Chapter 18

COGNITIVE PROCESSES IN STATED PREFERENCE METHODS

BARUCH FISCHHOFF

Department of Social and Decision Sciences, Department of Engineering and Public Policy,
Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Contents

Handbook of Environmental Economics, Volume 2. Edited by K.-G. Mäler and J.R. Vincent

DOI: 10.1016/S1574-0099(05)02018-8

Abstract

Cognitive psychology is best known, to many environmental economists, through the filter of acrimonious debates over the validity of contingent valuation methods (CVM). Psychologists' views on CVM reflect concerns that are deeply rooted in their profession's history and theories. Although psychologists have participated in some CVM studies, their roles have rarely allowed them to present a comprehensive design philosophy, illustrated in actual studies. This chapter sets psychologists' critiques and alternatives within a general cognitive perspective on value elicitation, including stated preferences for environmental goods. It begins with a historical review, organized around two converging streams of psychological research. One stream leads from psychophysics to attitude research. The second leads from decision theory to decision analysis and behavioral decision research. The next section reports some environmental valuation studies arising from each tradition. These studies do not directly monetize environmental goods. However, they can still directly inform policies that do not require monetization and indirectly inform policies that do, by shaping studies with that ambition. The following section considers the role of cognitive studies in helping investigators to know what issues matter to people and present them comprehensibly. The concluding section of the chapter presents a cognitive approach to stated preference methods for environmental values – one that could be developed most fully in collaboration with economists. It is built around a cognitive task analysis of the four main elements in any evaluation process: (a) specifying the valuation question, (b) understanding its terms, (c) articulating a value for that specific question (from more general basic values), and (d) expressing that value in a public form.

Keywords

environment, preferences, elicitation, values, uncertainty

JEL classification: Q51, C9

1. Overview

Cognitive psychology is best known, to many environmental economists, through the filter of acrimonious debates over the validity of contingent valuation methods (CVM) [Arrow et al. (1993), Cummings, Brookshire and Schulze (1986), Mitchell and Carson (1989, Chapter 11)]. Psychologists' views on CVM reflect concerns that are deeply rooted in their profession's history and theories. However, expressing these views through the filter of controversy precludes systematic exposition of the research motivating them. It also makes psychologists seem like unrelenting critics, unconcerned with the critical environmental policy issues that CVM research directly addresses, whatever its strengths and weaknesses.

Although psychologists have participated in some CVM studies, their roles have rarely allowed them to present a comprehensive design philosophy, illustrated in actual studies. Psychologists' own CVM-related studies often seem like destructive exercises, attempting to undermine CVM, without concern for the vital need that it addresses. For example, many experiments [Kahneman, Diener and Schwarz (1999)] are presented as demonstrating respondents' insensitivity to the "scope" of the environmental good being evaluated (e.g., the number of birds, or lakes, protected). That would represent a fundamental failure of measurement [which should show sensitivity to relevant features of the good being evaluated, and insensitivity to irrelevant features; Fischhoff (1988b)]. Demonstrating problems does not, however, provide alternative, putatively better methods for producing the quantitative valuations needed for cost–benefit analyses and compensatory damage assessments. It does not even clarify what policy makers should do if they accept these claims. In the absence of an alternative, rejecting CVM leaves a vacuum. Decisions still need to be made. Without accepted valuations, environmental goods may be neglected. As a result, psychologists might be seen as showing a radical skepticism toward CVM that embodies radical indifference toward the environment.

This chapter sets psychologists' critiques and alternatives within a general cognitive perspective on value elicitation, including stated preferences for environmental goods. It begins with a historical review, organized around two converging streams of psychological research. One stream leads from psychophysics to attitude research. The second leads from decision theory to decision analysis and behavioral decision research. The next section reports some environmental valuation studies arising from each tradition. These studies do not directly monetize environmental goods. However, they can still directly inform policies that do not require monetization and indirectly inform policies that do, by shaping studies with that ambition. The following section considers the role of cognitive studies in helping investigators to know what issues matter to people and present them comprehensibly. The concluding section of the chapter presents a cognitive approach to stated preference methods for environmental values – one that could be developed most fully in collaboration with economists. It is built around a cognitive task analysis of the four main elements in any evaluation process: (a) specifying the valuation question, (b) understanding its terms, (c) articulating a value for that specific question (from more general basic values), and (d) expressing that value in a public form.

2. Origins

Cognitive psychology's approach to stated preference methods draws on several research traditions. To a first approximation, these can be described as arising from psychophysics and decision theory. Each stream characterizes responses by comparing them with a normative standard, then looking for the psychological processes shaping (and limiting) performance.

2.1. Psychophysics stream

2.1.1. History

Evaluation has been a central topic in experimental psychology, since its inception in the mid-to-late 1800s. Emerging from the natural sciences, early psychologists focused on *psychophysics*, determining the sensory equivalent of physical stimuli. According to the simple model underlying these studies, various physiological and psychological mechanisms translate external stimuli into states of arousal. If asked, individuals can report those states, with a word, number, or action (e.g., squeezing a handgrip or adjusting an illumination level to equal the experience of a tone). Those reports could reflect a subjective magnitude (e.g., loudness) or valuation (e.g., pleasantness).

Over the intervening century-plus, researchers have discovered the complexity of these ostensibly simple processes. One family of complications arises from people's sensitivity to seemingly irrelevant procedural features. For example, a detection threshold might depend on whether successive stimuli (e.g., tones, weight differences, figures on a background) are presented in ascending or descending order [e.g., Woodworth and Schlosberg (1954)]. Table 1 collects some of these effects. Their discovery has often led to full-fledged research programs, examining their detailed operation. McGuire (1969) characterized this process as converting an *artifact* into a *main effect*. A classic example is tracing inconsistent celestial observations to differences in the reaction time of astronomers' assistants. Subsequent research identified underlying processes that are important in their own right (e.g., nerve conductance, concurrent distractions, speed-accuracy tradeoffs). Another family of productive artifacts involves the subtle ways that interviewers communicate their expectations. The methodological challenges of controlling these (often unwitting) cues has informed basic research into nonverbal communication.

Some of these effects primarily interest specialists in the relevant psychological or sensory systems. For example, an important auditory ability is accommodating to low ambient sound levels, making it easier to detect weak signals. (Indeed, under extremely quiet conditions, most people eventually pick up nonexistent sounds, even experiencing tinnitus in sound-proof rooms.) These effects should matter to stated preference researchers who must present auditory stimuli (e.g., in studies evaluating noise pollution levels). Similarly, the nuances of color and form discrimination are central to eliciting valuations for atmospheric visibility levels [e.g., Tolley et al. (1986)]. Those studies

Table 1
From artifact to main effect

Lability in judgment due to	Led to
Organism	
Inattention, laziness, fatigue, habituation, learning, maturation, physiological limitations, natural rhythms, experience with related tasks	Repeated measures Professional subjects Stochastic response models Psychophysiology Proactive and retroactive inhibition research
Stimulus presentation	
Homogeneity of alternatives, similarity of successive alternatives (especially first and second), speed of presentation, amount of information, range of alternatives, place in range of first alternative, distance from threshold, order of presentation, areal extent, ascending or descending series	Classic psychophysical methods The new psychophysics Attention research Range-frequency theory Order-effects research Regression effects Anticipation
Response mode	
Stimulus–response compatibility, naturalness of response, set, number of categories, halo effects, anchoring, very small numbers, response category labeling, use of end points	Ergonomics research Set research Attitude measurement Assessment techniques Contrasts of between- and within-subject design Response-bias research Use of blank trials
"Irrelevant" context effects	
Perceptual defenses, experimenter cues, social pressures, presuppositions, implicit payoffs, social desirability, confusing instructions, response norms, response priming, stereotypic responses, second-guessing	New look in perception Verbal conditioning Experimenter demand Signal-detection theory Social pressure, comparison, and facilitation research

Source: Fischhoff, Slovic and Lichtenstein (1980).

need visual displays that accurately represent the degree of light extinction associated with alternative policies (e.g., emission standards). The basic science of psychophysics can guide such designs (and determine the validity of previously conducted studies).

2.1.2. Design framework

In addition to effects that are specific to sensory modalities, psychophysics research has identified general effects, found in many studies. One large set pertains to the use of

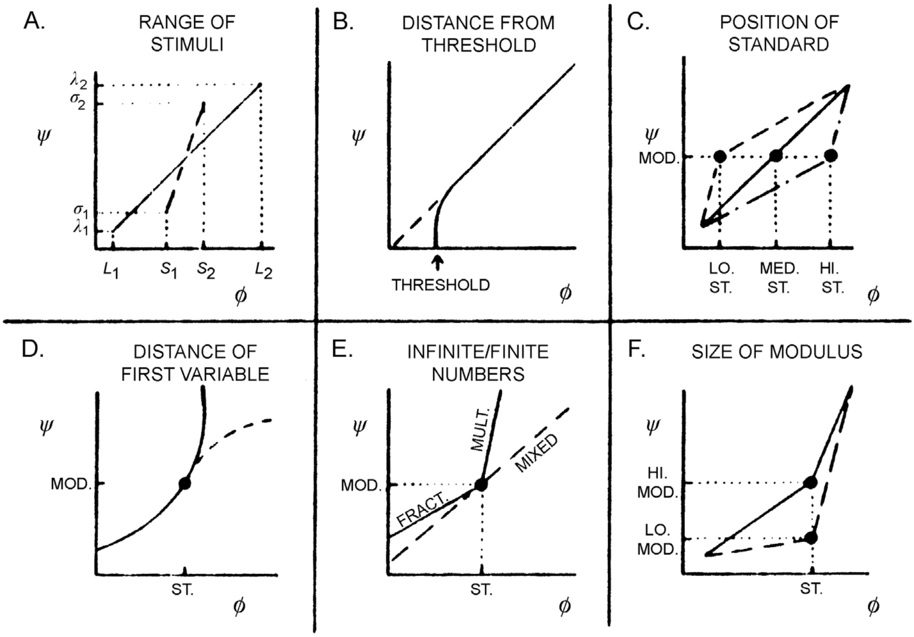

Figure 1. Six "laws of the new psychophysics," depicting the influence of experimental design on the numerical response used to describe the psychological state (ψ) equivalent to a physical stimulus (ϕ). Figure A shows that a narrower stimulus range (S_1, S_2) will use a proportionately larger portion of the response range than would the same stimuli, when embedded in a larger response range (L_1, L_2). Figure B considers the effects of assumptions regarding the treatment of stimuli below the threshold of perception or evaluation. Figure C considers the effects of where a standard stimulus falls in the response range, after it has been assigned a numerical valuation (or modulus). Figure F shows the reverse effects where a modulus value, for a given standard stimulus, falls within the response range. Figure D shows the effects of where the first judged stimulus is relative to the standard. Figure E shows the effects of using fractional or integer response values, for stimuli smaller than the standard. Fuller details are found in Poulton (1968), from which this exhibit was taken; they are elaborated and extended in Poulton (1989, 1994).

numbers – both ones that people produce by themselves and ones proposed by others ("Is that a 10?"). Figure 1 shows a summary, derived by Poulton (1968) from secondary analyses of research in the paradigm of Stevens (1975), perhaps the central figure in mid-20th century psychophysics. Stevens and colleagues attempted to estimate the rate of decreasing marginal sensitivity ("the shape of the Weber–Fechner law," to psychologists), for many sensory modalities. Based on the shapes of the curves, investigators proposed theories about the physiology of these modalities. Poulton argued, however, that these curves were strongly influenced by the methodological conventions of researchers studying each modality. These conventions induced consistent results among studies on a single modality, while limiting comparisons across modalities.

The figures summarize these artifacts, showing how responses are shaped by design features that investigators must somehow set, when creating tasks. For example, Figure 1A shows how a narrower stimulus range $[S_1, S_2]$ elicits a proportionately larger portion of the response range than do the same stimuli, embedded in a larger response range $[L_1, L_2]$. Figure 1C shows how values assigned to larger stimuli are cramped if the initial (or standard) stimulus is relatively large (and spread out if it is relatively small).

Such effects occur because, however well respondents understand their feeling regarding a stimulus, they must translate it into the investigators' terms. Where those terms differ from their natural mode of expression (or where there is none, because the task is so novel), respondents rely on general heuristics for using response scales. Because these *response preferences* are often widely shared [Tune (1964)], they lend an artifactual consistency to results. For example, other things being equal, people try to use the entire response scale; they look for patterns in randomly ordered stimuli; they deduce an expected level of precision, such as the tradeoff to make between speed and accuracy. Thus, their responses are shaped by a task's form as well as by its substance, when they are unsure what to say – and must seek clues in task details.

The importance of these effects depends on how responses are used. For example, the patterns in Figure 1 all maintain response order. Investigators requiring just ordinal information can ignore these design issues. When investigators require stronger metric properties, they must match study conditions to those in the analogous real-world context. For example, they should avoid fractional responses, if people rarely use them. They should show the full stimulus range, so that respondents need not infer it from the initial stimuli. If there is no (single) natural context, then investigators can use relationships like those in Figure 1 to extrapolate from a context that has been studied to other ones. For example, if a study used a restricted stimulus range, then one can assume that its largest members would receive lower values, if embedded in a larger range. Conversely, if a study uses a narrow range, one should expect respondents to make finer distinctions and spread out their responses, relative to a real-world situation with a narrower focus.

Psychophysical considerations complicate research design, by adding features that require explicit attention. However, they also simplify it, by providing an orderly, empirically based foundation for focusing on the (perhaps few) features that really matter when designing and interpreting studies. If people are, indeed, sensitive to these features, then (in the words of an American auto repair commercial) it is "pay now" (by explicitly addressing the features) or "pay later" (by having to disentangle their effects).

2.1.3. Evaluative extensions

As attitude research developed, especially during the second quarter of the past century, the psychophysical paradigm held great appeal, for attitude researchers. It had a distinguished pedigree, with some of psychology's most distinguished researchers having contributed theories, procedures, results, and analytical methods. It had been applied to diverse domains (e.g., the intensity of smell, cold, weight). Its method seemed relatively straightforward: pose a clear question; offer clear response options; report the responses. Moreover, attitude questions seem to avoid the response-mode ambiguity arising with physical stimuli having unfamiliar measures (e.g., lumens, decibels). Attitude scales can use terms that seem objective and consensually interpreted (e.g., anchors like "strongly agree" and "strongly disagree," or "very happy" and "very unhappy").

Unfortunately, such scales have often proven less objective and consensual than was hoped. It turns out that "there's 'very happy' and then there's 'very happy' " [Turner and Krauss (1978)]. Here, too, respondents seek contextual cues for interpreting their task. These cues might be found in a question's wording, the topics preceding it, the study's sponsor, or the consent form, among other places. Conversely, relatively clear response scales may be used to disambiguate unclear questions. As a result, respondents may attach different meanings to both the question and the answer of an apparently standard task [Budescu and Wallsten (1995), Fischhoff (1994), Tanur (1992)].

Schwarz (1999) offers an example with an 11-point scale, anchored at "not at all successful" and "extremely successful." Respondents' evaluations of their lives differ when these anchors were assigned the numbers [0, 10] or [−5, +5]. Apparently, negative numbers suggest the possibility of abject failure – whereas 0 suggests lack of success, as the worst possible outcome. Other contextual cues can signal whether to evaluate success relative to a comparison group (and, if so, which one) or relative to one's own dreams (or to one's realistic expectations). People can shift perspectives in the research context, just as they do in everyday life, as they try to determine (in the words of a former New York mayor), "How'm I doing?"

Such artifacts have long been known in the survey research literature. Their prevalence depends on the inferential task imposed on respondents [Fischhoff (1991), National Research Council (1982), Schwarz (1996), Turner and Martin (1984)]. Unambiguous questions and answers reduce the need for contextual cues. So does posing tasks for which respondents already have *articulated* values, in the sense of knowing what they want and how to express themselves. With a novel question, respondents must derive their preferences from the relevant *basic* values – just as they would when facing unfamiliar choices in everyday life. Table 2 offers conditions favorable to having articulated values – and, therefore, to stated preferences that are resistant to meaningless procedural and wording changes (and sensitive to meaningful ones).

According to Poulton (1968), Figure 1 captures the six key sources of interpretative cues for quantitative judgment tasks. No such summary is available, or perhaps even possible, for the myriad of verbal cues that people might use, when they read between the lines of verbal questions and answers – in surveys or everyday life. Creating a manageable science of task construal requires treating surveys as conversations. An increasingly influential approach assumes that respondents interpret survey questions as though the investigator has adhered to the conversational norms of their common culture, and that they will respond in kind (expecting their answers to be similarly interpreted). Those norms provide a structured way for investigators to analyze respondents' construal of questions and answers [Schwarz (1996, 1999)].

Table 2
Conditions favorable to respondents having articulated values

Aspects of the topic	Aspects of the consequences	Aspects of the respondent
Familiar issues Familiar formulation Publicly discussed – providing opportunities to hear and share views Stable – allowing the formation of tastes	Few – providing simplicity Similar – providing commensurability Previously experienced – allowing the formation of tastes Certain – reducing complexity	Able to play a single role – avoiding agency issues Able to generate alternative perspectives Motivated to consider issues Able to consider topic in isolation

Source: Fischhoff (1991).

A prominent summary of conversational norms is Grice's (1975) four maxims, which require speakers to be *relevant*, *concise*, *complete*, and *honest*. An investigator following these norms would ask whether a question–answer pair contains irrelevant details, says things that could go without saying (thereby lacking conciseness), omits things that respondents could not confidently infer, or misleads respondents (or even gives the appearance of dishonesty). The more familiar the topic and the respondents, the more confidently investigators can make these evaluations. If they need evidence to support their claims, it could come from dedicated studies (e.g., think-aloud protocols) or general studies – both considered below [Fischhoff, Welch and Frederick (1999)]. The return on that investment is reduced uncertainty about whether tasks were interpreted as intended.

2.2. Decision theory stream

Von Neumann and Morgenstern's (1947) landmark work prompted two lines of psychological research (often with collaborators from other disciplines) [Edwards (1954, 1961), Yates (1990)]. One line, *behavioral decision research*, studies the descriptive validity of the decision theoretic axioms [Dawes (1988), Fischhoff (1988a), Kahneman and Tversky (1984), McFadden (1999)]. These studies may directly test people's adherence to the axioms, characterize the cognitive skills that facilitate (and constrain) rationality, or identify behaviorally realistic approaches to decision making (e.g., reliance on heuristics). One general (and perhaps unsurprising) result is that people do best when they have conditions conducive to acquiring decision-making abilities as learned skills. These conditions include prompt, unambiguous feedback that rewards rationality (rather than, say, bravado or evasiveness). People also need enough feedforward to make the desired behaviors part of their repertoire. It is hard to master unheard-of or counterintuitive principles.

The second line of research, *decision analysis*, seeks to increase the descriptive validity of the axioms, by helping people to make more rational choices [Clemen (1996), Raiffa (1968), von Winterfeldt and Edwards (1986)]. It uses behavioral decision research to identify where people need help in achieving the performance standard set by the normative axioms. Decision analysts elicit individuals' probabilities and utilities for possible consequences of their action options. They hope to overcome their clients' known judgmental limitations in three ways: (a) Structuring the elicitation process, so that people make systematic use of what they know about themselves and their world; (b) Focusing people's limited attention on the most critical issues, as identified by procedures such as sensitivity analysis and value-of-information analysis; (c) Computing the expected utility of options, thereby avoiding the vagaries and omissions of mental arithmetic.

The success of this enterprise hinges on participants' ability to express their beliefs and values in the required form, probabilities and utilities. The conceptually clear structure of decision analytic questions should reduce the ambiguity endemic to attitude measurement. So should its reliance on explicit probability and utility scales, avoiding the well-documented vagueness of verbal quantifiers (e.g., rare, likely, possible), terms that can mean different things to different people in a single context and to a single person in different contexts [Lichtenstein and Newman (1967), Budescu and Wallsten (1995)]. The interaction between client and decision analyst constitutes an actual conversation (albeit a somewhat stilted one). That provides opportunities to identify and resolve residual miscommunications, beyond what is possible with the generic conversations of standardized surveys. In order to exploit these opportunities, decision analysts "look for trouble," presenting tasks from multiple perspectives, in order to ensure mutual understanding [Fischhoff (1980), Keeney (1996), von Winterfeldt and Edwards (1986)].

As a result, in decision analysis, elicitation is a *reactive* measurement procedure: It can change participants, as they reflect on their beliefs and values. That process should deepen their thinking, without imposing the elicitors' perspectives. That is, it should reduce random error, without adding systematic error. The process assumes that people sometimes need help, in order to understand what they believe and want. That help may include presenting a balanced selection of opinions, lest clients miss a critical perspective just because it did not occur to them at the time.

Such concern for completeness and reflection contrasts radically with the nonreactivity of psychophysical research. Investigators there seek immediate, "natural" responses to task stimuli. The procedure is seen as a neutral conduit for those responses. As a result, psychophysical studies have impassive interviewers, standardized questions, no clarification (beyond perhaps scripted paraphrasing), and limited time to think.

These two elicitation philosophies place different weights on sins of commission (inappropriately influencing respondents) and sins of omission (leaving respondents to misunderstand the issue or their position on it). Fearing the former discourages interaction with respondents; fearing the latter may require it. The degree of interaction should depend on the risk of respondents saying, subsequently, "I wish I had thought of that perspective," "I wish I had been able to keep it all in my head," or "How could I have forgotten that?" That risk should be small with tasks fulfilling the conditions of Table 2.

For example, people may know just what they want, after a vigorous political campaign thoroughly airs the issues. With complex, novel questions, though, people may need a chance to think and hear others' thinking, before they can formulate stable, independent views.

Behavioral decision research also provides alternative perspectives, for helping respondents triangulate on their values. For example, the framing effects of prospect theory [Kahneman and Tversky (1979, 1984)] suggest views that might not otherwise occur to them (e.g., "be sure to think about both the number of salmon that will be left and the number that will be lost"). Contrasting perspectives can be derived from any context effect from psychophysics research (Table 1, Figure 1). For example, making a number salient will bring responses toward it – even when that *anchor* is clearly chosen arbitrarily [e.g., Tversky and Kahneman (1974)]. Such anchoring (which partially underlies Figure 1's effects) can work in two (nonexclusive) ways: (a) directly "priming" that specific response, making it more available when people look for what to say, and (b) prompting respondents to evaluate that number's appropriateness, a search that disproportionately primes reasons justifying it. Investigators fearing these effects can provide multiple anchors (as a way to provide multiple perspectives) or try to avoid suggesting any number.

Eliciting consistent responses with a single procedure is necessary but not sufficient for demonstrating articulated values. Such consistency may just mean that respondents devised an efficient ad hoc strategy in order to get through their task (e.g., "I see that they are varying the price of the fictitious student apartments 'offered' in this study. I'll focus on that feature, since it's so easy to do."). Even if such strategies reflect real concerns (e.g., price does matter), the resulting consistency could fragment in more realistic contexts. Part of psychology's lore is how easily people find some way to answer any question that researchers pose, as seen in the low nonresponse rates to questions about fictitious issues [Plous (1993)]. Successful elicitation should capture the residual uncertainty and incoherence in respondents' preferences.

Behavioral decision research has also shaped value elicitation by identifying the limits to judgment under conditions of uncertainty [Dawes (1988), Gilovich, Griffin and Kahneman (2002), Kahneman, Slovic and Tversky (1982)]. With uncertain environmental choices, those limits require a special effort to ensure that respondents understand the facts of their choices [Fischhoff (1998, 1999), Slovic (1987)]. Assessing that understanding means eliciting respondents' uncertain beliefs. Generally speaking, that is straightforward: just use clear probability scales and well-defined events [Slovic and Lichtenstein (1971), Wallsten and Budescu (1983), Yates (1990)]. However, there are still anomalies that complicate interpreting expressed beliefs. One is that people sometimes use "50" to express *epistemic* uncertainty (not knowing what to say), rather than a numeric probability. As a result, they seem to overestimate small probabilities, when they are not actually giving a number. Such scale misuse is less common with structured response scales, controllable events, and children (until they learn the phrase "fifty–fifty") [Fischhoff and Bruine de Bruin (1999)]. A second anomaly arises from people not appreciating how small risks (or benefits) mount up through repeated exposure (or even that a long-term perspective is warranted). As a result, one cannot infer beliefs about single and multiple events from one another.

Less is known about communicating the events that the probabilities amplify [Fischhoff (1994), Fischhoff, Welch and Frederick (1999)]. As mentioned, events have many more possible cues for events than do probabilities, hence require much messier scientific accounts: Grice's maxims, rather than Poulton's laws. Furthermore, there is rather less research on substantive aspects of task descriptions, compared to quantitative ones. That imbalance reflects both personal predilections and the limited success of attempts to determine broadly relevant values [e.g., Rokeach (1973)].

2.3. Confluence

Two points of confluence might clarify the interrelationship of these research streams, as well as the balance sought by those eliciting values within each.

(a) Psychophysicists historically have sought immediate responses to stimuli, for comparison with the corresponding physical measurements. However, some complex stimuli are not easily summarized. Researchers needing a standard for comparison have sometimes trained people to be *human meters*, synthesizing sounds, lights, smells, pressures, or tastes. The training attempts to accelerate them along the path that others will take, as they experience a new auditorium (or chocolate or wine). A similar logic guides citizen juries, which try to determine public values where none exists, by simulating the learning process of a well-conducted public inquiry [Guston (1999)]. These procedures focus on participants' summary judgments, taking accompanying explanations with a grain of salt. Cognitive psychology has long been wary of introspective accounts, fearing that they reflect people's intuitive theories about cognitive processes, rather than their actual processes. Concurrent verbal reports, or think-aloud protocols, have greater credibility, as ways to catch thinking in the raw [Ericsson and Simon (1994), Nisbett and Wilson (1977)].

(b) Decision theory provided psychophysicists with a systematic approach to an old question: how do uncertain respondents decide what to say? (Did I really hear something? Is it really sweeter? Am I actually 'very happy'? Am I 70% or 80% sure?) The *theory of signal detection* (TSD) distinguishes between how well individuals discriminate among different states and how willing they are to risk different errors when expressing those feelings [Green and Swets (1966)]. The former reflects their evaluative (or diagnostic) ability, the latter their incentives. TSD produced new experimental designs, and disentangled seemingly inconsistent results. It revealed better ways to train "human meters" (e.g., as interpreters of complex medical images [Swets (1992)]). It increased psychologists' sensitivity to incentives, while revealing the difficulties of conveying them. For example, response time often varies greatly with small differences in error rate. Yet, it is hard to tell someone how hard to work in order to achieve a 1% vs. a 2% error rate, or to demonstrate the difference without providing feedback on a very large sample of behavior [Wickelgren (1977)].

3. Evaluating environmental changes

3.1. Psychophysics stream

Contemporary attitude research generally assumes a dual-process model of evaluation [Eagly and Chaiken (1993), Fiske (1998)]. One process involves more global, superficial, affective, connotative evaluations, as people ask themselves, "How do I feel about this, in general?" The second process involves more specific, detailed, cognitive, denotative evaluations, asking, "How much is it worth to me?" The processes are linked, so that evaluations at one level still activate the other. For example, an ostensibly simple request like, "Give me your first impression" may evoke more detailed questions like "What do they want from me?" and "What will they think about me, if I say . . . ?" Conversely, detailed questions can evoke powerful emotions, confounding the desire (or request) to be "rational." Research on dual-process models examines such topics as: What conditions evoke each mode of processing? How do the processes interact? What do people extract from stimuli, when directed to one level or the other?, and How do individuals differ in their general propensity to seek details (sometimes called their "need for cognition")?

Psychophysical studies of environmental values can address either level. Affect-level studies ask people how they feel about some environment: facing them, depicted, or just imagined. Cognitive-level studies ask for more explicit evaluations (e.g., how natural, attractive, rich . . . is it?) [e.g., Daniel and Meitner (2001)]. Contingent valuation falls in the latter category. As mentioned, the two processes can intrude on one another. For example, cognitive tasks can make people mad (as seen in the "protest responses" of some stated preference studies), whereas anger can engage their thinking. Environments judged more "natural" can make people feel better [Frumkin (2001)], heal faster during hospitalization [Ulrich (1984)], and require less healthcare [Moore (1981)]. The fact that researchers had to discover these effects shows a limit to introspection as a guide to environmental valuation [Nisbett and Wilson (1977)]: Individuals may not appreciate these "environmental services," just as they may not realize the negative impacts of seasonal affective disorder. Without such understanding, environmental valuations are incompletely informed.

A research strategy that addresses both attitude levels considers the features that people notice in environmental stimuli. For example, respondents might be asked to sort, by similarity or preference, pictures of scenes that vary in the roles of natural and built objects [Cantor (1977), Kelly (1955)]. Using pictures frees respondents from having to verbalize what they see and like. As a result, the process may capture attitudes for which people lack the right words or hesitate to use them (e.g., because of social unacceptability or uncertainty about terminology). One set of recurrent properties, identified in such studies, is *coherence*, *complexity*, *legibility*, and *mystery* [Kaplan and Kaplan (1989)].

Although this research strategy frees respondents from verbalization, it increases investigators' interpretative burden – and freedom. That is, investigators can – and must – decide: Just what were respondents looking at? Was their attention drawn to a feature that happened to vary a lot in the picture set, exaggerating its importance? How do respondents derive uncertainty from a concrete scene? Analogous interpretative concerns arise with other psychophysical procedures, such as direct ratings [e.g., the Scenic Beauty Estimation procedure; Daniel and Vining (1983)] or similarity judgments, subjected to multi-dimensional scaling [Berlyne (1971), Gärling (1976)]. These are the attitude-research equivalent of the specification problems facing revealed preference studies. Researchers must determine which of the (often correlated) features of complex situations best account for behavioral regularities [Dawes and Corrigan (1974), Leamer (1983)]. The many degrees of interpretative freedom mean that each study must be viewed in the context of related ones, using varied methods and stimuli – with only recurrent patterns being trusted.

Whatever their limits, these studies place important constraints on ones using other procedures. Features that play a robust role in attitude studies should be represented in stimuli used with other methods. Otherwise, respondents must infer the missing features from available ones. Even when respondents guess correctly, fragmenting their experience may prevent them from articulating values that integrate cognitive and affective responses. Environments have emergent properties, only partially captured by pictures, much less by verbal or vector-like descriptions of their characteristics [Brunswick (1947), Gibson (1979), Larkin and Simon (1987)]. Respondents naturally seek guidance in the investigators' choice of features ("If they think that it's important, so should I"), and overlook neglected ones.

The preferences observed in attitude studies should also be seen in research using other stated preference methods. If not, then one must question the robustness of either the studies or the preferences. As with revealed preferences, interpreting individual stated preference studies should consider the entire literature. For example, some studies have found an aversion to scenes that are hard to understand. That aversion could reduce the value assigned to environments depicted in cluttered pictures, relative to the value of the actual settings, where people could get their bearings (and understand them). If that "clutter" is necessary for ecosystem health, then using pictures will undervalue healthy ecosystems and overvalue manicured ones. Similarly, unless respondents understand what they are seeing, they may assign equal value to water cleared by reducing algae blooms and by introduced zebra mussels (or they may love pollution-induced red sunsets). Without background information, respondents may not realize the complexity of healthy environments.

The research literature on environmental attitudes is large and complex, reflecting the great variety of possible environments, representations, and respondents. As such, it frustrates the desire for simple summary statements of environmental attitudes. On the other hand, the diversity of studies increases the chances of finding ones relevant to the focal environment in any stated preference study. Their stimuli might be reused, taking advantage of previous development work; their results pose a consistency test for subsequent ones. Without a comprehensive perspective, anything goes – in the sense of taking results in isolation, without context or constraint. Points of entry to the environmental attitudes literature include Cassidy (1997), Dunlap et al. (2000), Fischhoff (2001), Gardner and Stern (1996), Kempton, Boster and Hartley (1995), and Nasar (1992).

Attitude researchers are often challenged to demonstrate "attitude–behavior consistency." It might be prompted by observing people who espouse positive attitudes toward other ethnic groups, but still associate primarily with their own kind – or by observing sloth among people who express positive attitudes toward exercise (not to mention "environmentalists" driving SUVs). In decision-making terms, of course, even deeply held, well-measured values may have little necessary relationship to many potentially related behaviors. Choices should depend on attitudes towards the other expected outcomes of possible actions. Thus, people who like exercise may not act on that attitude because of the implications of their attitudes towards time, health, family, career, etc. (The strength of their attitude toward exercise might still be captured in how badly they feel about not doing – even when that remorse cannot be observed directly.)

Studies in areas as diverse as health, race, and the environment find that "attitudes are more predictive of behavior when both are measured at the same level of specificity and when the behaviors are easier to perform" [Cassidy (1997, p. 209), summarizing Ajzen and Fishbein (1980), among others]. The same applies when the behaviors are personal actions (e.g., recycling) or public ones (e.g., voting for deposit laws). Stern (1992) summarizes predictors of environmental behavior, including personal norms (a generalized attitude), public commitment, perceived barriers, and perceived efficacy.

Thus, the desire for behavioral realism has drawn some attitude researchers to pay increased attention to situational details. That can lead to task specifications as detailed as those pursued by economists who design stated preference studies motivated by particular policy concerns. Incorporating these factors transforms attitudes into something like the values (or utilities) of decision-making models. (Progress in the other direction occurs when economists elaborate their notion of the individuals making choices.)

A critical departure from the decision-making perspective is that attitude research focuses on a single option, and not a choice among options. In principle, attitudes toward one option could incorporate the alternatives, in terms of the opportunity costs of forgoing them. For example, the strength of one's attitude toward a personal norm could incorporate the cost of violating it. If so, then evaluation would reflect the action's expected net benefits. Cognitively, though, the alternatives will be less clear than the focal option. With a large option set, the alternatives may not even be fully enumerated. With two complementary options (e.g., go/don't go), raising one should immediately call the other to mind. Nonetheless, thinking about one option may evoke different consequences than thinking about the other [Beyth-Marom et al. (1993)], not to mention different valuations.

Kahneman, Ritov and Schkade (1999) offer a general account of the costs and benefits of treating attitudes as values, presented as a critique of "dollar responses to public issues." They point to a robustness of attitudes that is both boon and bane to stated preference studies. On the one hand, attitudes are so strong and easily evoked that individuals can draw upon them, whatever question they face. On the other hand, that general view can be hard to overcome when a specific evaluation is required (e.g., how much one is willing to pay for a particular good in a particular policy context). Attitudes have an affective component, central to dual-process theories. Those feelings can draw respondents into a task, at the price of distracting them from its details.

When stated preference tasks evoke attitudes, that should reduce the chance of responses showing the *scope sensitivity* sought by contingent valuation researchers [Arrow et al. (1993), Mitchell and Carson (1989), Frederick and Fischhoff (1998), Kahneman and Knetsch (1992)]. Envisioning the death of wildfowl makes many people sad, affecting some more than others. Envisioning the death of many wildfowl should create more sadness than the death of a few. However, that number should have a weaker representation, both cognitively and affectively, than the associated event (birds dying). Dual-process theories show the challenge of evoking full evaluations. On the one hand, one must engage respondents in ways that challenge their intellect, so that they absorb the details essential to policy-related decisions (e.g., the scope of the damages). On the other hand, one must preserve the affective component of valuation, lest the process "anaesthetiz[e] moral feeling" [Tribe (1972)], and evoke preferences that respondents do not fully understand or endorse.

3.2. Decision theory stream

Behavioral decision research entered environment valuation through a back door. The early modern environmental movement confronted many technology managers with opposition that they could not, or would not, understand. One natural response was to view their opponents as irrational. Whatever truth they might hold, such claims could also reflect biased observations. Technologists may not realize that they see their opponents in unrepresentative circumstances (e.g., primarily in the heat of battle); they may want to believe the worst about them. The ensuing research revealed a more complicated picture. Nonetheless, strong claims regarding public competence persist [Okrent and Pidgeon (1998)].

Initial studies used psychophysical methods. They found patterns that have borne up fairly well over time. (a) Lay and statistical estimates of annual fatalities from different sources tend to be strongly correlated. (b) The best-fit curve, relating lay and statistical estimates tends to be flat, with lay estimates spanning a smaller range than statistical ones. (c) There are systematic deviations from that curve, partially attributable to the relative *availability* of different causes of death (because people hear more about some causes than others of equal frequency). (d) There is considerable ordinal (and even ratio scale) consistency among estimates elicited with different methods (e.g., estimates of fatalities or of relative death rates). (e) There is considerable inconsistency in absolute estimates, across response modes (reflecting anchoring, among other things). Thus, respondents reveal a fairly robust feeling for relative risk levels, emerging however these unusual questions are asked. Nonetheless, the task is sufficiently unfamiliar that contextual cues strongly affect responses (consistent with general psychophysical principles) [Lichtenstein et al. (1978)]. [As an aside, the flat curve (point (b)) is often cited as showing that people overestimate small risks and underestimate large ones. However, its slope and intercept are procedure dependent, precluding inferences about absolute judgments from any one study.]

Later studies presented similar stimuli, but elicited judgments of "risk," rather than of "fatalities in an average year" – and found rather different estimates. These differences were traced to the multi-attribute character of "risk," so that technologies' "riskiness" depends on more than just average-year fatalities [Slovic, Fischhoff and Lichtenstein (1979, 1980)]. Catastrophic potential was quickly identified as one potentially important additional attribute. That is, other things being equal, people may be more averse to technologies that can claim many lives at once. This hypothesis emerged from anecdotal observation of preferences apparently stated or revealed in public actions, such as concern over plane crashes or nuclear power (given citizens' recognition that, in an average year, few people die from commercial flight or nuclear power). Catastrophic potential was taken seriously enough to be mooted in the US Nuclear Regulatory Commission's attempt to set explicit safety goals for nuclear power plants [Bier (1988), Fischhoff (1984)]. One proposal was to apply an exponent to the number of deaths from an accident sequence in risk analyses. Some critics objected to that proposal as an immoral preference for how deaths are "packaged." Others objected to how the value of the exponent could dominate regulatory proceedings.

Subsequent research suggested, however, that catastrophic potential per se did not drive lay risk concerns [Slovic, Lichtenstein and Fischhoff (1984)]. Rather, people are averse to the uncertainty often surrounding technologies that can produce catastrophic accidents. That represents a different ethical principle, with different public policy implications. For example, regulating catastrophic potential would mean encouraging small, remotely located technologies. Regulating uncertainty would mean promoting research that sharpens risk estimates and discouraging innovative technologies (which can't be known that well).

The role of catastrophic potential continued to be pursued, as one of many possible attributes of risk. The impetus to this research program was Starr's (1969) claim that, for a given level of benefit, the public tolerates higher risk levels for voluntary activities (e.g., skiing) than for involuntary ones. He backed this claim with a sort of revealed preference analysis, plotting estimates of societal risks and economic benefits from eight activities. He sketched two parallel "acceptable risk" lines, an order of magnitude apart, for voluntary and involuntary risks. Although the paper's integration of concepts was seminal, its technical treatment was but a first approximation [Fischhoff et al. (1981)]. The estimates reflected statistical estimates, rather than the lay judgments that, presumably, drove society's risk–benefit tradeoffs. Lowrance (1976) noted that voluntariness is only one feature of risk that might affect preferences. A straightforward research response to these two concerns asks citizens to evaluate technologies in terms of risk, benefits, and other attributes (such as catastrophic potential, dread, controllability, known to science, known to the public – and voluntariness). A first study [Fischhoff et al. (1978), Slovic, Fischhoff and Lichtenstein (1979)] found

(a) substantial discrepancies between lay and statistical risk and benefit estimates;
(b) a weak correlation between lay estimates of current risks and current benefits (so that society is not seen as having exacted greater benefit from riskier technologies);
(c) no greater correlation between current risk and benefit judgments, after partialing out judgments of voluntariness – or any other attribute (so that society is not seen as having set a double standard);
(d) a belief that most (but not all) technologies had unacceptable current risk levels (contrary to the hypothesis of societal revealed preferences);
(e) a significant correlation between judgments of current benefits and of acceptable risks (indicating willingness to incur greater risk in return for greater benefit);
(f) an increased correlation between perceived benefits and acceptable risks after partialing out voluntariness, and many other attributes (indicating a willingness to have double standards for qualitatively different risks).

These stated preferences indicate a willingness to accept risk–benefit tradeoffs – which might surprise some critics of public rationality. They suggest the risk attributes to consider when designing technologies or regulatory mechanisms. They imply that stated preference studies need to characterize risks in multi-attribute terms [Keeney and Raiffa (1976), von Winterfeldt and Edwards (1986)]. Fischhoff, Watson and Hope (1984) showed how the relative riskiness of energy technologies could depend on how one weighted these attributes, as well as more conventional morbidity and mortality measures. Subsequent studies have found a desire to regulate more strongly technologies with disliked attributes [McDaniels, Kamlet and Fischer (1992), Slovic (1987)].

Dealing with many attributes is unwieldy, cognitively and analytically. As a result, Fischhoff et al. (1978) looked for redundancy in ratings of its 9 attributes. Simple factor analysis found that two factors accounted for much of the variance in the ratings. In one factor, the central concept was something like how well risks are known; in the second, something like how much the risks are dreaded. Uncertainty and catastrophic potential load heavily on the former, as does voluntariness. Catastrophic potential and dread load on the second. The details of these correlations have been explored in dozens of studies, examining many attributes, elicitation procedures, data reduction methods, respondent populations, and target technologies [reviewed by Jenni (1997)]. By and large, the same two factors emerged. A third factor, when found, seems to center on present and future catastrophic potential (pulling that attribute out of the other factors).

Most subsequent studies compared attribute ratings with overall risk evaluations. As such, they fall in the psychophysical, attitude research stream – assigning numbers to objects. Given the robustness of these results, stated preference studies would be missing something if they failed to characterize risks in multi-attribute terms. The key dimensions should be clear in the stimuli, and considered when analyzing preferences. Investigators who assume that risk is just about fatalities have only part of the story. A multi-attribute representation is natural to studies using conjoint measurement and related techniques. In some cases, just mentioning a technology evokes attribute knowledge (e.g., how uncertain or dreadful it is). If can be established that an attribute goes without saying, then precious bandwidth can be saved for describing other task features.

When a multi-attribute representation is used, investigators must decide how to represent each factor. Although empirically correlated, voluntariness and uncertainty may suggest different ethical principles and evoke different preferences. For example, opposition to nuclear power sounds different when attributed to how casualties are aggregated (catastrophic potential) or to how well it is understood (uncertainty). Voluntariness sounds more egocentric than the closely correlated equity.

Morgan et al. (1996) proposed a procedure for representing attributes without drowning respondents in detail. They attempted to formalize the process used by the US Environmental Protection Agency (1990a, 1990b) in its 50 or so state and regional efforts to set risk priorities [Davies (1996)]. These exercises assembled groups of diverse citizens, for extended periods, and produced (more or less) consensual reports, ranking heterogeneous risks. These stated preferences had enough internal credibility for EPA's Scientific Advisory Board to undertake its own ranking, as a guide to agency policies. Nonetheless, it was unclear how any panel had weighted any attribute. Without transparency, it is hard to justify their rankings to nonparticipants or to aggregate rankings across exercises.

Although some standardization is needed, the panels' ability to structure their own work seemed critical to their success. Morgan et al.'s (1996) procedure uses the risk factor research to create a flexible form of standardization. It offers a fixed set of attributes, representing each risk factor, from which respondents could choose the one(s) closest to their concerns. A standard display (like Table 3) characterizes each hazard in these terms. The displays draw on risk communication research and were extensively pretested, to ensure comprehensibility and reduce sensitivity to formally irrelevant differences in display designs. Values are elicited with multiple procedures in order to help respondents absorb information and derive the implications of their basic values for these specific risks [following Gregory, Lichtenstein and Slovic (1993), Keeney (1996), McDaniels, Kamlet and Fischer (1992), National Research Council (1996), Payne, Bettman and Schkade (1999)]. Because it is an overtly reactive process, the methodology includes guidance on providing balanced, accurate materials.

The process uses two pairs of triangulating operations. One pair has respondents both perform holistic rankings and provide attribute weights from which implicit priorities are computed. They can then reconcile any inconsistencies as they see fit. The second pair alternates self-study and group discussion, with procedures designed to ensure respect for both perspectives: views are recorded in private and summarized for the group; instructions note that personal views may differ from ones expressed when seeking consensus; moderators are trained to restrain dominating personalities. Studies have found that the process increases participants' satisfaction, the agreement of holistic judgments and ones derived from attribute weights, and the agreement of public and private preferences.

Table 3
A general framework for characterizing environmental risks

Number of people affected	Degree of environmental impact	Knowledge	Dread
Annual expected number of fatalities:	Area affected by ecosystem stress or change:	Degree to which impacts are delayed:	Catastrophic potential:
0–**450**–600 (10% chance of zero)	**50** km^2	**1–10** years	**1000** times expected annual fatalities
Annual expected number of person-years lost:	Magnitude of environmental impact:	Quality of scientific understanding:	Outcome equity:
0–**9000**–18000 (10% chance of zero)	**modest** (15% chance of large)	**medium**	**medium** (ratio = 6)

Each column includes two markers for a dimension of risk, found to affect overall judgments of riskiness, desire for regulation, and other evaluative judgments [Slovic (1987, 2001)]. For example, catastrophic potential and outcome equity are two attributes that tend to be correlated with one another, with judgments of the dread that a hazard evokes, and with other attributes often described representing a dimension called “dread.” Source: Morgan et al. (1996).

Studies have also found increasing agreement among participants. However, a valid procedure could also reveal and clarify disagreements [Florig et al. (2001), Morgan et al. (2001a, 2001b)]. This convergence would be a sign of success, if one had reason to believe that participants shared underlying preferences, a sign of failure if one had reason to believe that they did not (suggesting that the process somehow manipulated them to agree). The British government has recommended a variant on this procedure for structuring deliberations over risk [HM Treasury (2005)].

4. A cognitive approach to eliciting stated preferences for environmental outcomes

Thus, cognitive approaches to preference elicitation draw on two of psychology’s major streams, as developed over the past century-plus. The psychophysical stream envisions stimuli as evoking evaluative feelings, which then must be translated into policy-relevant terms, through an intellectual process. The decision theory stream envisions specific values as being derived intellectually from basic values. However, these inferences must connect with the feelings and trial-and-error learning of everyday life, lest they be but transient artifacts. Thus, researchers in each stream have struggled to overcome its limits. Attitude research has become more cognitively complex, while decision theory has elaborated the roles of effect, experience, and reflection in choice processes [Kahneman, Diener and Schwarz (1999), Lerner and Keltner (2000), Slovic (2001), Loewenstein (1996)].

Despite their different emphases, both streams recognize common steps toward formulating meaningful values:

(1) Encode the task – in order to understand the choices;
(2) Access relevant personal values – in order to structure the evaluation process;
(3) Interpret those values in the specific context – in order to construct a preference;
(4) Translate that preference into suitable terms – in order to state a preference.

Of course, everyday life itself may not provide these conditions, witness individuals' occasional confusion regarding what they want from various (central and peripheral) decisions in their lives. However, investigators owe it to their respondents and their readers to create such conditions and to evaluate their success – in order to demonstrate that they have elicited stated preferences worth taking seriously.

Researchers' aspirations can be arrayed along a continuum, ranging from *gist* to *contractual* studies. *Gist* studies claim to elicit general answers to general questions, such as how much people support the environment, worry about pollution, trust industry, or dislike regulation. *Contract* studies claim to elicit valuations for specific transactions, such as "willingness to pay 7% for 'green' products" (with "green" suitably defined) or "to incur a 0.2% rise in unemployment in order to meet Kyoto obligations."

Researchers have an obligation to provide respondents with the conditions needed to produce the valuations they seek [Fischhoff (2000)]. Otherwise, they (and those who use their work) may misrepresent respondents. Gist responses should provide only vague direction. Respondents haven't said very much; no one should read very much into their responses. Doing more is akin to politicians basing strong, specific mandates on diffuse electoral indicators. Gist researchers should oppose misuse of their work, in terms such as, "All they (the public) said was that they wanted cleaner cars; they didn't say that they wanted to mandate that particular fuel system," or "Respondents' general support for 'free trade' does not imply knowledge and advocacy of all WTO environmental provisions." Contractual claims bear a much greater burden of proof, for demonstrating that respondents have completed the four tasks – hence really understand the agreement they are endorsing. With a fixed budget, meeting these demands will mean smaller samples. That means achieving statistical power through more precise measurement, rather than through increased sample size (whose effects are more easily estimated).

The remainder of this section considers how cognitive psychologists approach the challenge posed by contractual studies – which have greater interest for policy makers (and environmental economists). With gist studies, the primary risk is asking respondents to reflect too hard on their values, taking them beyond their initial gut reactions, perhaps even throwing them into question [Fischhoff (1991, 2000), Wilson et al. (1993)].

In many ways, eliciting environmental values is no different than eliciting any other values. For example, it is generally true that: (a) Multiple methods are needed when eliciting preferences, in order to demonstrate method invariance. (b) Constructive elicitation is needed – unless respondents have articulated values that they can "read off" for the specific question. (c) Communication studies are needed when respondents lack an understanding of the issues that they are evaluating. Eliciting stated preferences for environmental changes is unique in the substance of its problems and the frequent severity of these design challenges. When policy questions define contractual evaluations, respondents often must master details that are unfamiliar in that specific combination, even when familiar in general. They may also face deceptively simple questions, as when researchers demand a contractual response to a nuanced general principle (e.g., discount rate, equity) [Frederick, Loewenstein and O'Donoghue (2002)]. Omitting features does not keep respondents from reading them between the lines of a task description [Fischhoff, Welch and Frederick (1999)].

The first step to ensuring shared understanding is creating a full task specification, addressing each feature important to the investigators or respondents. Table 4 shows a framework for eliciting contractual stated preferences for environmental changes. It holds that the respondent and the investigator must agree about the meaning of the *good*, the *payment*, and the *social context* of the proposed *transaction*. If not, then respondents are answering a different question than the one being asked. Fischhoff and Furby (1988) describe how each feature might matter when evaluating changes in atmospheric visibility.

Table 4
A framework for defining transactions (features that may require specification, if an evaluation task is to be understood similarly by respondents, investigators, and policymakers)

The good (e.g., visibility)
Substantive definition (aspects of proposed change in good that may matter to evaluators)
Focal attribute(s)
(e.g., haze intensity, visual range, plume color, light extinction)
Context (giving particular value to attribute)
(e.g., natural or built, judged uniqueness, associated activities (such as hiking, viewing, playing), significance (such as religious, cultural, historical))
Source of change (in focal attribute)
predominantly natural (e.g., vegetation, forest fires, dust storms, humidity) or human (e.g., power plant, other factory, field burning, slash burning, motor vehicles)
Formal definition (specifying extent of change in valued focal attributes)
Reference and target levels (of good, before and after change)
magnitude and direction of change, statistical summary, form of representation (mode, richness, organization)
Extent of change
geographical, temporal
Timing of change (when will it happen?)
Certainty of provision (will it really happen?)

(*continued on next page*)

Table 4
(*Continued*)

- The value measure (e.g., money, time, discomfort, effort)
  - *Substantive definition* (aspects of proposed change in payment that may matter to evaluators)
    - Focal attribute(s)
      - dollars (for money)
      - foregone leisure or work (for time)
      - physical or emotional toll (for discomfort or effort)
    - Context
      - electric bill, sales tax, income tax, park entry fee, environmental fund (for money)
      - when convenient, when demanded (for time)
      - when rested, when exhausted (for effort)
    - Constituency
  - *Formal definition* (specifying extent of change in valued focal attributes)
    - Reference and target levels
      - magnitude and direction of change, statistical summary, elicitation procedure (response mode, response format, cues, feedback)
    - Extent of payment
      - frequency, duration
    - Timing of payment (when will it happen?)
    - Certainty of payment (will it really happen?)
- The social context
  - *Other people involved*
    - Provider of the good
    - Others present
  - *Resolution mechanism* (determining whether transaction will actually occur)
    - Determining parties
    - Iterations, constraints
  - *Other stakes involved*
    - Externalities
    - Precedents
    - Legitimacy of process

Source: Fischhoff and Furby (1988).

Table 5 presents a second framework, for specifying tasks involving time preferences. It shows the various things that might differ when a good could be provided at two times – in addition to the utility of receiving it. For example, respondents could believe that they are less likely to receive a good at a latter time, investigators' reassurances notwithstanding. If so, then they should value the future prospect less, even if the good itself would have equal utility at either time. These concurrent changes in task features confound interpreting choices over time as expressing pure time preferences.

Table 5
Reasons for evaluating goods differently at different times

Model	Corresponding description in words
DU (time preference only)	Future utility should be discounted because we should care less about the later parts of our life (for some unexplained reason)
DU + probability	Future utility should be weighted by the probability that the consequence that gives rise to the utility will actually occur
DU + changes in objective consequence	The objective properties of some coarsely defined consequence may depend on the time at which it occurs
DU + changes in utility function	The subjective utility associated with a particular objective consequence may change over time
DU + utility from anticipation	The utility at a given moment may be influenced by the anticipation of future utility
DU + utility from memory	The utility at a given moment may be influenced by the recollection of past utility
DU + opportunity cost	Utility depends on the current consumption level, and the potential consumption level depends on current income and past investment

Source: Frederick, Loewenstein and O'Donoghue (2002).

There is, of course, some circularity in needing to know how much features matter in a context before evaluating environmental changes with those features in that context. That circle is broken by the cumulative empirical record of how much those features have typically been found to matter for that class of changes. Formal properties such as the scope and probability of the change (see Table 4) always need to be specified.

The effort needed to convey a task definition depends on how familiar respondents are with it. As mentioned, when voting on a widely debated referendum, many citizens know what contractual commitment it implies – and could respond reliably to an opinion poll. Similarly, many citizens know the verdict that they want from one of our periodic show trials (Clinton, Simpson), and how that would vary with changed evidence or charges. In such situations, many details "go without saying." Omitting them maintains conversational norms and leaves time for communicating less obvious features.

Investigators can approach their communication challenge in a piecemeal or holistic way. That is, they can try either to convey the few most important individual features or to create a meaningful whole that facilitates recalling and inferring features [Fischhoff (1999, 2000)]. The piecemeal strategy creates a supply curve for features, focusing on those that respondents most need to learn [Merz, Small and Fischbeck (1992), Riley, Small and Fischhoff (2000)]. That perspective frames communications adequacy in terms of how much has been conveyed. Sometimes, only a few things really matter. Sometimes, many do. One might expand the envelope of comprehension through more intensive, interactive procedures [e.g., Gregory, Lichtenstein and Slovic (1993), Whittington (1998)]. Or, one might settle for a gist study, getting across just the rough idea.

People absorb information more quickly, when they can organize it into *chunks*, processed cognitively as units [Miller (1956)]. Mnemonists take this skill to high art [Luria (1968)], creating chunks from diverse elements, by integrating them into highly flexible templates. Ordinary people organize information into less coherent, domain-specific *mental models*. Activating a mental model allows some task features to go without saying, while making others easier to absorb. Mental models can also prompt unintended inferences and hinder the processing of unexpected features. For example, "referendum" is a widely used metaphor in CVM studies, asking respondents to imagine voting on a proposed transaction [McDaniels (1996)]. Respondents might naturally infer a government sanctioned, legally binding, take-it-or-leave-it choice, decided by majority rule and open to all citizens. If so, then the single term predictably evokes multiple features. That is good, if they are legitimate inferences; bad, if they are not (e.g., if they lead respondents to exaggerate the probability of the promised good or payment being delivered).

Psychology has studied mental models in many domains [Bartlett (1932), Gentner and Stevens (1983), Rouse and Morris (1986)]. These studies take advantage of the coherence of natural systems to create holistic pictures, allowing respondents to integrate fragmentary beliefs, absorb new features and infer unspoken ones. These results, and the methods that produced them, provide a resource for communicating evaluation tasks [Fischhoff (1999), Morgan et al. (2001a, 2001b)].

Once investigators have finished communicating, they must assess their success. One standard assessment method is the *think-aloud protocol* [Ericsson and Simon (1994), Schkade and Payne (1994), Schriver (1989)]; whereby respondents describe whatever comes into their minds, as they read their task. The interviewer requests enough elaboration to be sure that respondents' meaning is understood. In addition to cases of obvious ambiguity, prompts are needed when people are known to use a term in different ways – for example, "safe sex" [McIntyre and West (1992)], "climate" [Read et al. (1994)], "employed," or "room" [Turner and Martin (1984)].

Placed at the end of a task, *manipulation checks* ask respondents to report their understanding of critical features. Table 6 shows results of three manipulation checks, administered after a short stated preference task. A plurality of respondents reported the actual value for the first check. They did less well on the two other checks. Despite the task's brevity, most respondents did not hear, believe, or remember these features. They were, in effect, answering a different question than the one that was asked.

When performed in pretests, think-aloud protocols and manipulation checks allow predicting how well the task will be mastered by participants in the actual study, who work equally hard. When performed on actual respondents, these assessments show where respondents fall relative to the acceptable level of misunderstanding, hence what conclusions the study can support.

Table 6
Manipulation checks (percent of respondents choosing each offered value of the stimulus feature; bold indicates the value actually in the stimulus)

Condition	Offered values of each stimulus feature			
	Miles of river in proposed cleanup			
	0–100 miles	101–1000 miles	1001–10000 miles	Don't know
30 miles	**62**	9	1	29
1000 miles	4	**34**	32	31
	Feasibility of proposed cleanup			
	Eliminated completely	Good headway	Not much progress	Did not think
30 miles	**35**	46	4	13
1000 miles	**18**	52	10	18
	Payment vehicle for proposed cleanup			
	Taxes	Higher prices	Donations	Other
30 miles	63	**25**	22	19
1000 miles	65	**31**	28	16

The task
In a phone interview, Pittsburgh-area respondents were asked about their willingness to pay (in higher prices for goods and services) to complete cleaning up an environment problem. For the 1000-mile condition they were told "Presently, a large number of the rivers in Pennsylvania are seriously polluted. These rivers include the Delaware, Susquehanna, Monongahela, Allegheny, Ohio, Clarion, Schuylkill and Lehigh. All together, there are more than 3,000 miles of rivers, of which more than 1,000 miles are polluted. Authorities caution against swimming in or eating fish caught from these polluted portions of the rivers." The 30-mile condition mentioned only the Susquehanna River. The first manipulation check asked which of the three ranges contained the value that had been read to them in the 30-mile condition, 62% of respondents reported a value in the correct (0–100 mile) range. The rest reported incorrect values or did not know [source: Fischhoff et al. (1993)].

Even a clear specification and a diligent presentation will leave some gaps between the question being asked and the one being answered. There are three ways to deal with the residual imperfections: (a) *Disqualify* respondents whose task construal strays too far from the intended one; (b) *Adjust* stated preferences to undo the effects of the misconstrual (e.g., double the value assigned to a good that the task promised but a respondent saw but a 50% chance of actually receiving); (c) *Accommodate* misconstruals, when reporting study results (e.g., separate the preferences stated by respondents demonstrating different degrees of task mastery; note common forms of disbelief). Participants in Table 6's study showed greater scope sensitivity when their responses were analyzed in terms of the question that they reported answering, rather than the one actually presented.

5. Conclusions

In the short run, the costs of all this design work can be reduced by exploiting solutions already in the cognate literatures (e.g., proven ways to describe an ecosystem). The analytical frameworks of Tables 4 and 5 represent two ways to hasten the accumulation of such knowledge. Each identifies essential features of tasks, for which regularities can be sought. Over the long run, design costs should decline as a cumulative empirical record is created, on issues central to eliciting environmental values (e.g., how to convey the time period for a payment or the magnitude of a change, how to convince respondents that a change will really happen). As mentioned, such research can simplify otherwise complex tasks, by showing which features can go without saying or need no explanation. However, design research may also reveal investigators to be victims of a *curse of cleverness*. We prize novel tasks, whose formulation captures nuances that eluded our colleagues or address emerging policy concerns. Such tasks are necessarily even more novel for respondents. The greater the novelty, the greater the need for explanatory exposition and constructive elicitation – if respondents are to understand the choice being posed and their own preferences. The reward for such efforts is improving the signal from stated preference studies, by reducing respondents' uncertainty about the investigator's question and their answer. The research literature of cognitive psychology both demonstrates the reality of these challenges and offers resources for addressing them.

6. Contingent valuation: a postscript

As mentioned, cognitive psychology is best known to many environmental economists through controversies over contingent valuation. The potential constructive contributions of the other literatures cited here are rarely mentioned, while the controversies have a polemic character, ill suited to fostering collaboration and understanding [Driver, Peterson and Gregory (1988), Fischhoff and Furby (1986), Furby and Fischhoff (1988), Kahneman and Knetsch (1992), Kopp, Pommerehne and Schwarz (1997), Schkade and Payne (1994)]. From the tenor of these interactions, one might infer that cognitive psychologists are fundamentally opposed to eliciting stated preferences for environmental goods. However, their opposition is to the methods used to ask questions, not to pursuit of the answers. As long as contingent valuation adopts the psychophysics paradigm, most cognitive psychologists will remain skeptical of its claims – believing that attitude-research methods cannot elicit values supporting the contractual claims sought by CV researchers [e.g., Fischhoff (1991, 2000), Gregory, Lichtenstein and Slovic (1993), Kahneman, Knetsch and Thaler (1991), Kahneman, Diener and Schwarz (1999), Payne, Bettman and Johnson (1993), Payne, Bettman and Schkade (1999)]. Eliciting more than just the gist of environmental preferences will require constructive procedures, rooted in the decision theory stream. Creating them will require accepting reactive measurement, and the philosophy of science that supports it. It will demand better-specified tasks and more extensive manipulation checks than has been common – in order to ensure that respondents answer the specific question that interests policy makers. It will afford and require opportunities for collaboration among psychologists, economists, and others.

References

Ajzen, I., Fishbein, M. (1980). Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs, NJ.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R., Schuman, H. (1993). “Report of the NOAA panel on contingent valuation”. Federal Register 58, 4601–4614.

Bartlett, R.C. (1932). Remembering. Cambridge University Press, Cambridge.

Berlyne, D.E. (1971). Aesthetics and Psychobiology. Appleton-Century-Crofts, New York, NY.

Beyth-Marom, R., Austin, L., Fischhoff, B., Palmgren, C., Quadrel, M.J. (1993). “Perceived consequences of risky behaviors”. Developmental Psychology 29, 549–563.

Bier, V.M. (1988). “The US Nuclear Regulatory Commission’s safety goal policy: A critical review”. Risk Analysis 8, 563–568.

Brunswick, E. (1947). Systematic and Representative Design of Psychological Experiments. University of California Press, Los Angeles, CA.

Budescu, D.F., Wallsten, T.S. (1995). “Processing linguistic probabilities: General principles and empirical evidence”. In: Busemeyer, J.R., Hastie, R., Medin, D.L. (Eds.), Decision Making from a Cognitive Perspective. Academic Press, New York, pp. 275–318.

Cantor, D. (1977). The Psychology of Place. Architectural Press, London, UK.

Cassidy, T. (1997). Environmental Psychology. Taylor and Frances, London, UK.

Clemen, R.T. (1996). Making Hard Decisions: An Introduction to Decision Analysis. Duxbury, Belmont, CA.

Cummings, R.G., Brookshire, D.S., Schulze, W.D. (Eds.) (1986). Valuing Environmental Goods: An Assessment of the Contingent Valuation Method. Rowman & Allanheld, Totowa, NJ.

Daniel, T.C., Meitner, M.M. (2001). “Representational validity of landscape visualizations: the effects of graphical realism on perceived scenic beauty of forest landscapes”. Journal of Environmental Psychology 21, 61–72.

Daniel, T.C., Vining, J. (1983). “Methodological issues in the assessment of landscape quality”. In: Altman, I., Wohlwill, J.F. (Eds.), Human Behavior and Environment, vol. 16. Plenum Press, New York, NY, pp. 39–84.

Davies, C. (Ed.) (1996). Comparing Environmental Risks. Resources for the Future, Washington, DC.

Dawes, R.M. (1988). Rational Choice in an Uncertain World. Harcourt, Brace Jovanovich, San Diego, CA.

Dawes, R.M., Corrigan, B. (1974). “Linear models in decision making”. Psychological Bulletin 81, 95–106.

Driver, B., Peterson, G., Gregory, R. (Eds.) (1988). Evaluating Amenity Resources. Venture, New York.

Dunlap, R.E., Van Liere, K.D., Mertig, A.G., Jones, R.E. (2000). “Measuring endorsement of the new ecological paradigm: A revised NEP scale”. Journal of Social Issues 56, 425–442.

Eagly, A.H., Chaiken, S. (1993). The Psychology of Attitudes. Harcourt, Brace & Jovanovic, Orlando.

Edwards, W. (1954). “A theory of decision making”. Psychological Bulletin 54, 380–397.

Edwards, W. (1961). “Behavioral decision theory”. Annual Review of Psychology 12, 473–498.

Ericsson, A., Simon, H. (1994). Verbal Reports as Data, 2nd ed. MIT Press, Cambridge, MA.

Fischhoff, B. (1980). “Clinical decision analysis”. Operations Research 28, 28–43.

Fischhoff, B. (1984). “Setting standards: A systematic approach to managing public health and safety risks”. Management Science 30, 823–843.

Fischhoff, B. (1988a). “Judgment and decision making”. In: Sternberg, R.J., Smith, E.E. (Eds.), The Psychology of Human Thought. Cambridge University Press, New York, pp. 153–187.

Fischhoff, B. (1988b). “Specifying value measurements”. In: Driver, B., Peterson, G., Gregory, R. (Eds.), Evaluating Amenity Resources. Venture, New York, pp. 107–116.

Fischhoff, B. (1991). “Value elicitation: Is there anything in there?”. American Psychologist 46, 835–847.

Fischhoff, B. (1994). "What forecasts (seem to) mean". International Journal of Forecasting 10, 387–403.
Fischhoff, B. (1998). "Communicate unto others . . .". Reliability Engineering and System Safety 59, 63–72.
Fischhoff, B. (1999). "Why (cancer) risk communication can be hard". Journal of the National Cancer Institute Monographs 25, 7–13.
Fischhoff, B. (2000). "Informed consent in eliciting environmental values". Environmental Science and Technology 38, 1439–1444.
Fischhoff, B. (2001). "Environmental cognition, perception, and attitudes". In: Baltes, P.B., Smelser, N.J. (Eds.), International Encyclopedia of the Social and Behavioral Sciences, vol. 7. Pergamon, Kidlington, UK, pp. 4596–4602.
Fischhoff, B., Bruine de Bruin, W. (1999). "Fifty/fifty = 50?". Journal of Behavioral Decision Making 12, 149–163.
Fischhoff, B., Furby, L. (1986). A review and critique of Tolley, Randall et al. "Establishing and valuing the effects of improved visibility in the Eastern United States". ERI Technical Report 86-8. Eugene, OR, Eugene Research Institute.
Fischhoff, B., Furby, L. (1988). "Measuring values: A conceptual framework for interpreting transactions". Journal of Risk and Uncertainty 1, 147–184.
Fischhoff, B., Slovic, P., Lichtenstein, S. (1980). "Knowing what you want: Measuring labile values". In: Wallsten, T. (Ed.), Cognitive Processes in Choice and Decision Behavior. Erlbaum, Hillsdale, NJ, pp. 117–141.
Fischhoff, B., Watson, S., Hope, C. (1984). "Defining risk". Policy Sciences 17, 123–139.
Fischhoff, B., Welch, N., Frederick, S. (1999). "Construal processes in preference elicitation". Journal of Risk and Uncertainty 19, 139–164.
Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S.L., Keeney, R.L. (1981). Acceptable Risk. Cambridge University Press, New York.
Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., Combs, B. (1978). "How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits". Policy Sciences 8, 127–152.
Fischhoff, B., Quadrel, M.J., Kamlet, M., Loewenstein, G., Dawes, R., Fischbeck, P., Klepper, S., Leland, J., Stroh, P. (1993). "Embedding effects: Stimulus representation and response modes". Journal of Risk and Uncertainty 6, 211–234.
Fiske, S. (1998). Social Psychology. Wiley, New York.
Florig, H.K., Morgan, M.G., Morgan, K.M., Jenni, K.E., Fischhoff, B., Fischbeck, P.S., DeKay, M. (2001). "A test bed for studies of risk ranking". Risk Analysis 21, 913–922.
Frederick, S., Fischhoff, B. (1998). "Scope insensitivity in elicited values". Risk Decision and Policy 3, 109–124.
Frederick, S., Loewenstein, G., O'Donoghue, T. (2002). "Time discounting and time preference: A critical review". Journal of Economic Literature 40, 351–401.
Frumkin, H. (2001). "Beyond toxicity: The greening of environmental health". American Journal of Preventive Medicine 20, 47–53.
Furby, L., Fischhoff, B. (1988). "Specifying subjective evaluations. A critique of Dickie et al.'s interpretation of their contingent valuation results for reduced minor health symptoms". USEPA Cooperative Agreement No. CR814655-01-0. Eugene Research Institute, Eugene, OR.
Gardner, G.T., Stern, P.C. (1996). Environmental problems and human behavior. Allyn & Bacon, Boston, MA.
Gärling, T. (1976). "The structural analysis of environmental perception and cognition". Environment and Behavior 8, 258–263.
Gentner, D., Stevens, A.L. (Eds.) (1983). Mental Models. Erlbaum, Hillsdale, NJ.
Gibson, J.J. (1979). The Ecological Approach to Visual Perception. Houghton Mifflin, Boston, MA.
Gilovich, T., Griffin, D., Kahneman, D. (Eds.) (2002). The Psychology of Judgment: Heuristics and Biases. Cambridge University Press, New York.
Green, D.M., Swets, J.A. (1966). Signal Detection Theory and Psychophysics. Wiley, New York.
Gregory, R., Lichtenstein, S., Slovic, P. (1993). "Valuing environmental resources: A constructive approach". Journal of Risk and Uncertainty 7, 177–197.

Grice, H.P. (1975). “Logic and conversation”. In: Davidson, D., Harman, G. (Eds.), The Logic of Grammar. Dickenson, Encino.

Guston, D. (1999). “Evaluating the first US consensus conference: The impact of the citizens’ panel on telecommunications and the future of democracy”. Science, Technology and Human Values 24, 451–482.

HM Treasury (2005). Managing Risks to the Public. HM Treasury, London.

Jenni, K. (1997). Attributes for Risk Evaluation. Unpublished doctoral dissertation. Department of Engineering and Public Policy, Carnegie Mellon University.

Kahneman, D., Knetsch, J.L. (1992). “Valuing public goods: The purchase of moral satisfaction”. Journal of Environmental Economics and Management 22, 57–70.

Kahneman, D., Tversky, A. (1979). “Prospect theory: An analysis of decision under risk”. Econometrica 47, 263–281.

Kahneman, D., Tversky, A. (1984). “Choices, values, and frames”. American Psychologist 39, 341–350.

Kahneman, D., Diener, E., Schwarz, N. (Eds.) (1999). Well-being: The Foundations of Hedonic Psychology. Russell Sage Foundation Press, New York.

Kahneman, D., Knetsch, J.L., Thaler, R.H. (1991). “The endowment effect, loss aversion, and status quo bias”. Journal of Economic Perspectives 5, 193–206.

Kahneman, D., Ritov, I., Schkade, D. (1999). “Economic preferences or attitude expression?”. Journal of Risk and Uncertainty 19, 203–242.

Kahneman, D., Slovic, P., Tversky, A. (Eds.) (1982). Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, New York.

Kaplan, R., Kaplan, S. (1989). The Experience of Nature: A Psychological Perspective. Cambridge University Press, Cambridge, UK.

Keeney, R. (1996). Value-focused Thinking. MIT Press, Cambridge, MA.

Keeney, R., Raiffa, H. (1976). Decisions with Multiple Objectives: Preferences and Value Trade-offs. Wiley, New York.

Kelly, G.A. (1955). The Psychology of Personal Constructs. Norton, New York, NY.

Kempton, W., Boster, J.S., Hartley, J.A. (1995). Environmental Values in American Culture. MIT Press, Cambridge, MA.

Kopp, R., Pommerehne, W.W., Schwarz, N. (Eds.) (1997). Determining the Value of Nonmarketed Goods. Kluwer, New York.

Larkin, J.H., Simon, H.A. (1987). “Why a diagram is (sometimes) worth 10,000 words”. Cognitive Science 11, 65–100.

Leamer, E. (1983). “Let’s take the con out of econometrics”. American Economic Review 72 (1), 31–43.

Lerner, J.S., Keltner, D. (2000). “Beyond valence: Toward a model of emotion-specific influences on judgment and choice”. Cognition and Emotion.

Lichtenstein, S., Newman, J.R. (1967). “Empirical scaling of common verbal phrases associated with numerical probabilities”. Psychonomic Science 9, 563–564.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., Combs, B. (1978). “Judged frequency of lethal events”. Journal of Experimental Psychology: Human Learning and Memory 4, 551–578.

Loewenstein, G. (1996). “Out of control: Visceral influences on behavior”. Organizational Behavior and Human Decision Processes 65, 272–292.

Lowrance, W.W. (1976). Of Acceptable Risk: Science and the Determination of Safety. William Kaufman, Los Altos, CA.

Luria, A.R. (1968). The Mind of a Mnemonist. Basic Books, New York.

McDaniels, T. (1996). “The structured value referendum: Eliciting preferences for environmental policy alternatives”. Journal of Policy Analysis and Management 15, 227–251.

McDaniels, T., Kamlet, M.S., Fischer, G.W. (1992). “Risk perception and the value of safety”. Risk Analysis 12, 495–503.

McFadden, D. (1999). “Rationality for economists?”. Journal of Risk and Uncertainty 19, 73–110.

McGuire, W. (1969). “Suspiciousness of experimenter’s intent”. In: Rosenthal, R., Rosnow, R.L. (Eds.), Artifact in Behavioral Research. Academic Press, New York.

McIntyre, S., West, P. (1992). "What does the phrase "safer sex" mean to you? Understanding among Glaswegian 18-year olds in 1990". AIDS 7, 121–126.

Merz, J.F., Small, M., Fischbeck, P. (1992). "Measuring decision sensitivity: A combined Monte Carlo – logistic regression approach". Medical Decision Making 12, 189.

Miller, G.A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review 63, 81–97.

Mitchell, R.C., Carson, R.T. (1989). Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington, DC.

Moore, E.O. (1981). "A prison environment's effect on health care service demands". Journal of Environmental Systems 11, 17–34.

Morgan, K.M., DeKay, M.L., Fischbeck, P.S., Morgan, M.G., Fischhoff, B., Florig, H.K. (2001a). "A deliberative method for ranking risks: Evaluating validity and usefulness". Risk Analysis 21, 923–938.

Morgan, M.G., Fischhoff, B., Bostrom, A., Atman, C. (2001b). Risk Communication: The Mental Models Approach. Cambridge University Press, New York.

Morgan, M.G., Fischhoff, B., Lave, L., Fischbeck, P. (1996). "A proposal for ranking risks within federal agencies". In: Davies, C. (Ed.), Comparing environmental risks. Resources for the Future, Washington, DC, pp. 111–147.

Nasar, J.L. (Ed.) (1992). Environmental Aesthetics: Theory Research and Applications. Cambridge University Press, New York, NY.

National Research Council (1982). Survey Measure of Subjective Phenomena. National Academy Press, Washington, DC.

National Research Council (1996). Understanding Risk. The Council, Washington, DC.

Nisbett, R.E., Wilson, T.D. (1977). "Telling more than we know: Verbal reports on mental processes". Psychological Review 84, 231–259.

Okrent, D., Pidgeon, N. (Eds.) (1998). Actual versus Perceived Risk. Reliability Engineering and System Safety 59 (1998). Special issue.

Payne, J.W., Bettman, J.R., Johnson, E. (1993). The Adaptive Decision Maker. Cambridge University Press, Cambridge.

Payne, J.W., Bettman, J.R., Schkade, D. (1999). "Measuring constructed preferences: Toward a building code". Journal of Risk and Uncertainty 19, 243–270.

Plous, S. (1993). The Psychology of Judgment and Decision Making. McGraw–Hill, New York.

Poulton, E.C. (1968). "The new psychophysics: Six models for magnitude estimation". Psychological Bulletin 69, 1–19.

Poulton, E.C. (1989). Bias in Quantifying Judgment. Lawrence Erlbaum, Hillsdale, NJ.

Poulton, E.C. (1994). Behavioral Decision Making. Lawrence Erlbaum, Hillsdale, NJ.

Raiffa, H. (1968). Decision Analysis: Introductory Lectures on Choices under Uncertainty. Addison-Wesley, Reading, MA.

Read, D., Bostrom, A., Morgan, M.G., Fischhoff, B., Smuts, T. (1994). "What do people know about global climate change? Part 2. Survey studies of educated laypeople". Risk Analysis 14, 971–982.

Riley, D.M., Small, M.J., Fischhoff, B. (2000). "Modeling methylene chloride exposure–reduction options for home paint-stripper users". Journal of Exposure Analysis and Environmental Epidemiology 10 (3), 240–250.

Rokeach, M. (1973). The Nature of Human Values. The Free Press, New York.

Rouse, W.B., Morris, N.M. (1986). "On looking into the black box: Prospects and limits in the search for mental models". Psychological Bulletin 100, 349–363.

Schkade, D., Payne, J.W. (1994). "How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation". Journal of Environmental Economics and Management 26, 88–109.

Schriver, K.A. (1989). Plain Language for Expert or Lay Audiences: Designing Text Using Protocol Aided Revision. Communications Design Center, Carnegie Mellon University.

Schwarz, N. (1996). Cognition and Communication: Judgmental Biases, Research Methods and the Logic of Conversation. Erlbaum, Hillsdale, NJ.

Schwarz, N. (1999). “Self-reports: How the questions shape the answers”. American Psychologist 54, 93–105.
Slovic, P. (1987). “Perceptions of risk”. Science 236, 280–285.
Slovic, P. (Ed.) (2001). The Perception of Risk. Earthscan, London.
Slovic, P., Lichtenstein, S. (1971). “Comparison of Bayesian and regression approaches to the study of information processing in judgment”. Organizational Behavior and Human Performance 6, 649–744.
Slovic, P., Fischhoff, B., Lichtenstein, S. (1979). “Rating the risks”. Environment 21 (4), 14–20, 36–39.
Slovic, P., Fischhoff, B., Lichtenstein, S. (1980). “Facts and fears: Understanding perceived risk”. In: Schwing, R., Albers, W.A. Jr. (Eds.), Societal Risk Assessment: How Safe Is Safe enough?. Plenum Press, New York, pp. 181–214.
Slovic, P., Lichtenstein, S., Fischhoff, B. (1984). “Modeling the societal impact of fatal accidents”. Management Science 30, 464–474.
Starr, C. (1969). “Social benefit versus technological risk”. Science 165, 1232–1238.
Stern, P.C. (1992). “Psychological dimensions of global environmental change”. Annual Review of Psychology 43, 269–302.
Stevens, S.S. (1975). Psychophysics: Introduction to Its Perceptual, Neural and Social Prospects. Wiley, New York, NY.
Swets, J.A. (1992). “The science of choosing the right decision threshold in high-stakes diagnostics”. American Psychologist 47, 522–532.
Tanur, J.M. (Ed.) (1992). Questions about Questions. Russell Sage Foundation, New York.
Tribe, L.H. (1972). “Policy science: Analysis or ideology?”. Philosophy and Public Affairs 2, 66–110.
Tolley G. et al. (1986). “Establishing and valuing the effects of improved visibility in the Eastern United States”. Report to the Environmental Protection Agency.
Tune, G.S. (1964). “Response preferences: A review of some relevant literature”. Psychological Bulletin 61, 286–302.
Turner, C.F., Krauss, E. (1978). “Fallible indicators of the subjective state of the nation”. American Psychologist 33, 456–470.
Turner, C.F., Martin, E. (Eds.) (1984). Surveying Subjective Phenomena. Russell Sage Foundation, New York.
Tversky, A., Kahneman, D. (1974). “Judgment under uncertainty: Heuristics and biases”. Science 185, 1124–1131.
Ulrich, R.S. (1984). “View through a window may influence recovery from surgery”. Science 224, 420–421.
US Environmental Protection Agency (1990a). Reducing Risk: Setting Priorities and Strategies. US Environmental Protection Agency, Washington, DC.
US Environmental Protection Agency (1990b). Comparative Risk. US Environmental Protection Agency, Washington, DC.
von Neumann, J., Morgenstern, O. (1947). Theory of Games and Economic Behavior, 2nd ed. Princeton University Press, Princeton.
von Winterfeldt, D., Edwards, W. (1986). Decision Analysis and Behavioral Research. Cambridge University Press, New York.
Wallsten, T., Budescu, D. (1983). “Encoding subjective probabilities: A psychological and psychometric review”. Management Science 29, 135–140.
Whittington, D. (1998). “Administering contingent valuation surveys in developing countries”. World Development 26, 21–30.
Wickelgren, W. (1977). “Speed–accuracy trade-off and information processing dynamics”. Acta Psychologica 41 (1), 67–85.
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 Social Psychology Bulletin 19, 331–339.
Woodworth, R.S., Schlosberg, H. (1954). Experimental Psychology. Holt, New York.
Yates, J.F. (1990). Judgment and Decision Making. Wiley, New York.