37–44. doi: 10.1177/0272989X010210010
- Loewenstein, G., Weber, E. U., Hsee, C. K., & Welch, E. S. (2001). Risk as feelings. *Psychological Bulletin*, 127, 267–286.
- Montague, R. P., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36, 265–284. doi: 10.1016/S0896-6273(02)00974-1
- Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*. New York, NY: Hill and Wang.
- Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring constructed preferences: Towards a building code. *Journal of Risk and Uncertainty*, 19, 243–270.
- Peters, E. (2006). The functions of affect in the construction of preferences. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preference* (pp. 454–463). New York, NY: Cambridge University Press.
- Peters, E., Diekmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, 64, 169–190. doi: 10.1177/10775587070640020301
- Peters, E., Hess, T. M., Västfjäll, D., & Auman, C. (2007). Adult age differences in dual information processes: Implications for the role of affective and deliberative processes in older adults' decision making. *Perspectives on Psychological Science*, 2, 1–23. doi: 10.1111/j.1745-6916.2007.00025.x

- Peters, E., Hibbard, J. H., Slovic, P., & Dieckmann, N. F. (2007). Numeracy skill and the communication, comprehension, and use of risk and benefit information. *Health Affairs*, 26, 741–748. doi: 10.1377/hlthaff.26.3.741
- Peters, E., Lipkus, I., & Diefenbach, M. A. (2006). The functions of affect in health communications and in the construction of health preferences. *Journal of Communication*, 56, S140–S162. doi: 10.1111/j.1460-2466.2006.00287.x
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17, 407–413. doi: 10.1111/j.1467-9280.2006.01720.x
- Rothbart, M., Davis-Stitt, C., & Hill, J. (1997). Effects of arbitrarily placed category boundaries on similarity judgments. *Journal of Experimental Social Psychology*, 33, 122–145. doi: 10.1006/jesp.1996.1315
- Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, 14, 294–301.
- Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, 81, 75–86.
- Shiv, B., Loewenstein, G., Bechara, A., Damasio, A. R., & Damasio, A. (2005). Investment behavior and the dark side of emotion. *Psychological Science*, 16, 435–439.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York, NY: Cambridge University Press.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, 24, 1–12. doi: 10.1111/j.0272-4332.2004.00433.x
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- Trafimow, D., & Sheeran, P. (2004). A theory about the translation of cognition into affect and behavior. In G. R. Maio & G. Haddock (Eds.), *Contemporary perspectives in the psychology of attitudes: The Cardiff symposium* (pp. 57–75). London, England: Psychology Press.
- Verplanken, B., Hofstee, G., & Janssen, H. J. W. (1998). Accessibility of affective versus cognitive components of attitudes. *European Journal of Social Psychology*, 28, 23–36. doi: 10.1002/(SICI)1099-0992(199801/02)28:1<23::AID-EJSP843>3.0.CO;2-Z
- Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, 18, 516–523. doi: 10.1111/j.1467-9280.2007.01932.x
- Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin*, 19, 331–339. doi: 10.1177/0146167293193010
- Woloshin, S., Schwartz, L. M., Black, W. C., & Welch, H. G. (1999). Women's perceptions of breast cancer risk: How you ask matters. *Medical Decision Making*, 19, 221–229. doi: 10.1177/0272989X9901900301
- Zikmund-Fisher, B., Fagerlin, A., Keeton, K., & Ubel, P. (2007). Does labeling prenatal screening test results as negative or positive affect a woman's responses? *American Journal of Obstetrics and Gynecology*, 197, 528.e1–528.e6.

Received July 25, 2008

Revision received May 5, 2009

Accepted May 5, 2009 ■