NEW YORK UNIVERSITY
LAW REVIEW

VOLUME 81 DECEMBER 2006 NUMBER 6

ARTICLES

WELFARE POLLS: A SYNTHESIS

MATTHEW D. ADLER*

"Welfare polls" are survey instruments that seek to quantify the determinants of human well-being. Currently, three welfare polling formats are dominant: contingent valuation (CV) surveys, quality-adjusted life year (QALY) surveys, and happiness surveys. Each format has generated a large, specialized, scholarly literature, but no comprehensive discussion of welfare polling as a general enterprise exists. This Article seeks to fill that gap.

Part I describes the trio of existing formats. Part II discusses the current and potential uses of welfare polls in governmental decisionmaking. Part III analyzes in detail the obstacles that welfare polls must overcome to provide useful well-being information, and concludes that they can be genuinely informative. Part IV synthesizes the case for welfare polls, arguing against two types of challenges: the revealed-preference tradition in economics, which insists on using behavior rather than surveys to learn about well-being; and the civic republican tradition in political theory, which accepts surveys but insists that respondents should be asked to take a "citizen" rather than "consumer" perspective. Part V suggests new directions for welfare polls.

 Leon Meltzer Professor of Law, University of Pennsylvania Law School. Many thanks to Abby Wright for superb research assistance and to Paul Dolan, Eric Posner, Valerie Tiberius, and participants in a workshop on National Well-Being Indicators held during the 2006 Medici Conference at the University of Pennsylvania Positive Psychology Center, for their comments.

INTRODUCTION

What are the avenues for citizen participation in the administrative state? Citizen participation may be an intrinsic good, and at a minimum surely can be instrumentally valuable in improving the

quality of governmental decisions. But the traditional avenues for citizen participation in administrative decisionmaking have familiar flaws. Agency adjudications may be accompanied by oral hearing rights for the targeted individuals,[1] who will typically have strong incentives to exercise those rights. Yet agencies, with rare exceptions, need not conduct oral hearings when they issue regulations or undertake other general decisions[2]—and, in any event, trial-like process in these cases could be expected to generate low and unrepresentative citizen participation, given free rider problems.

Free rider problems also beset the participatory mechanisms that do currently accompany rulemakings. Citizens can lobby their legislators, who in turn can pressure administrators; they can join the notice-and-comment process that agencies are required to conduct for most legally binding rules, sending written comments that the agency will be required to read and address;[3] and they can show up and speak at the informal public meetings that agencies often hold prior to the promulgation of important regulations.[4] But in each case the rational-apathy dynamic will set in. Most individuals reasonably expect to have little chance, via the mechanisms just described, of changing the outcome of the administrative decision. Thus they do better by remaining uninvolved. Those who do become involved will be a self-selected, statistically biased sample of the public, and they will also tend to be uninformed and make relatively little effort to understand the issues at hand.[5]

These are not novel observations, of course, and much recent scholarly work has been undertaken that contemplates innovative participatory devices—devices to produce citizen involvement in administrative decisionmaking that is better informed, more thorough, and representative of the citizenry as a whole. Most of this scholarship is inspired by the "deliberative democracy" or civic republican tradition

---

[1] *See generally* 1 RICHARD J. PIERCE, JR., ADMINISTRATIVE LAW TREATISE 530–43 (4th ed. 2002) (discussing scope of statutory oral hearing rights in agency adjudications).

[2] *See id.* at 415–24.

[3] *See generally id.* at 424–63 (discussing notice-and-comment process in informal rulemakings).

[4] On agency use of informal public meetings, see, for example, Daniel J. Fiorino, *Citizen Participation and Environmental Risk: A Survey of Institutional Mechanisms*, 15 SCI. TECH. & HUM. VALUES 226, 230–31 (1990).

[5] *See* Mariano-Florentino Cuéllar, *Rethinking Regulatory Democracy*, 57 ADMIN. L. REV. 411, 423 (2005) (summarizing skeptical literature on notice-and-comment rulemaking). To be sure, some citizens do participate in rulemakings, and sometimes their comments are quite sophisticated. *See id.* at 468–72, 486–89. But these facts are consistent with the propositions that participants are self-selected, rather than a randomly selected sample of the public, and that even those who participate are not generally very informed.

in political theory.[6] The tradition envisions a particular kind of citizen participation: public-spirited, concerned with advancing the public good rather than personal preferences or interests. A variety of concrete formats have been proposed for civic republican citizen deliberation, including citizen advisory committees or review panels, citizen juries, and "deliberative polling."[7]

But federal agencies have shown little interest in the deliberative democrats' proposals. Advisory committees are frequently convened, but (at least at the federal level) usually consist of technical experts or interest group representatives rather than ordinary citizens.[8] Citizen jury and deliberative polling formats are more ambitious than citizen advisory committees. These formats require a highly structured process led by convenors, whereby representative citizens become informed, deliberate, and then vote or state their views. Citizen juries or deliberative polls have been used only very occasionally by U.S. governmental entities, state or federal.[9]

---

[6] The literature on civic republicanism and deliberative democracy is vast. For representative contributions, see Cass R. Sunstein, *Beyond the Republican Revival*, 97 YALE L.J. 1539 (1988), and DELIBERATIVE DEMOCRACY (James Bohman & William Rehg eds., 1997).

[7] James Fishkin is a leading proponent of "deliberative polling." *See* JAMES S. FISHKIN, THE VOICE OF THE PEOPLE 161–76 (1995) [hereinafter FISHKIN, THE VOICE OF THE PEOPLE]; James S. Fishkin, *Toward Deliberative Democracy: Experimenting with an Ideal*, *in* CITIZEN COMPETENCE AND DEMOCRATIC INSTITUTIONS 279 (Stephen L. Elkin & Karol E. Soltan eds., 1999); Robert C. Luskin et al., *Considered Opinions: Deliberative Polling in Britain*, 32 BRIT. J. POL. SCI. 455 (2002) (article co-authored by James Fishkin). Ned Crosby is a leading proponent of "citizen juries." *See* Ned Crosby, *Using the Citizens Jury® Process for Environmental Decision Making*, *in* BETTER ENVIRONMENTAL DECISIONS 401 (Ken Sexton et al. eds., 1999) [hereinafter Crosby, *Using the Citizens Jury® Process*]; Ned Crosby, *Citizens Juries: One Solution for Difficult Environmental Questions*, *in* FAIRNESS AND COMPETENCE IN CITIZEN PARTICIPATION 157 (Ortwin Renn et al. eds., 1995); Ned Crosby et al., *Citizens Panels: A New Approach to Citizen Participation*, 46 PUB. ADMIN. REV. 170 (1986). Other scholarship on citizen advisory committees, citizen juries, deliberative polling, and similar formats includes: ANNA COOTE & JO LENAGHAN, CITIZENS JURIES: THEORY INTO PRACTICE (1997); FAIRNESS AND COMPETENCE IN CITIZEN PARTICIPATION, *supra*; THE POLL WITH A HUMAN FACE (Maxwell McCombs & Amy Reynolds eds., 1999); Jonathan Aldred, *Citizens and Wetlands: Evaluating the Ely Citizens' Jury*, 34 ECOLOGICAL ECON. 217 (2000); John S. Applegate, *Beyond the Usual Suspects: The Use of Citizens Advisory Boards in Environmental Decisionmaking*, 73 IND. L.J. 903 (1998); Thomas C. Brown et al., *The Values Jury to Aid Natural Resource Decisions*, 71 LAND ECON. 250 (1995); Wendy Kenyon et al., *Citizens' Juries: An Aid to Environmental Valuation?*, 19 ENV'T & PLAN. C: GOV'T & POL'Y 557 (2001).

[8] *See* Applegate, *supra* note 7, at 922, 927–28. Citizen advisory committees appear to be somewhat more widely used in the states. *See* Frances M. Lynn & Jack D. Kartez, *The Redemption of Citizen Advisory Committees: A Perspective from Critical Theory*, *in* FAIRNESS AND COMPETENCE IN CITIZEN PARTICIPATION, *supra* note 7, at 87, 88–90.

[9] On the use of citizen juries, see Crosby, *Using the Citizens Jury® Process*, *supra* note 7, at 404. On deliberative polls, see Luskin et al., *supra* note 7, at 461.

This Article takes a different tack. It describes a set of participatory devices that are much more widespread in actual governmental practice than the formats proposed by deliberative democrats, yet have eluded sustained theoretical attention. I will call these "welfare polling formats," or "welfare polls" for short, to be contrasted with "policy deliberation formats" such as citizen juries, citizen panels, or deliberative polls. The thrust of this Article is that welfare polls have substantial instrumental value by providing information about well-being that is relevant in a wide range of governmental contexts.[10]

Welfare polls ask ordinary citizens about well-being, not policy. Citizens are not asked for their all-things-considered views about what government should do. Rather, they are posed questions that will help measure the impacts of governmental choices on a scale of human well-being. These well-being questions are, in crucial ways, narrower and less ambitious than those contemplated by the deliberative democrats. They do not ask citizens to bracket their own interests and preferences. They do not ask citizens to take a stance about the appropriate goals of government, for example, about the tradeoff between equity and efficiency, or welfare and rights. Instead, welfare polls start from the premise that welfare matters to governmental choice. The participants in these surveys are brought into the conversation not to interrogate this premise—to rethink normative fundamentals—but rather (more narrowly) to help determine what exactly well-being means.

Welfare polls can use a variety of metrics and can inquire about different aspects of welfare. Currently three specific formats are dominant: *contingent valuation* (CV) surveys, which ask citizens for money valuations and have been applied to value fatality risks, health, psychological states, recreation, environmental goods, artistic and cultural goods, and virtually every other aspect of welfare; *quality-adjusted life year* (QALY) surveys, which ask citizens[11] to rank health states on a nonmonetary scale with one representing perfect health and zero death; and *happiness* surveys, which ask citizens to rank their own "happiness" or "life-satisfaction" on various nonmonetary scales.[12]

Each of these techniques has generated vast scholarly literatures.[13] Further, the results of CV surveys and, increasingly, QALY

---

[10] *See infra* Part IV.

[11] As clarified below, QALY surveys are sometimes administered to doctors or other health care professionals, but citizen surveys are also common. *See infra* text accompanying notes 27–28.

[12] *See infra* Part I (describing these formats).

[13] *See infra* notes 20–22, 25, 31.

surveys play a substantial role in agency policy analysis. The Environmental Protection Agency (EPA), the Army Corps of Engineers, the Forest Service, and a number of other federal agencies that regulate health, safety, or environmental hazards, or fund projects with environmental impacts, have long relied on the results of CV surveys in cost-benefit analyses. The Food and Drug Administration (FDA) has pioneered the practice of incorporating QALY-based valuations into cost-benefit analysis—taking a QALY valuation of a health state and translating that into a dollar figure through a conversion factor. Pursuant to regulations promulgated by the Department of the Interior (DOI) and the National Oceanic and Atmospheric Administration (NOAA), CVs are currently employed for purposes of natural resource damage assessments. They are also used by various agencies in preparing environmental impact statements under the National Environmental Policy Act (NEPA).[14]

Happiness surveys, the third leg of the welfare polling triad, have yet to play the role in U.S. governmental practice that CV and QALY surveys do. But plausible scholarly proposals for happiness-based policy analysis have been advanced. In any event, it is clear that *some* of the main welfare polling formats (CV and QALY surveys) already figure importantly in administrative decisionmaking—much more so than the policy deliberation formats favored by deliberative democrats—and that the potential role of welfare polls is yet larger.

So why has no one written about welfare polls? More precisely, why has no one written about welfare polling *as such*? There is plenty of writing about QALYs, CVs, and happiness surveys. Each of these particular techniques has generated a vast outpouring of primary and secondary work. But the writing almost always focuses on a particular kind of welfare poll, rather than viewing QALYs, CVs, and happiness surveys as instantiations of a more general category; and it is almost always done by applied economists rather than political, legal, or moral theorists. No one has described and evaluated welfare polls as a generic structure for citizen participation in governance.

Why not? Political, legal, and moral theorists tend not to be welfarists. Economists, who are welfarists, tend to be more interested in modeling and measurement than in political, legal, or moral theory. Some administrative law scholars are welfarists—but the ones who care most about citizen participation often are not and assume that

[14] *See infra* Part II (describing use of CVs and QALYs by federal agencies).

the novel forms of participation worth discussing should be modeled on the civic republican ideal.[15]

So, welfare polling has slipped under the theoretical radar. This Article aims to redress that. Part I describes the existing polling techniques: CV surveys, QALY surveys, and happiness surveys. Part II surveys the range of contexts in which these techniques currently inform decisionmaking by administrative agencies, and suggests other possible uses.

Part III examines a range of technical, but critical, problems in designing welfare polls. Welfare pollsters, like their civic republican counterparts, need to overcome the obstacle of rational apathy. More generally, there are a range of valuational and communicative conditions that must be fulfilled for welfare polls to have substantial informational content: Respondents must be sufficiently well informed; their preferences must not be distorted; they must be focused on well-being (i.e., self-interested); they must engage in mental effort; they must understand the question asked; they must answer the question truthfully, or at least in a way that is correlated with the truthful answer; and they must constitute a sufficiently representative sample of the public at large. These conditions pose critical, practical problems for those who conduct welfare polls. Even more fundamentally, they are critical to a normative evaluation of welfare polling as a practice.

That normative evaluation is undertaken in Part IV. Synthesizing the material from previous parts, I provide a normative case for the use of welfare polls, grounded in the moral and legal relevance of well-being. It defends a moral framework, "weak welfarism," which recognizes that non-welfare considerations may play a role in moral evaluation, but insists that moral factors which make essential reference to well-being are an integral *part* of morality. These latter factors include, at a minimum, overall well-being, and also plausibly include distributive factors framed in terms of the distribution of well-being. Law and morality are not, of course, identical; but Part IV explains why administrative officials are in fact *legally* required or permitted (not just morally required) to attend to well-being.

But even if well-being information *is* relevant to governmental decisionmaking, why should that information be secured through surveys? And aren't welfare polls inconsistent with policy deliberation formats? Part IV addresses both of these questions. It argues,

---

[15] *See, e.g.*, Jim Rossi, *Participation Run Amok: The Costs of Mass Participation in Deliberative Agency Decisionmaking*, 92 NW. U. L. REV. 173, 203–07 (1997) (discussing rise of civic republican theory within administrative law scholarship).

against the revealed-preference tradition in welfare economics, that social planners have reason to rely on surveys and not just behavior in estimating individual valuations. It further argues, against the deliberative-democratic tradition in political theory, that these surveys need not always ask citizens to put on the hat of policymaker. Welfare polling complements, rather than displaces, policy deliberation formats. The two capture different kinds of citizen judgment, rather than being mutually exclusive.

Part V looks to the future. It describes a variety of novel formats with which welfare pollsters should experiment. The trio of CV, QALY, and happiness surveys will surely remain dominant for some time, but should be supplemented with new approaches.

I
WELFARE POLLS: EXISTING FORMATS (CV, QALY, AND HAPPINESS SURVEYS)

Welfare polls or surveys, as I conceptualize them, have a number of defining features. The respondents are lay people, not experts. The respondents are not queried about their policy views or moral judgments, but instead are asked to evaluate some human's life, or a change in some human's life (the respondent's own life, or someone else's), with respect to well-being. And the respondents are invited to express this judgment quantitatively, in terms of some numerical scale.

This definition highlights what is both distinctive and normatively attractive about certain existing survey practices, namely, CV, QALY, and happiness surveys. At the same time, it leaves much room for experimentation and improvement. What information should respondents be given? How should they be debiased? Should they be asked to think about their answers alone, or to deliberate together about well-being? What scale should be used? These sorts of questions will be examined in Parts III and V. The current welfare polling formats may be far from optimal. Still, it is important to see that the project of welfare polling is not a utopian one. The project is already well underway, with CV, QALY, and happiness surveys as the leading examples.

A. Contingent Valuation (CV) Surveys

CV surveys were invented by environmental economists in the 1960s.[16] They are now conducted not just for ecological goods, but for

[16] Good reviews of the CV technique include: IAN BATEMAN ET AL., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES (2002); A. MYRICK FREEMAN III, THE MEASUREMENT OF ENVIRONMENTAL AND RESOURCE VALUES 161–87 (2d ed. 2003);

virtually every aspect of well-being: recreation, noise, smell, visibility, fatality risks, health states, psychological states, cultural amenities, and aesthetic values.[17] Respondents are selected members of the citizenry or some subset of the citizenry—for example, the population that uses some amenity or that is exposed to some hazard—and may be randomly or nonrandomly selected, depending on the study design.[18] Mail, telephone, and in-person surveys are all common, and the rise of the Internet has created yet another possible way to administer CV surveys.[19] Surveys are typically undertaken by academic researchers, usually applied economists, or by government agencies or contractors working for agencies.

It is estimated that thousands of CV surveys have been undertaken.[20] The secondary literature is correspondingly large.[21] Whole

---

ROBERT CAMERON MITCHELL & RICHARD T. CARSON, USING SURVEYS TO VALUE PUBLIC GOODS: THE CONTINGENT VALUATION METHOD (1989); Kevin J. Boyle, *Contingent Valuation in Practice*, *in* A PRIMER ON NONMARKET VALUATION 111 (Patricia A. Champ et al. eds., 2003). Two important anthologies are VALUING ENVIRONMENTAL PREFERENCES: THEORY AND PRACTICE OF THE CONTINGENT VALUATION METHOD IN THE US, EU, AND DEVELOPING COUNTRIES (Ian J. Bateman & Kenneth G. Willis eds., 1999), and CONTINGENT VALUATION: A CRITICAL ASSESSMENT (Jerry A. Hausman ed., 1993). Helpful recent literature reviews are L. Venkatachalam, *The Contingent Valuation Method: A Review*, 24 ENVTL. IMPACT ASSESSMENT REV. 89 (2004), and Richard T. Carson et al., *Contingent Valuation: Controversies and Evidence*, 19 ENVTL. & RESOURCE ECON. 173 (2001). Two other excellent reviews, published after this Article was drafted, are HANDBOOK ON CONTINGENT VALUATION (Anna Alberini & James R. Kahn eds., 2006), and Richard T. Carson & W. Michael Hanemann, *Contingent Valuation*, *in* 2 HANDBOOK OF ENVIRONMENTAL ECONOMICS 821 (Karl-Göran Mäler & Jeffrey R. Vincent eds., 2005).

[17] *See, e.g.*, Richard T. Carson et al., *Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods*, 72 LAND ECON. 80, 81 (1996) (reviewing large bibliography of contingent valuation papers, and noting that "[t]he goods valued are various forms of recreation (most outdoor), changes in health risks, and changes in environmental amenities such as air pollution, noise pollution, water pollution, or parks"); Maureen L. Cropper, *Has Economic Research Answered the Needs of Environmental Policy?*, 39 J. ENVTL. ECON. & MGMT. 328, 332, 335, 338, 340 (2000) (discussing use of CV surveys to value fatality risks, morbidity, water quality, and visibility); Matthew D. Adler, *Fear Assessment: Cost-Benefit Analysis and the Pricing of Fear and Anxiety*, 79 CHI.-KENT L. REV. 977, 1029–30 (2004) (discussing use of CV surveys to value fear and anxiety, either directly or as independent variable in predicting willingness-to-pay or willingness-to-accept (WTP/WTA) for other goods); B.R. Bamber & G.A. Khoury, *Contingent Valuation of Landscape*, 135 PROC. INSTITUTION CIV. ENGINEERS—TRANSPORT 185 (1999) (discussing contingent valuation of landscape); Eric Thompson et al., *Valuing the Arts: A Contingent Valuation Approach*, 26 J. CULTURAL ECON. 87, 88–89 (2002) (discussing contingent valuation of cultural amenities).

[18] *See infra* text accompanying notes 207–10.

[19] *See* BATEMAN ET AL., *supra* note 16, at 89–111; Patricia A. Champ, *Collecting Survey Data for Nonmarket Valuation*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 59, 69–80.

[20] *See* Ståle Navrud & Gerald J. Pruckner, *Environmental Valuation—To Use or Not to Use?*, 10 ENVTL. & RESOURCE ECON. 1, 8 (1997).

journals are focused on publishing primary CV studies or discussing methodology.[22]

CV surveys employ a *monetary* scale of well-being. The basic thrust of the methodology is to get the respondent to imagine some change in the world that affects her well-being, and to determine how much she is willing to pay for that change (if it benefits her) or how much she would be willing to accept in return for it (if it harms her).

CV researchers have devised various ways to elicit monetary valuations. The simplest and oldest technique is to ask, "How much are you willing to pay (or accept) in return for ____?" One variation on this technique presents the respondent with a series of "payment cards," displaying different sums of money, and asks her to point to the card that shows the amount she is willing to pay or accept. Or, in the so-called "auction" format, the respondent is presented with an initial "bid" amount ("Would you be willing to pay at least ____?"), and that amount is increased until the respondent says no. The simplest technique is quite cognitively demanding; the payment card approach helps, but the cognitive load is still substantial; and the auction technique leads respondents to anchor on the initial bid. Thus many researchers now favor yet a different approach, which is to present each respondent with a single "dichotomous choice" question—"Are you willing to pay $\$X$ for ____?"—varying the $\$X$ amounts among the survey group, and using econometric techniques to estimate an average valuation from the pattern of responses.[23]

Readers familiar with the CV approach may object to my characterization of CV surveys as welfare-focused. Current practice is to ask respondents for their willingness-to-pay or willingness-to-accept (WTP/WTA) for various outcomes, given the totality of their preferences. Typically, no effort is made to screen out moral, altruistic, or otherwise disinterested preferences. Is it not, therefore, more accurate to characterize CVs as a strange kind of policy survey, rather than a welfare poll? The answer to this important objection is that CV surveys *are* effectively welfare-focused when used to value goods

---

[21] *See* Wiktor L. Adamowicz, *What's It Worth? An Examination of Historical Trends and Future Directions in Environmental Valuation*, 48 AUSTRALIAN J. AGRIC. & RESOURCE ECON. 419, 420–25 (2004).

[22] *See* V. Kerry Smith, *JEEM and Non-Market Valuation: 1974–1998*, 39 J. ENVTL. ECON. & MGMT. 351 (2000) (discussing role of that journal in developing nonmarket valuation, particularly contingent valuation).

[23] For a discussion of CV elicitation techniques, see BATEMAN ET AL., *supra* note 16, at 135–45, and Boyle, *supra* note 16, at 134–43. Variations on the straight dichotomous choice question have also been developed; for example, the "one-and-a-half bound" and "double bounded" dichotomous choice formats. *See* BATEMAN ET AL., *supra* note 16, at 141.

(such as recreation, smell, noise, health, psychological states, or fatality risks) where self-interested preferences predominate—in contrast with environmental "nonuse" values. Further, the CV methodology might in the future incorporate discursive techniques to screen out disinterested preferences. These points are developed in Part III.[24]

B. Quality-Adjusted Life Year (QALY) Surveys

QALYs were invented by public health scholars in the 1970s and are now a cornerstone of research both in that field, and in the related field of health economics.[25] Unlike CV surveys, which are applicable to all types of welfare impacts, QALYs only measure health effects—although researchers often adopt an inclusive definition of health, encompassing pain, emotional distress, and mental handicaps as well as physical changes.[26]

QALY surveys ask respondents to place a given health state on a 0–1 scale, with zero representing death and one perfect health. These surveys are sometimes given to experts (namely, health care professionals) but expert surveys are now viewed skeptically in the field,[27] and QALY surveys of laypersons (either patients or members of the general public) are the preferred technique[28]—hence my categorization of QALYs as a kind of welfare poll. Like CV surveys, these can be done in person, by phone, through the mail, or using the Internet.[29] A number of standard techniques are used for eliciting QALY rankings: the time tradeoff method (where the respondent contemplates the prospect of living a certain amount of time T^* in the health state, and is then asked to determine the amount of time T_0 such that she would be indifferent between living T^* in the health state and living T_0 in perfect health); the standard gamble method (where the respondent is asked for the probability p that makes her indifferent between

---

[24] *See infra* text accompanying notes 106–16.

[25] For overviews of the QALY method, see Matthew D. Adler, *QALYs and Policy Evaluation: A New Perspective*, 6 YALE J. HEALTH POL'Y L. & ETHICS 1, 1–2 n.1 (2006) (citing sources). On the size of the QALY literature, see *id.* at 3.

[26] *See id.* at 48–50.

[27] *See* Paul Dolan, *Whose Preferences Count?*, 19 MED. DECISION MAKING 482, 482 (1999); G. Ardine de Wit et al., *Sensitivity and Perspective in the Valuation of Health Status: Whose Values Count?*, 9 HEALTH ECON. 109, 110 (2000).

[28] *See* Paul Dolan, *The Measurement of Health-Related Quality of Life for Use in Resource Allocation Decisions in Health Care*, *in* 1B HANDBOOK OF HEALTH ECONOMICS 1723, 1738 (Anthony J. Culyer & Joseph P. Newhouse eds., 2000) (discussing whether patients or members of the general public should be surveyed).

[29] *See, e.g.*, J. Brazier et al., *A Review of the Use of Health Status Measures in Economic Evaluation*, 3 HEALTH TECH. ASSESSMENT 1, 114–32 (1999) (listing numerous QALY surveys, including both interviewer- and self-administered studies).

living a given amount of time in the health state, and a lottery with probability *p* of living in perfect health for the same amount of time and *1-p* of dying instantly); and a simple rating task, which instructs the respondent to rank the state on a scale of zero to one hundred.[30]

C. *Happiness Surveys*

Let us turn, finally, to happiness surveys.[31] The U.S. General Social Survey, conducted annually or biannually for more than thirty years, surveys a large random sample (1500 or so) of the U.S. population about a range of topics. Since its inception, it has included the following question: "[T]aken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?"[32] A parallel large-scale survey conducted several times a year in European Union member states, the Eurobarometer Survey Series, asks "[O]n the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?"[33] Similar questions have been asked in a host of nonperiodic surveys, conducted by academics, governments, or other organizations, in the United States and elsewhere, often involving very large samples.[34] The general format is to ask respondents to express their happiness or satisfaction with their life (or perhaps some aspect of their life) on a numerical scale, such as a scale from one to three, one to seven, or one to ten; or to subsume their happiness or life-satisfaction in one of an ordered set of categories (for example, "very satisfied," "fairly satisfied," "not very satisfied," and "not at all satisfied").[35]

---

30 On QALY elicitation methods, see *id.* at 23–56, and Dolan, *supra* note 28, at 1732–37.

31 For overviews of happiness surveys and the literature they have generated, see BRUNO S. FREY & ALOIS STUTZER, HAPPINESS AND ECONOMICS (2002); Ed Diener et al., *Subjective Well-Being: Three Decades of Progress*, 125 PSYCHOL. BULL. 276 (1999); Richard M. Ryan & Edward L. Deci, *On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being*, 52 ANN. REV. PSYCHOL. 141 (2001); and Norbert Schwarz & Fritz Strack, *Reports of Subjective Well-Being: Judgmental Processes and their Methodological Implications*, *in* WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY 61 (Daniel Kahneman et al. eds., 1999) [hereinafter WELL-BEING].

32 David G. Blanchflower & Andrew J. Oswald, *Well-Being over Time in Britain and the USA*, 88 J. PUB. ECON. 1359, 1363–66 (2004); NAT'L OPINION RESEARCH CTR., UNIV. OF CHI., GENERAL SOCIAL SURVEYS, 1972–2004: CUMULATIVE CODEBOOK, at v, 224 (2005).

33 *See* Blanchflower & Oswald, *supra* note 32, at 1367–69; European Comm'n, Public Opinion Analysis: Methodolgy, http://europa.eu.int/comm/public_opinion/description_en.htm (last visited Oct. 23, 2006).

34 *See, e.g.*, Frank M. Andrews & John P. Robinson, *Measures of Subjective Well-Being*, *in* MEASURES OF PERSONALITY AND SOCIAL PSYCHOLOGICAL ATTITUDES 61, 65–68 (John P. Robinson et al. eds., 1991); Michael Argyle, *Causes and Correlates of Happiness*, *in* WELL-BEING, *supra* note 31, at 353.

35 *See* Andrews & Robinson, *supra* note 34, at 70–73.

Psychologists pioneered happiness research and have undertaken most of these surveys and generated much of the secondary literature on happiness. But happiness has recently become a hot topic in economics. There is now a large and growing body of work by economists that analyzes surveys to identify and quantify the determinants of happiness, discusses the econometrics of these inferences, or makes policy recommendations for increasing happiness.[36]

II
WELFARE POLLS: CURRENT AND POTENTIAL GOVERNMENTAL USES

Some intellectual tools that are influential in the academy never make it into the public sector. But that is not the case with welfare polls. This Part surveys their current governmental uses and interweaves a discussion of potential uses. My focus here is on federal agencies and to that extent is underinclusive as it does not include state governments[37] or governments abroad.[38] The actual and potential uses surveyed here fall into three broad categories: First, welfare polls can be used for *policy analysis* (specifically, cost-benefit analysis, which currently incorporates both CV and QALY surveys; happiness-based cost-benefit analysis; and alternative policy analysis). Second, welfare polls can be used to *calibrate individual obligations or entitlements* that are defined in terms of the welfare impact of individuals' activities (specifically, by using welfare polls to determine natural resource damages; other sorts of damages, fines, and fees; and credits and allotments for environmental trading markets). Third, welfare polls can be used to shape *government communications* (such as environmental impact statements, "statements of basis and purpose" accompanying rulemakings, and national well-being accounts).

---

[36] This work is summarized in FREY & STUTZER, *supra* note 31.

[37] State governments do use CV studies. *See* John B. Loomis, *Contingent Valuation Methodology and the US Institutional Framework*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 613, 618–20. Oregon relied on QALYs in a notorious episode fifteen years ago, but it appears that state governments do not use QALYs much. *See* Adler, *supra* note 25, at 3–4. As far as I am aware, happiness surveys have not yet been employed by either state or federal governments.

[38] "To date, techniques for the monetary valuation of environmental damage and benefits [in particular CVs] have been more extensively developed and applied in the United States than in Europe." François Bonnieux & Pierre Rainelli, *Contingent Valuation Methodology and the EU Institutional Framework*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 585, 593. In contrast, QALYs play a larger role in health policymaking in certain foreign governments than in the U.S. government. *See* Adler, *supra* note 25, at 3–4.

A. Policy Analysis

1. Current Use: Cost-Benefit Analysis

Since the days of President Reagan, federal executive agencies have been required by presidential order to perform full-blown cost-benefit analyses of major rules, for review by the Office of Management and Budget (OMB), and to conform all regulations to a cost-benefit standard where statutorily permissible.[39] In traditional cost-benefit analysis, the money valuations of goods are derived using "revealed-preference" techniques, which look to market prices or nontransactional behaviors. However, as elaborated in Part IV, these techniques are far from perfect,[40] and agencies now regularly incorporate the results of CV surveys into their cost-benefit analyses.

In particular, the EPA routinely relies on CV surveys when conducting cost-benefit analysis for rulemaking—and it does so not only to quantify "nonuse" values, but also "use" values,[41] in particular mortality risk, health effects, visibility, and water quality.[42] Agencies

[39] Exec. Order No. 12,291, 3 C.F.R. 127 (1982), *reprinted as amended in* 5 U.S.C. § 601 (2000); Exec. Order No. 12,866, 3 C.F.R. 638 (1994), *reprinted in* 5 U.S.C. § 601 (2000). On the methodology of cost-benefit analysis and its use by the federal government, see generally MATTHEW D. ADLER & ERIC A. POSNER, NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS (2006) [hereinafter ADLER & POSNER, NEW FOUNDATIONS]; Matthew D. Adler & Eric A. Posner, *Implementing Cost-Benefit Analysis When Preferences Are Distorted*, *in* COST-BENEFIT ANALYSIS 269 (Matthew D. Adler & Eric A. Posner eds., 2001) [hereinafter Adler & Posner, *Implementing Cost-Benefit Analysis*]; and Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165 (1999) [hereinafter Adler & Posner, *Rethinking Cost-Benefit Analysis*].

[40] *See infra* text accompanying notes 240–62.

[41] The distinction between "use" and "nonuse" values is discussed below. *See infra* text accompanying notes 106–16. As I explain there, the measurement of nonuse values with CVs is problematic because such values are likely to be grounded in moral preferences. But there is no such difficulty with use values.

[42] The following are regulatory impact analyses (obtained from the AEI-Brookings Joint Center database, see http://www.aei.brookings.org/publications), or Federal Register notices accompanying rulemakings, in which the EPA has explicitly relied upon CV studies to quantify use values. Because the Joint Center database is incomplete, and because the RIAs and Federal Register notices are not always explicit about whether CV or revealed-preference techniques were employed to estimate values, the list here is surely not comprehensive. *See* National Primary Drinking Water Regulations: Long Term 2 Enhanced Surface Water Treatment Rule, 71 Fed. Reg. 654, 732 (Jan. 5, 2006) (mortality risk); Regulatory Impact Analysis for the Stationary Internal Combustion Engine (RICE) NESHAP, at 8-24 to -25 (Feb. 2004) (mortality risk); Effluent Limitations Guidelines and New Source Performance Standards for the Concentrated Aquatic Animal Production Point Source Category, 67 Fed. Reg. 57,872, 57,913 (Sept. 12, 2002) (water quality, including use values); National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, 65 Fed. Reg. 38,888, 38,945–46 (June 22, 2000) (mortality risk); Asbestos Worker Protection, 65 Fed. Reg. 24,806, 24,817 (Apr. 27, 2000) (mortality risk); Control of Air Pollution from New Motor Vehicles: Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements, at VII-38, VII-46 to -48, VII-57 (Feb. 10, 2000) (mortality risk, chronic bronchitis,

also perform cost-benefit analysis outside the rulemaking context, and CV surveys have been used here, particularly by the U.S. Army Corps of Engineers (in evaluating public works projects) and the Forest Service (in evaluating forest plans), as well as by the Bureau of Reclamation, the Fish and Wildlife Service, and the National Parks Service.[43]

How exactly do CV studies percolate into these agencies' cost-benefit analyses? Who conducts the study? How wide will the study's scope be? The answer is, "It depends." Sometimes, an agency or its contractors will perform a CV study for a particular policy. But this can be quite expensive, and it bears emphasis that there are alternative techniques that economize on decision costs. For example, an agency might look to a study—its own or some other agency's—of a similar policy. Or the agency might break down the policy's effects into different dimensions, and turn to the academic literature for CV studies regarding each dimension. A related idea: Some CV studies inquire not just about an individual's WTP/WTA amount for some policy but about the various individual characteristics and policy effects on the individual that presumably determine the WTP/WTA amount. From these studies one might estimate a "benefits function," correlating WTP/WTA with its determinants, and apply that function to the particular policy at hand.[44]

It is not particularly surprising that CV surveys are employed by administrative agencies in performing cost-benefit analysis. After all,

---

asthma, visibility); Industrial Laundries, at 10-29, 10-59 (Aug. 18, 1999) (recreational benefits, mortality risk); Regional Haze Rule, at 9-13, 9-30, 9-37 (Apr. 22, 1999) (visibility, mortality risk, upper respiratory symptoms); National Primary Drinking Water Regulations: Disinfectants and Disinfection Byproducts, at 4-20 (Dec. 16, 1998) (bladder cancer based on CV study of chronic bronchitis); Proposed Effluent Limitations Guidelines and Standards for the Pharmaceutical Manufacturing Industry, at 7-8 (Sept. 21, 1998) (mortality risk); National Pollutant Discharge Elimination System—Proposed Regulations for Revision of the Water Pollution Control Program Addressing Storm Water Discharges, 63 Fed. Reg. 1536, 1602 (Jan. 9, 1998) (water quality, including recreational values); Activities in Target Housing and Child-Occupied Facilities, at 6-35, 6-39 (Aug. 29, 1996) (mortality risk); Effluent Limitations Guidelines, Pretreatment Standards, and New Source Performance Standards: Metal Products and Machinery, 60 Fed. Reg. 28,210, 28,261 (May 30, 1995) (mortality risk); Municipal Solid Waste Landfills: NSPS, at 12-10 (Apr. 1994) (odors); Sacramento Nonattainment Area, South Coast Nonattainment Area, and Ventura County, at VII-11 (Feb. 15, 1994) (visibility); Pesticides and Ground Water Strategy: A Survey of Potential Impacts, at 26 (Feb. 22, 1991) (groundwater contamination); Listing of Surface Coal Mines for New Source Review, at VI-16 (Sept. 1985) (visibility).

[43] *See* Loomis, *supra* note 37, at 613–18.

[44] The alternative techniques described in this paragraph are all forms of "benefit transfer," generally discussed in FREEMAN, *supra* note 16, at 453–56; Ståle Navrud, *Value Transfer and Environmental Policy*, *in* THE INTERNATIONAL YEARBOOK OF ENVIRONMENTAL AND RESOURCE ECONOMICS 2004/2005 at 189 (Tom Tietenberg & Henk Folmer eds., 2004); and Randall S. Rosenberger & John B. Loomis, *Benefit Transfer*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 445.

cost-benefit analysis employs a *monetary* scale for evaluating policies. The welfare impacts, positive and negative, of a potential policy are reduced to dollar figures; those dollar amounts are then aggregated to determine whether the policy has net costs or benefits relative to the status quo. The CV technique also uses a dollar scale, so the technique fits hand-in-glove with cost-benefit analysis.

But welfare polling formats that employ a nonmonetary scale can also feed into cost-benefit analysis: The valuations they yield can be translated into dollars via a conversion factor. This, in fact, has been the primary way that QALYs have figured in agency decisionmaking in the United States. Over the last decade, in several dozen cost-benefit analyses accompanying major rulemakings, the FDA has valued deaths or morbidity through QALY-to-dollar conversions. Loss of health or loss of life is measured in QALYs. For example, if a food substance causes 100 healthy individuals to die prematurely, each losing 30 years of life, and causes 10,000 more healthy individuals to suffer a temporary, 6-month disease that reduces health from level 1 to level 0.8, then the total QALY loss caused by the substance is 4000 quality-adjusted life years ($100 \times 30 + 10{,}000 \times 0.5 \times 0.2$). The FDA then translates the QALY measure of changes to health or longevity into a dollar figure, to be incorporated in cost-benefit analysis, using a conversion factor such as $100,000 or $300,000 per QALY.[45]

Why would the FDA be justified in employing QALY-to-dollar conversions, rather than ordinary WTP/WTA amounts elicited in CV studies or inferred from market or behavioral evidence, in undertaking cost-benefit analysis? I have addressed this question at length elsewhere.[46] The answer, very briefly, is twofold. First, certain cognitive distortions that interfere with CV studies can be circumvented by QALY surveys.[47] Second, even if elicited without distortion, WTP/WTA amounts will not be perfect proxies for welfare. For example, wealthier individuals will tend to have higher WTP/WTA amounts for a given health impact or risk of death, not because that disease or risk has a greater effect on their well-being, but because money is less useful for them. Money's marginal utility is deflated by their wealth. By contrast, QALYs are invariant to wealth. While Donald Trump's WTP to avoid a year of emphysema is likely to be vastly greater than my own, the QALY value of Donald's emphysema and mine are exactly the same.

---

[45] *See* Adler, *supra* note 25, at 57–60.

[46] *See id.* at 24–42.

[47] For more on this point, see *infra* text accompanying notes 162–63.

Thus, under certain conditions, QALY-based money valuations of health or risk can be more accurate welfarist measures than WTP/WTA amounts. Cost-benefit analysis incorporating these nontraditional money measures may be less likely to go astray—i.e., to pick policies that actually reduce overall welfare—than cost-benefit analysis incorporating the traditional, WTP/WTA measures. My work on QALYs elaborates this point and discusses how to set the QALY-to-dollar conversion factor.[48]

2. *Potential Use: Happiness-Based Cost-Benefit Analysis*

Happiness surveys, still confined to the academy in the United States, have various potential roles in governmental decisionmaking. One such role parallels the FDA's practice with respect to QALYs. Just as the FDA currently incorporates QALYs into cost-benefit analysis via QALY-to-dollar conversions, so welfare impacts valued on a happiness scale could be translated into money with a conversion factor and then fed into cost-benefit analysis.

Some academic work in this vein has been undertaken. For example, economists Bernard Van Praag and Ada Ferrer-i-Carbonell derived money valuations of the noise impact of the Amsterdam airport from happiness surveys.[49] They surveyed a random sample of individuals living near the airport, inquiring both about their happiness and about other characteristics, including noise exposure and income, and then estimated an equation explaining happiness values (the dependent variable) as a function of these characteristics (the independent variables). Happiness was negatively correlated with noise, and positively correlated with income.

The ratio of the coefficients establishes a noise-to-dollar tradeoff ratio R, which can be used to monetize the noise-reduction benefits of a policy to deal with airport noise. If a policy reduces the amount of noise by N noise units, the monetary equivalent of this reduction is RN. Since R is itself the product of the noise-to-happiness tradeoff rate and the happiness-to-dollar tradeoff rate, the Van Praag/Ferrer-i-Carbonell technique is equivalent to translating the noise-reduction effects of a policy into happiness units, and then converting those happiness units into dollars.[50] This approach for monetizing welfare

---

[48] *See* Adler, *supra* note 25, at 57–74.

[49] *See* BERNARD VAN PRAAG & ADA FERRER-I-CARBONELL, HAPPINESS QUANTIFIED 219–38 (2004).

[50] The description in the text above of the Van Praag/Ferrer-i-Carbonnell technique is not quite accurate, since in fact they estimated happiness using a functional form that allowed interaction effects between independent variables rather than a straight linear form, and measured the independent variables using a logarithmic scaling. But the simpli-

impacts—let's call it "happiness-to-dollar conversions"—is directly analogous to the FDA's technique of valuing health effects on a QALY scale and then converting the QALY amounts into dollars.

Andrew Clark and Andrew Oswald have generalized the Van Praag/Ferrer-i-Carbonell approach. With data from the British Household Panel Survey, they estimate a happiness function that depends both on income and other characteristics, and use that function to calculate happiness-to-dollar conversion amounts for changes in employment, health, and marital status. And they point out that the method, in principle, is applicable to any life change—the relevant characteristics simply need to be documented, along with happiness levels and income, in a happiness survey.[51]

The rationale for using happiness-to-dollar conversions, in lieu of or in addition to traditional WTP/WTA amounts,[52] as an input to cost-benefit analysis is parallel to the QALY case. Consider the example of noise. Being exposed to noise is not a health impact, and is not picked up by current QALY surveys; thus, the QALY-to-dollar technique is unavailable. The variation in housing prices as between noisier and quieter neighborhoods might be used to estimate WTP/WTA for noise, but this revealed-preference technique will be accu-

---

fied description communicates the basic idea of their technique. Imagine that our survey provides data on respondents' happiness as well as income, noise exposure, and other characteristics, and we use the survey data to estimate an individual's happiness level as a simple linear function of income, noise, and other characteristics. The coefficient on the noise variable is $-H$, which means that reducing noise by one unit increases happiness by H units. The coefficient on the income variable is D, which means that increasing income by one dollar increases happiness by D units. Define R, the noise-to-dollar conversion rate, as H/D. The Van Praag/Ferrer-i-Carbonnell technique monetizes a noise reduction of N units by multiplying it by R. But note that $RN = (H/D) \times N = (HN)/D$. So this is the same as translating the noise reduction into happiness units (HN) and then translating that happiness change into dollars (HN/D).

[51] Andrew E. Clark & Andrew J. Oswald, *A Simple Statistical Method for Measuring How Life Events Affect Happiness*, 31 INT'L J. EPIDEMIOLOGY 1139 (2002). Similarly, in a recent paper Frey and Stutzer use happiness data to monetize the effect of terrorism. Bruno S. Frey & Alois Stutzer, *Happiness Research: State and Prospects*, 62 REV. SOC. ECON. 207, 220–23 (2005).

[52] Happiness-to-dollar conversions based on happiness surveys that focus on the respondents' positive and negative affects are clearly distinct from conversions derived from CV surveys, since CV surveys inquire about an individual's WTP/WTA for a policy given the totality of its effects on the individual's well-being, while these sorts of happiness-to-dollar conversions yield the amount of money sufficient to produce an affective impact on the individual counterbalancing the affective impact of the policy. Happiness-to-dollar conversions based on life-satisfaction questions are closer to CV-based conversions, although even here the valuations may be different—for example, if CV and life-satisfaction surveys are differentially affected by various preference distortions, or if respondents to the two formats have different informational deficits. *See infra* note 60 (discussing affective versus life-satisfaction conceptions of happiness surveys); *infra* Part III.B–C (discussing informational deficits and preference distortions).

rate only if relocation costs are low.[53] CV studies are of course available, but individuals may have some difficulty determining their WTP/WTA to avoid noise, and wealth effects may skew these valuations. Even if cognitive distortions and wealth effects are not expected to be large, happiness-to-dollar conversions furnish useful, additional information for the cost-benefit analyst.[54]

3. *Potential Use: Alternative Policy Analysis*

Cost-benefit analysis is the dominant technique for policy analysis in the United States. By "policy analysis," I mean some technique for evaluating governmental choices: the choice of issuing one or another regulation, the choice of undertaking some project or doing nothing. But cost-benefit analysis is not, and should not be, the sole policy-analytic technique. Alternative methods may be legally required, or morally preferable, and the nonmonetary valuations furnished by certain welfare polling formats (such as QALYs or happiness surveys) can provide numerical inputs into these alternative methods.

Health and safety regulation provides an obvious example. Some important statutory provisions, such as the provision governing noncarcinogenic pollutants in the Clean Air Act, the food additive licensing provision in the Food, Drug and Cosmetic Act, and the toxins provision in the Occupational Safety and Health Act, preclude cost-benefit analysis. Instead, these provisions instruct the agency to protect the public health and safety, perhaps with some gross cutoff (for example, maximize public health and safety up to the point that is technologically feasible, or up to the point that firms begin to go bankrupt).[55] Although fatalities furnish a crude index of public health and safety, air pollutants, workplace toxins, and dangerous foods can cause all manner of nonfatal diseases. QALYs, which subsume both death and other health impacts, provide a better index of public health and safety than total fatalities, the level of fatality risk, or total population longevity without health adjustments. QALY-maximization (perhaps with a cost or feasibility cutoff) is therefore the most attractive way to interpret health- and safety-focused statutes such as

---

[53] *See* VAN PRAAG & FERRER-I-CARBONELL, *supra* note 49, at 220–24.

[54] *Cf.* Paul Dolan & Tessa Peasgood, Valuing Non-Market Goods: Does Subjective Well-Being Offer a Viable Alternative to Contingent Valuation? (June 2006) (unpublished paper, on file with the *New York University Law Review*) (suggesting that happiness surveys are a promising alternative to CVs for valuing nonmarket goods, but that lack of robust relationship between income and happiness undermines happiness-to-dollar conversions).

[55] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 73–80; Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. 1293, 1414–17 (2003).

those just mentioned.[56] This point, it should be stressed, applies not only to agencies such as the FDA that have traditionally interested public health researchers, but to *any* agency implementing a statute that requires the maximization of health or safety—for example, the EPA or the OSHA.[57]

A different reason for departing from cost-benefit analysis is moral rather than legal. Eric Posner and I have argued at length, in various publications, that cost-benefit analysis is morally defensible as a decision procedure implementing overall well-being.[58] But the sum of WTP/WTA amounts becomes an increasingly imperfect proxy for overall well-being as wealth effects and the variable marginal utility of money become pronounced.[59] Consider an extreme example: tax-and-transfer policy. A tax scheme that would raise $100 million dollars from the middle and upper classes and transfer that money to the poor, with $10 million in administrative costs, will be viewed by traditional cost-benefit analysis as an inadvisable project, with $10 million in net costs. The total WTP of the impoverished persons who would benefit from the scheme is $100 million, and the total WTA of the taxpayers who would fund the scheme is also $100 million. Thus the transfer itself is seen by cost-benefit analysis as a wash; add administrative costs, and the tax scheme looks like a loss. But of course, if the

---

[56] QALY-maximization should be seen as a sophisticated variant of "risk-risk" analysis: one that takes account of health quality as well as loss of life. On "risk-risk" analysis, see, for example, RISK VERSUS RISK: TRADEOFFS IN PROTECTING HEALTH AND THE ENVIRONMENT (John D. Graham & Jonathan Baert Wiener eds., 1995). To be sure, the language of the safety-focused statute might prohibit a risk-risk approach, requiring the agency instead to focus on the reduction of certain kinds of risks. *See* Whitman v. Am. Trucking Ass'ns, 531 U.S. 457, 464–71 (2001) (interpreting section 109 of Clean Air Act to preclude risk-risk approach). But even in this sort of case QALYs can be useful in quantifying the degree to which the relevant kind of risk has been reduced. And in any event, a strong case can be made that health and safety statutes should be read to permit risk-risk analysis absent a clear congressional statement to the contrary.

As noted in the text, an agency might couple QALY-maximization with a feasibility or cost cutoff. The latter sort of procedure—maximizing QALYs within a cost budget—is a kind of "cost-effectiveness analysis." Cost-effectiveness analysis is the approach to health policy choice often favored by public health scholars. *See* Adler, *supra* note 25, at 8–17.

[57] *See* Am. Trucking Ass'ns v. EPA, 175 F.3d 1027, 1039–40 (D.C. Cir. 1999) (suggesting that EPA can cure constitutional difficulties in Clean Air Act by measuring benefits of air pollution regulations using QALYs), *rev'd in part*, 531 U.S. 457 (2001); Rafael A. Ponce et al., *Quality Adjusted Life Years (QALYs) and Dose-Response Models in Environmental Health Policy Analysis—Methodological Considerations*, 274 SCI. TOTAL ENV'T 79 (2001) (discussing use of QALYs for risk assessment with heterogeneous health impacts).

[58] ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 39; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39.

[59] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 95–100, 130–31, 142–46; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 39, at 286–87, 300–05; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39, at 224.

marginal utility of money decreases with income, a transfer of money from the middle and upper classes to the poor may increase overall welfare, and the tax scheme as a whole may do so even with administrative costs.

One way to reduce the inaccuracy of cost-benefit analysis in tracking overall welfare is to adjust WTP/WTA amounts using so-called distributive weights. (These would deflate WTP/WTA amounts as the affected individuals become wealthier.) Another technique, discussed in the previous Section, is to evaluate a policy by monetizing certain of its welfare effects through QALY-to-dollar or happiness-to-dollar conversions, adding them to the WTP/WTA numbers valuing the policy's remaining effects. A third possibility is to circumvent dollars entirely and measure *all* of the policy's effects as negative or positive amounts on some nonmonetary scale. Which of these three techniques is best for maximizing overall welfare is a complicated matter, one that is beyond the scope of this Article. But it seems at least plausible that policy evaluation with a nonmonetary scale should be considered as an alternative or supplement to cost-benefit analysis, not only in contexts where cost-benefit analysis is legally precluded, but even in contexts where it is not—even in contexts where statutes permit or require agencies to maximize overall welfare.

What would the nonmonetary scale be? It could be a QALY scale, but since there are many aspects of well-being that are distinct from physical or mental health, QALY maximization is more easily justified as implementing a statutory mandate to focus on health and safety than as a proxy for overall welfare. Happiness maximization is probably a better proxy for overall welfare (although not a perfect one).[60] A number of scholars have proposed that government evaluate policies by determining which one maximizes happiness. For example, Bruno Frey and Alois Stutzer write:

> The use of measures of happiness allows for a new way of evaluating the effects of government expenditure. . . . The problem has been approached scientifically by using cost-benefit analysis. The

[60] The standard view in the happiness-survey literature is that the psychological item being measured "consists of three interrelated components: life satisfaction, pleasant affect, and unpleasant affect. Affect refers to pleasant and unpleasant moods and emotions, whereas life satisfaction refers to a cognitive sense of satisfaction with life." Ed Diener & Eunkook Suh, *Measuring Quality of Life: Economic, Social, and Subjective Indicators*, 40 SOC. INDICATORS RES. 189, 200 (1997). Maximizing positive and negative affect is not the same as maximizing overall well-being, because well-being is not just a matter of mental states. *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 28–31, 50. Maximizing the extent to which individuals are satisfied with their lives is the same as maximizing well-being only if individuals are accurate in perceiving and valuing their achievements with respect to well-being.

> benefits are the recipients' marginal willingness to pay, which is best measured by a contingent valuation analysis. . . . This method is best suited to relatively small and isolated public projects, but it breaks down when it comes to more extensive expenditure policies. Simulations using microeconomic happiness functions with a large number of determinants may be better able to evaluate the widespread effects of such policies.[61]

Thomas Griffith has suggested that happiness surveys be employed to help set tax policy. Tax policy scholarship often begins with a utilitarian "social welfare function" that maximizes the sum of individual utilities, in turn calculated as the logarithm of individual income.[62] This function is mathematically tractable and increases in income at a decreasing rate, thus justifying redistributive taxation,[63] but it is not based in any systematic research on how income translates into well-being. Griffith argues that the survey data on the correlation between income and happiness confirm the basic supposition of declining marginal income utility and should be used to determine the specific shape of the social welfare function.[64]

Admittedly, tax-and-transfer policy is more the domain of legislatures than administrative agencies. Even so, happiness-based valuation of income redistribution could have some place in agency practice—for example, at agencies that provide foreign aid or that administer domestic welfare programs. A different and more broadly applicable approach to happiness-based policy analysis builds on work by Ruut Veenhoven, a leading happiness scholar, who proposes that "happy life expectancy" be used as a metric for comparing different nations. The happy life expectancy (HLE) of a given country is simply average longevity multiplied by average happiness levels expressed in

---

61 FREY & STUTZER, *supra* note 31, at 176 (emphasis omitted); *see also* NICK DONOVAN & DAVID HALPERN, CABINET OFFICE, UNITED KINGDOM, LIFE SATISFACTION: THE STATE OF KNOWLEDGE AND IMPLICATIONS FOR GOVERNMENT 35–36 (2002), *available at* http://www.e-democracy.gov.uk/knowledgepool/default.htm?mode=1&pk_document=28 (suggesting that happiness surveys could be used for policy analysis).

62 Thomas D. Griffith, *Progressive Taxation and Happiness*, 45 B.C. L. REV. 1363, 1367–68 (2004).

63 A utility function is increasing in income if higher income maps onto higher utility, and increases at a decreasing rate if it has declining *marginal* utility—in other words, if the increase in utility for an incremental dollar of income is lower at higher income levels. Intuitively, the relation between money and well-being should have these two characteristics, and they provide a strong case for redistributive taxation—one that even the utilitarian can accept.

64 Griffith, *supra* note 62, at 1397–98. For other suggestions that happiness surveys be used to set tax policy or (relatedly) to measure poverty or inequality, see FREY & STUTZER, *supra* note 31, at 176–77, and VAN PRAAG & FERRER-I-CARBONELL, *supra* note 49, at 239–317.

surveys.[65] HLE policy analysis would be an analogue to QALY-maximization. The aim in both cases is to maximize quality-adjusted longevity. In the latter case, longevity is adjusted for health quality, using QALY surveys; in the former case, it would be adjusted for happiness, using happiness surveys.[66]

The Nobel prize–winning psychologist Daniel Kahneman suggests a policy-analytic technique which is broadly similar to HLE analysis.[67] Kahneman is skeptical of the standard happiness surveys. He prefers a moment-based format that asks people to express the quality of different momentary experiences on a numerical scale, rather than the standard format which elicits individual statements of overall happiness or life-satisfaction.[68] Still, the basic idea is the same as HLE analysis, namely happiness maximization. Kahneman's approach predicts how policies will change individual experiences; translates those changes into a happiness scale via survey data (in Kahneman's case, momentary data); and picks the policy with the biggest net happiness benefit.

B. Calibrating Individual Obligations and Entitlements

Up until this point, I have discussed the actual or potential use of welfare polling data in policy analysis: either cost-benefit analysis or some alternative policy-analytic technique. But welfare polls have additional uses. Imagine that a wrongdoer injures some person or resource. How much should the wrongdoer pay in compensation? CV surveys, or perhaps other kinds of welfare polls, can help answer the question.

---

[65] Ruut Veenhoven, *Happy Life-Expectancy: A Comprehensive Measure of Quality-of-Life in Nations*, 39 SOC. INDICATORS RES. 1, 29–31 (1996).

[66] Veenhoven himself does not propose that the HLE measure be used to evaluate policies. *See id.* at 45. But HLE maximization would seem to be at least as plausible as QALY maximization, which certainly has been proposed by many. *See* Adler, *supra* note 25, at 8–10 (discussing view of many public health scholars that QALYs should be effectiveness metric for purposes of health policy cost-effectiveness analysis, i.e., that QALYs should be maximized subject to cost constraint).

[67] Daniel Kahneman, *Experienced Utility and Objective Happiness: A Moment-Based Approach*, *in* CHOICES, VALUES, AND FRAMES 673, 689–92 (Daniel Kahneman & Amos Tversky eds., 2000).

[68] *Id.*; *see also* Daniel Kahneman, *Objective Happiness*, *in* WELL-BEING, *supra* note 31, at 3 [hereinafter Kahneman, *Objective Happiness*]; Daniel Kahneman et al., *Back to Bentham? Explorations of Experienced Utility*, 112 Q.J. ECON. 375 (1997) [hereinafter Kahneman et al., *Back to Bentham*].

1. *Current Use: Natural Resource Damage Assessment*

The Superfund statute (CERCLA),[69] the Oil Pollution Act (OPA),[70] and the Clean Water Act (CWA)[71] create a federal liability regime for oil spills and other releases of hazardous substances that harm publicly owned natural resources.[72] The statutes define natural resources broadly as "land, fish, wildlife, biota, air, water, ground water, drinking water supplies, and other such resources."[73] State or federal governments are authorized to assess damages for such harms and to sue polluters for recovery of these damages. Regulations issued by the DOI guide assessments under CERCLA and CWA,[74] and the NOAA regulates assessments under OPA.[75]

An early version of the DOI regulations contemplated the use of CV studies. This aspect of the regulations was challenged by industry but upheld by the D.C. Circuit in an important 1989 decision that solidified the role of CVs for damage assessment.[76] CVs were given a further boost in 1993 when a high-profile advisory panel convened by NOAA, including Kenneth Arrow and other eminent economists, endorsed their use if conducted in accordance with the panel's guidelines.[77] Current DOI regulations explicitly permit the use of CVs in Type B assessments (those that do not use a standard computer model).[78] While NOAA regulations do not explicitly authorize the use of CVs, an appendix lists CVs as a potential tool.[79]

CV surveys were in fact employed in the damage assessment for the 1989 Exxon Valdez oil spill; the multibillion dollar damage estimates generated by the studies helped induce Exxon's large settle-

---

69 42 U.S.C. §§ 9601–9675 (2000).

70 33 U.S.C. §§ 2701–2761 (2000).

71 33 U.S.C. §§ 1251–1387 (2000).

72 *See* Kevin R. Murray et al., *Natural Resource Damage Trustees: Whose Side Are They Really On?*, 5 ENVTL. LAW. 407, 413–18 (1999) (describing relationship between statutes).

73 42 U.S.C. § 9601(16); 33 U.S.C. § 2701(20). The statutes and implementing regulations are summarized in Charles B. Anderson, *Damage to Natural Resources and the Costs of Restoration*, 72 TUL. L. REV. 417 (1997), and Dale B. Thompson, *Valuing the Environment: Courts' Struggles with Natural Resource Damages*, 32 ENVTL. L. 57 (2002).

74 43 C.F.R. pt. 11 (2005).

75 15 C.F.R. pt. 990 (2005).

76 Ohio v. U.S. Dep't of the Interior, 880 F.2d 432, 474–81 (D.C. Cir. 1989).

77 *See* Report of the NOAA Panel on Contingent Valuation, 58 Fed. Reg. 4602, 4610–11 (Jan. 15, 1993).

78 43 C.F.R. § 11.83(c)(2)(vii) (2005).

79 61 Fed. Reg. 496, 498–99 (Jan. 5, 1996). The D.C. Circuit has upheld the use of CVs in Oil Pollution Act assessments. Gen. Elec. Co. v. U.S. Dep't of Commerce, 128 F.3d 767, 772–74 (D.C. Cir. 1997).

ment.[80] CVs have also been used in a number of less high-profile cases.[81]

2. *Potential Use: Damages, Fines, Fees, Taxes*

Using CVs in natural resource damage assessments is only the tip of the iceberg: Whenever monetary damages are meant to compensate for a welfare loss, CVs have a potential role. They can be used on a one-off basis to make natural resource damage assessments, as has in fact occurred under the federal regime just described. Alternatively, they might be used to generate a schedule of natural resource damages, or some other such standardized procedure. The federal regime includes a standard computer model for smaller pollution spills, and a number of states use schedules to calculate natural resource damages.[82]

Damages, of course, need not be limited to natural resources; they might compensate harms to persons, too. Of course, given tort law's solicitude for case-by-case decisionmaking, the notion of using CVs or other welfare polls to establish a damages schedule for death, physical injury, pain and suffering, and other nonpecuniary losses seems quite unrealistic.[83] More feasible is expert testimony at the damages phase that is grounded in CV surveys.[84] These surveys might also inform workers' compensation schedules, which to a limited

---

[80] Loomis, *supra* note 37, at 622; *see also* Richard T. Carson et al., *Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill*, 25 ENVTL. & RESOURCE ECON. 257, 278 (2003).

[81] Loomis, *supra* note 37, at 621; *cf.* Thompson, *supra* note 73, at 78–84 (noting that CV studies have been prepared in number of natural resource damages cases, often leading to settlement, but that in the very few adjudicated cases, courts have rejected CV evidence).

Although it appears CVs are most frequently used in this area to estimate damage to nonuse values, estimating nonuse values has not been their exclusive use and certainly need not be. *See, e.g.*, Loomis, *supra* note 37, at 621 (discussing Colorado survey of residents near hazardous mine and planned California survey to estimate both use and existence values of beaches and wetlands). The DOI regulations, in fact, prefer CVs for use values as opposed to nonuse values. 43 C.F.R. § 11.83(c)(2)(vii) (2005). The NOAA appendix, which suggests CVs as a possible method of analysis, mentions their use for both direct and passive use values. 61 Fed. Reg. at 499.

[82] *See* Anderson, *supra* note 73, at 457–63 (discussing standard computer model for federal natural resource damage assessments); Murray B. Rutherford et al., *Assessing Environmental Losses: Judgments of Importance and Damage Schedules*, 22 HARV. ENVTL. L. REV. 51, 76–80 (1998) (discussing state schedules). Indeed, CVs were, at least at one point, used as inputs to the federal schedule. *See* Loomis, *supra* note 37, at 620; Rutherford et al., *supra*, at 78.

[83] *See* Rutherford et al., *supra* note 82, at 74–76 (discussing personal injury scheduling abroad and proposals to do so in United States).

[84] *See, e.g.*, Brendan I. Koerner, *What's Your Happiness Worth?*, 3 LEGAL AFF. 56, 56–58 (Jan./Feb. 2004) (describing economist who testifies as expert in tort cases regarding size of hedonic damages using WTP measure).

extent cover not merely lost wages and out-of-pocket medical costs (where welfare polls would not be useful), but also pain and suffering (where they would).[85]

To be sure, tort law and workers' compensation are the province of the states. On the other hand, setting fines or fees for behavior that threatens life, health, natural resources, or other determinants of well-being falls within the jurisdiction of various federal agencies. And there is a strong economic justification for doing so where these welfare impacts are "externalities" of the behavior (i.e., where the transaction costs of negotiations between actor and cost-bearers are high). "Fines" have the flavor of a sanction that is clearly established prior to the welfare-affecting behavior; "fees" (or taxes) have the flavor of a price that is clearly established ex ante; "damages" have the flavor of a sanction that is established ex post. But these are fairly thin distinctions. The basic idea is that regulators have good reason for measuring the welfare effects of certain private behaviors on a money scale and making the actors pay those amounts (either to the state, or to the harmed parties).

CV studies, in turn, can be used to help determine what the amounts should be. In one illustrative study, Mauzerall, Kim, Sultan, and Bradford show how to calculate fees for nitrogen oxides emissions from power plants. They correlate the location of the plant with predicted morbidity and mortality effects per unit of pollution (depending on meterorological conditions and demographics at that location), and then attach a price to predicted deaths and illnesses using WTP/WTA for these effects derived from CV as well as revealed-preference studies.[86]

3. *Potential Use: Environmental Trading Markets*

Environmental trading markets (ETMs) such as "cap-and-trade" pollution regimes or wetlands banks present a similar potential application for CV studies. The basic idea of such markets is that actors are allotted limits to the amount of environmental damage they can produce, and can comply with these limits either by reducing their own harmful behavior or by purchasing credits from other actors.[87] A crucial issue for any ETM is the "currency" for the market. Are actors allotted limits, and assigned credits for reductions in environ-

---

[85] *See* Rutherford et al., *supra* note 82, at 72–73.

[86] Denise L. Mauzerall et al., *Charging NO_x Emitters for Health Damages: An Exploratory Analysis* 23–25 (CESifo Working Paper No. 1442, 2005) *available at* http://ssrn.com/abstract=706782.

[87] *See* James Salzman & J.B. Ruhl, *Currencies and the Commodification of Environmental Law*, 53 STAN. L. REV. 607, 616–22 (2000).

mental harm, that are expressed in physical units (tons of pollutant, acres of wetland)? In nonmonetary units of environmental harm (fatalities caused, acres of wetland adjusted for environmental quality in some sense)? Or perhaps in monetary units? The last practice, not typically employed in current ETMs, might seem unrealistic given the costs of CV studies.[88] But this objection overlooks the possibility of an ETM function or schedule that converts physical impacts into dollars, depending on characteristics of the environmental resources, the affected population, and so on. For example, rather than telling a polluting firm that it cannot emit more than X tons of nitrogen oxides and giving it a tradeable credit for every ton that its emissions are below the limit, the polluter might instead be told not to emit more than Y dollar-equivalents of nitrogen oxides, and given a tradeable one dollar credit for every dollar-equivalent its emissions are below the allotment. This approach, in contrast with the use of physical units (the main approach in practice), has the virtue of recognizing that polluting activities with identical physical impacts can have quite heterogeneous welfare effects.

A final important note: CVs would seem to be the natural welfare polling format for setting damages, fines, and fees, which after all are dollar amounts. But QALY-to-dollar and happiness-to-dollar conversions could in principle be used here, as in cost-benefit analysis—and with similar justification. Heinz Welsch has in fact used happiness surveys to quantify the monetary cost of nitrogen dioxide pollution, using a methodology very similar to Van Praag/Ferrer-i-Carbonell's happiness-based monetization of noise.[89]

C. *Government Communications*

1. *Current Use: Environmental Impact Statements and Rulemaking Notices*

The National Environmental Policy Act (NEPA) requires federal agencies to prepare an environmental impact statement to accompany all "proposals for . . . major Federal actions significantly affecting the quality of the human environment."[90] This brief language has made environmental scoping a pervasive aspect of federal agency decision-making, as evidenced by the large body of federal case law about the

---

[88] *See id.* at 634–35.

[89] *Compare* Heinz Welsch, *Preferences over Prosperity and Pollution: Environmental Valuation Based on Happiness Surveys*, 55 KYKLOS 473 (2002), *with* VAN PRAAG & FERRER-I-CARBONELL, *supra* note 49. Note also that QALYs or happiness units could, in principle, be used as the ETM currency rather than dollars.

[90] 42 U.S.C. § 4332 (2000).

impact-statement requirement[91] and by the sheer number of impact statements and preparatory documents. It is estimated that roughly five hundred impact statements and fifty thousand environmental assessments (preliminary documents that consider whether the federal action requires an impact statement) are issued by federal agencies every year.[92]

It is now clear that purely human impacts do not trigger NEPA. An action must have an effect on the physical environment—on "the air, land, [or] water"[93]—to come within the scope of the statute.[94] But, once triggered, the statute requires an impact statement that describes the health, economic, and social effects of the agency action that are proximately caused by its physical impact—not merely the physical impact itself.[95] The bottom line is that the kinds of effects described by the NEPA statements are multifold, including health and mortality, land transformation, changes in land use, changes to water or air quality, effects on basic services (schools, police, fire), ecological impacts, noise and vibration, effects on transportation systems, aesthetics, recreation, and even housing quality or employment.[96] CV surveys are an obvious technique for quantifying these sorts of effects, and indeed agencies have used CVs to prepare environmental impact statements under NEPA.[97]

NEPA is the quintessential example of a "communication-forcing" statute: It does not give substantive priority to environmental considerations, but simply requires agencies to publicly

---

91 *See generally* DANIEL R. MANDELKER, NEPA LAW AND LITIGATION (2d ed. Release 4, 2006) (reviewing NEPA case law).

92 NICHOLAS C. YOST, NEPA DESKBOOK 9 (3d ed. 2003).

93 Douglas County v. Babbitt, 48 F.3d 1495, 1505 (9th Cir. 1995).

94 Metro. Edison Co. v. People Against Nuclear Energy, 460 U.S. 766, 772–73 (1983); *see also* 40 C.F.R. § 1508.14 (2005) (defining "human environment" for purposes of NEPA).

95 Balt. Gas & Elec. Co. v. Natural Res. Def. Council, Inc., 462 U.S. 87, 106–07 (1983).

96 *See, e.g.*, CHARLES H. ECCLESTON, ENVIRONMENTAL IMPACT STATEMENTS: A COMPREHENSIVE GUIDE TO PROJECT AND STRATEGIC PLANNING 175–82 (2000); R.K. JAIN ET AL., ENVIRONMENTAL ASSESSMENT 239–79 (2d ed. 2002).

97 Loomis, *supra* note 37, at 614–16; John B. Loomis, *Use of Contingent Values of Wildlife and Habitat Preservation in Policy and Benefit-Cost Analyses*, *in* HANDBOOK ON CONTINGENT VALUATION 292 (Anna Alberini & James R. Kahn eds., 2006); E-mail from Charles Eccleston, Member, Bd. of Dirs., Nat'l Ass'n of Envtl. Prof'ls, to author (Dec. 22, 2005) (on file with the *New York University Law Review*); *see also* Venkatachalam, *supra* note 16, at 89 ("The CV method is a widely used nonmarket valuation method especially in the areas of . . . environmental impact assessment."); DAVID JAMES, *Economic Valuation Techniques*, *in* THE APPLICATION OF ECONOMIC TECHNIQUES IN ENVIRONMENTAL IMPACT ASSESSMENT 63, 86–90 (1994) (discussing potential use of CVs in environmental impact assessment); RICHARD K. MORGAN, ENVIRONMENTAL IMPACT ASSESSMENT: A METHODOLOGICAL PERSPECTIVE 230–32 (1998) (same). It is difficult to quantify how frequently CVs are used under NEPA, because there is no searchable database of impact statements.

communicate environmental effects, which might be useful insofar as it forces the agency to give those effects the weight required by existing statutes, or mobilizes political action by interested groups to amend the statutes. An even broader communication-forcing statute is § 553(c) of the Administrative Procedure Act, which generally obliges an agency to provide a public "statement of basis and purpose"—typically in the Federal Register—when it enacts legally binding rules.[98] When welfare polls figure in internal agency deliberations preceding the enactment of a rule, the § 553(c) statement accompanying the rule may well discuss the polls.

The most frequent cases involve cost-benefit analysis—because, to date, the leading function of welfare polls has been to inform cost-benefit analysis. For example, the FDA's Federal Register statements frequently use QALYs to describe the health effects of rules, since the FDA frequently incorporates QALY-to-dollar conversions in its cost-benefit analyses.[99] Or the EPA, having relied on a CV survey in conducting its cost-benefit analysis of a rule, may then publish this analysis in the Federal Register.[100] The communicative role of welfare polls in this context flows from their function in policy analysis, but is conceptually distinct. It is easy enough to imagine a nonpublic process of policy analysis, as indeed can occur for decisions that are not § 553(c) rules or covered by NEPA or some other communication-forcing mandate.

2. *Potential Use: Other Government Communications*

Government communications to the public are multifold, including but hardly limited to communications that describe proposed policies. Consider communications about governmental agendas, structures, or laws, or communications about the world (the state of the polity, say). GNP reports are an obvious example of the latter. Because welfare is morally and legally relevant in many contexts, it will often be appropriate for government communications to include facts about well-being, potentially bringing welfare polls into play.

---

[98] 5 U.S.C. § 553(c) (2000). Rules of "agency organization, procedure, or practice" are exempted from the notice-and-comment requirement, *id.* § 553(b), as are the categories of legislative rules described in 5 U.S.C. § 553(a). However, in the case of rules relating to "public property, loans, grants, benefits, or contracts," *id.* § 553(a)(2), agencies often voluntarily choose to follow the § 553(c) procedures. PIERCE, *supra* note 1, at 505.

[99] *See* Adler, *supra* note 25, at 58 n.195 (citing Federal Register statements where FDA has used QALY-to-dollar conversions).

[100] *See supra* note 42 (citing Federal Register statements where EPA relied on CVs for cost-benefit analysis).

I will not attempt to discuss these potential communicative functions systematically, but here are some exemplary proposals. Kahneman and co-authors propose the creation of "national well-being accounts," analogous to GNP. Total well-being would be calculated based on the time spent by U.S. citizens in different activities, multiplied by the happiness measures for those activities, as derived using Kahneman's experiential surveys:

> The goal of public policy is not to maximize measured GDP, so a better measure of well-being could help to inform policy. Here we propose measuring national well-being by weighting the time allocated to various activities by the subjective experiences associated with those activities. . . .
>
> The [national well-being account] can be used to summarize the average affective well-being of a population. Three potential uses are the following: (i) Changes in well-being in a country over time can be tracked . . . . (ii) For subpopulations (e.g., rich vs. poor) at a given time, differences in well-being can be attributed to [differences in time allocation plus differences in affect from a given activity]. (iii) Differences in well-being between countries can likewise be compared and decomposed.[101]

The psychologist Ed Diener, one of the leading happiness scholars, has a parallel proposal for a "national index of subjective well-being," which would incorporate data from more traditional happiness surveys.[102] Finally, numerous scholars propose environmental accounts that would track the state of the environment, and some have suggested that CVs could be used in preparing these.[103]

III
DO WELFARE POLLS PROVIDE SUBSTANTIAL INFORMATION ABOUT WELFARE?

The preceding Part described a range of contexts in which government officials currently do, or potentially might, rely on valuations derived from welfare polls. Whether officials ought to do so, of course, depends on the informational content of these polls. Do they indeed provide substantial evidence about human well-being?

Think of the worry this way. As Eric Posner and I have argued at length elsewhere, well-being consists in the satisfaction of preferences

---

101 Daniel Kahneman et al., *Toward National Well-Being Accounts*, 94 AM. ECON. ASS'N PAPERS & PROC. 429, 433 (2004).

102 Ed Diener, *Subjective Well-Being: The Science of Happiness and a Proposal for a National Index*, 55 AM. PSYCHOLOGIST 34, 40 (2000).

103 *See, e.g.*, Navrud & Pruckner, *supra* note 20, at 15–16 (discussing possibility of environmental accounting, including use of CVs).

that are self-interested (this rules out moral and other disinterested preferences) and that survive some degree of idealization.[104] Combining these conditions on preferences with conditions for surveys to evidence the preferences that respondents actually have, one might worry (1) that the preferences driving the survey are disinterested. One might also worry that, even if they are self-interested, the preferences are nonideal in the sense of being (2) poorly informed, (3) distorted by cognitive bias, or (4) not the result of sufficient mental effort. Further, the respondent's preferences might be self-interested and sufficiently idealized, but there may be slippage between the preferences and the answer provided in the survey, either because (5) the respondent is behaving strategically and not answering the question truthfully, or (6) is answering a different question from the one literally asked by the survey. Finally, it might be objected that, even if each and every respondent in the survey has truthfully revealed her self-interested and sufficiently idealized preferences, the sample of respondents is (7) not representative of the population that will be affected by the policy that the survey's numbers will inform.

There is in fact a large scholarly literature concerning the validity of CV, QALY, and happiness surveys, and most of the specific objections raised in this literature fall under one of the seven headings just stated. This Part surveys these difficulties. I conclude that none disables the enterprise of welfare polling but that many point to ways in which simplistic polling formats should be improved.

A different worry is philosophical: Perhaps I have specified the idealizing conditions for valuation incorrectly. Arguably, the respondent's preferences might be self-interested, undistorted, fully informed, and the product of sufficient mental effort and *still* not track her well-being. In particular, "objective list" theorists of well-being might insist that individual interest consists in the attainment of goods such as accomplishment or friendship, not in the satisfaction of preferences (even ideal ones). This is an important concern that I have grappled with at length in other work and will only briefly address here.[105] First, the distance between a full-information prefer-

---

[104] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 28–39; Matthew D. Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 FLA. ST. U. L. REV. 241, 264–67, 289–302 (2000); Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 39; Adler, *supra* note 25, at 17–30; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39, at 197–204.

[105] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 28–39, 51–52; Adler, *supra* note 104, at 264–67, 297–302; Adler, *supra* note 55, at 1303–10.

A third standard account of well-being is hedonism, which construes well-being as pleasure and the avoidance of pain or (somewhat more inclusively) the satisfaction of the subject's preferences regarding her own mental states. This theory, even in the more inclu-

entialist and an objectivist account of well-being may be very small; quite plausibly, objective goods are just those features of life-histories that individuals who have full information and satisfy other idealizing conditions generally *converge* in preferring. Second, even if it is possible for an individual's self-interested, undistorted, fully informed, and effortful preferences to diverge from her well-being, there is surely some substantial correlation between the two. So even the objectivist should concede that surveys that satisfy the idealizing conditions surveyed in this Part will provide substantial (if not perfect) information about well-being.

But it is an open question whether surveys *can* indeed be designed to sufficiently approximate those idealizing conditions: to screen out disinterested or distorted preferences, provide respondents with sufficient information, induce them to make a mental effort, and ensure they understand the question and answer it truthfully—and to do all this with a sufficiently representative sample of respondents. These are genuine obstacles to the enterprise of welfare polling—even for the preferentialist about well-being, let alone the objectivist. It is these obstacles that are surveyed in this Part.

A. Moral and Other Disinterested Preferences

Much of the scholarly criticism of CV surveys concerns moral preferences.[106] This is of particular concern when it comes to the valuation of the environment. Environmental economists distinguish between "use" values and "nonuse" values. A subject's use value is her WTP/WTA for an impact on some part of the environment with which she physically interacts, such as a park that she visits or a lake in which she fishes. Her nonuse value is her WTP/WTA for some environmental change that does not physically affect her, such as degradation in some wilderness area that she never plans to visit, or the extinction of an endangered species that she has never seen.

---

sive version, is too narrow as a full account of well-being, because events and states outside the subject's mind can be intrinsically good or bad for her welfare. *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 28–31; *supra* note 60.

106 For critical scholarship that points to the role of moral or otherwise "noneconomic" preferences in producing CV values, and the related problem of scope or embedding effects, see, for example, Peter A. Diamond & Jerry A. Hausman, *Contingent Valuation: Is Some Number Better than No Number?*, 8 J. ECON. PERSP. 45 (1994) [hereinafter Diamond & Hausman, *Is Some Number Better than No Number?*]; Peter A. Diamond & Jerry A. Hausman, *On Contingent Valuation Measurement of Nonuse Values*, *in* CONTINGENT VALUATION, *supra* note 16, at 3; and Daniel Kahneman et al., *Economic Preferences or Attitude Expressions? An Analysis of Dollar Responses to Public Issues*, 19 J. RISK & UNCERTAINTY 203 (1999). This scholarship is surveyed in Carson et al., *supra* note 16, at 177, 181–84.

CVs for environmental nonuse values often display certain anomalies. First, stated valuations are often extreme. Respondents claim an infinite WTA for the disappearance of the good, or a zero WTP to preserve it, or refuse to answer the question entirely. Second, nonextreme stated valuations are often insensitive to the scope or scale of the good.[107] For example, one well-known study by Desvousges told three different groups of respondents that some number N of migrating birds die each year by drowning in uncovered waste-oil ponds, and inquired about WTP to save the birds by putting covers on the ponds. The number N was varied among the groups: The first group was asked about WTP to save 2000 birds, the second 20,000, the third 200,000. Mean WTP values for the different surveys were virtually identical, despite the ten-fold differences in the size of the bird population saved: $80, $78, and $88.[108] Similarly, Kahneman and Knetsch found that Toronto residents were willing to pay only slightly more to clean up all the polluted lakes in Ontario than to clean up polluted lakes in one part of Ontario.[109]

Moral preferences plausibly explain, or help explain, both of these anomalies.[110] By "moral preference," I mean some sort of choice-relevant attitude that is directly based in the respondent's moral views, opinions, beliefs, and so on, rather than concern for her own interests. Moral preferences may well be lexicographic: Moral prohibitions on degrading the environment may be seen as absolute, or at least never overridable by benefit to the respondent. This explains infinite WTAs. A perceived moral prohibition on degradation might translate into an objection to the very enterprise of contingent valuation and thus "protest votes": refusals to answer, or zero WTPs, or (once again) infinite WTAs.

---

107 *See* Carson et al., *supra* note 16, at 181–84; Venkatachalam, *supra* note 16, at 95–102.

108 William H. Desvousges et al., *Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability*, *in* CONTINGENT VALUATION, *supra* note 16, at 91, 100.

109 *See* Kahneman et al., *supra* note 106, at 213 (discussing this study).

110 *See* sources cited *supra* note 106; Jonathan Baron, *Biases in the Quantitative Measurement of Values for Public Decisions*, 122 PSYCHOL. BULL. 72, 74–77, 82–84 (1997); Russell K. Blamey, *Citizens, Consumers and Contingent Valuation: Clarification and the Expression of Citizen Values and Issue-Opinions*, *in* FORESTRY, ECONOMICS AND THE ENVIRONMENT 103 (W.L. Adamowicz et al. eds., 1996); Brett R. Gelso & Jeffrey M. Peterson, *The Influence of Ethical Attitudes on the Demand for Environmental Recreation: Incorporating Lexicographic Preferences*, 53 ECOLOGICAL ECON. 35 (2005); Clive L. Spash & Nick Hanley, *Preferences, Information and Biodiversity Protection*, 12 ECOLOGICAL ECON. 191 (1995); Thomas H. Stevens et al., *Measuring the Existence Value of Wildlife: What Do CVM Estimates Really Show?*, 67 LAND ECON. 390 (1991); Arild Vatn, *Environmental Valuation and Rationality*, 80 LAND ECON. 1, 11–13 (2004).

Further, and a bit more subtly, scholars have identified a number of mechanisms by which moral preferences could produce scope-insensitivity. Respondents might understand the CV survey as asking about their willingness to make a charitable contribution in the service of their moral preferences, which is limited by their perceived budget for charity. They might get a "warm glow" from promoting those preferences, a kind of positive feeling occasioned by charitable acts, and state a valuation which is really their WTP/WTA for that warm glow, not for the object of the preferences. Or the preferences might be weakly lexicographic, ordering any degree of degradation over any money loss to the subject, up to some threshold.

Critics are right to worry about the extreme-value and scope anomalies. But it is hard to see how the role of moral preferences in fueling anomalous valuations in CV studies targeted at nonuse values would justify a general disavowal of CV surveys. Rather, it justifies a narrowing of the surveys' focus. Respondents should be focused on their self-interested preferences; moral and other disinterested preferences should be screened out.

My position, it should be stressed, is not that citizens' moral and other disinterested preferences have no role to play in the political process. That would be an absurd position. The claim, rather, is that CV studies are not the appropriate mechanism for rendering governmental choice sensitive to such preferences. Other mechanisms (for example, deliberative polls) are better. CV surveys ask the respondent to express her WTP/WTA for policies, taking into consideration her existing wealth. Because money is a "primary good" (i.e., generically useful for well-being), this is a plausible, if rough, way to capture the impact of the policies on her well-being.[111] By contrast, it is very hard to see why an individual's WTP/WTA for a policy is the correct measure (even roughly) of the degree of influence that her moral preference for the policy should have. Both the democratic procedure of one person/one vote, and deliberative procedures that (in effect) weight moral preferences in proportion to how persuasive and cogent they are, constitute procedures for incorporating citizens' moral preferences into governmental choice that have a much stronger normative grounding than the CV procedure. In any event, that is the position taken here—that citizens' disinterested preferences surely ought to influence governmental choice, but not via CV studies, which are best defended as a mechanism for measuring welfare impacts. CV

111 *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 62–123 (defending cost-benefit analysis as proxy for overall well-being); Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39, at 216–38 (same).

studies should therefore be structured to screen out moral and other disinterested preferences.

To be sure, there is a need for much research, theoretical and applied, on how to perform the screening. First, there is some fuzziness, as a theoretical matter, about where the boundary between self-interest and disinterest lies. For example, are "altruistic" preferences concerning friends or family members welfare-enhancing or disinterested?[112] But at a minimum, it seems clear that purely moral preferences fall within the disinterested category.

Second, how should surveys be structured to wash out moral preferences? Should respondents simply be reminded to direct their attention to their own well-being? Will exhortations to provide self-regarding valuations work to screen out moral preferences—or will they trigger a protest reaction by respondents? The applied economists who work on CV design have done very little to answer this second set of questions[113] because of the mistaken orthodoxy in economics that simply denies a distinction between disinterested and self-interested preferences.[114]

At a minimum, moral preferences can be screened out in a rough and ready way by limiting the survey population to those who (on our best current theory of well-being) have a welfare stake in the project or resource. In the case of environmental goods—again, the area where CVs have bumped up against moral preferences most violently—the distinction between use and nonuse works fairly well.[115]

---

112 *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 39.

113 Economists have used second-order techniques to determine whether moral preferences are influencing valuations, for example "includ[ing] questions in the survey to probe respondents' understanding and motivations." Boyle, *supra* note 16, at 145 (citing examples). It is also fairly routine to ignore extreme valuations. *See, e.g.*, Kevin J. Boyle & John C. Bergstrom, *Doubt, Doubts, and Doubters: The Genesis of a New Research Agenda?*, *in* VALUING ENVIROMENTAL PREFERENCES, *supra* note 16, at 183, 196–99. However, eliminating numerical outliers is not a full solution, since moral preferences can also produce nonextreme values, for example through a charitable contribution or "warm glow" effect.

There is a literature on the use of so-called "cheap talk scripts" to reduce "hypothetical bias" in CVs (i.e., the tendency of respondents to overstate what they actually would pay). *See, e.g.*, James J. Murphy & Thomas H. Stevens, *Contingent Valuation, Hypothetical Bias, and Experimental Economics*, 33 AGRIC. & RESOURCE ECON. REV. 182, 186–87 (2004). Researchers have not conceptualized these scripts as a way to screen out moral preferences, but in fact some of them might (inter alia) do that. *See, e.g.*, James T. Murphy et al., *An Empirical Study of Hypothetical Bias in Voluntary Contribution Contingent Valuation: Does Cheap Talk Matter?* 1–2 (U. Mass., Working Paper No. 2003-2, 2003) (describing "cheap talk script" that enjoined respondents not to articulate fair price for given good).

114 *See, e.g.*, W. Michael Hanemann, *Valuing the Environment Through Contingent Valuation*, 8 J. ECON. PERSP. 19, 33 (1994).

115 I say "fairly well" because nonuse values subsume not merely existence values but also option values—self-interested preferences to preserve some environmental good that

Those who do not physically interact with some environmental amenity should not be asked about their WTP/WTA for it; those who do should be asked questions that are targeted at the interaction (e.g., how much are you WTP/WTA to visit the park, see the view, etc.) and not at the sheer existence of the amenity. In point of fact, administrative agencies already implicitly do this outside the area of environmental law.[116] The Department of Agriculture does not ask animal rights activists for their WTP/WTA to have slaughterhouses closed; the Postal Service does not ask religious activists for their WTP/WTA to prevent the shipping of pornography through the mail; the FDA does not ask libertarians for their WTP/WTA to prevent the imposition of paternalistic regulations on others.

The discussion to this point has focused on the CV instrument. What about QALY and happiness surveys? QALY surveys can inquire about the respondent's own actual or hypothetical health state (as in patient or general population surveys), or about someone else's health state (as in surveys where physicians are asked to rate a patient's health).[117] In principle, the same is true of happiness states: Someone might be asked to rate her own happiness or someone else's. It is easy to see how the latter sort of QALY and happiness surveys might elicit preferences that are not welfare-focused. Health professional *H*, asked to evaluate patient *P*'s health state on a scale from zero to one, might give a number that expresses (1) the contribution that the health state makes to *P*'s well-being; (2) how healthy the state is, as a matter of pure "healthiness," detached from well-being;[118] or (3) how morally important it is to redress the state. Similarly (although the point is harder to see), psychologist *H* asked to rate *P*'s

---

the respondent does not currently use but might use in the future. Nonuse values might also incorporate bequest values, but (as with existence values) these *will* presumably be substantially moralized, involving a sense of obligation to future generations.

116 *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 126–27, 133–36; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 39, at 282; Diamond & Hausman, *Is Some Number Better than No Number?*, *supra* note 106, at 59.

117 Another example of a QALY format that asks the respondent to value someone else's health is the so-called "person tradeoff" (PTO) format, which asks about tradeoffs between programs that benefit different numbers of persons in different health states; naturally invites the respondent to make a moral judgment; and is not much used in practice. *See, e.g.*, Brazier et al., *supra* note 29, at 26–27, 39–41. Unlike CVs, PTOs may well be an appropriate way to elicit citizen moral preferences. That is not an issue I will pursue here—since survey instruments such as deliberative polls, or perhaps PTOs that focus on moral preferences, and welfare polls are complementary rather than mutually exclusive. *See infra* text accompanying notes 277–82. The critical point for my purposes is that standard QALY survey formats, namely time tradeoff and standard gamble questions that ask about the respondent's health, seem well-suited to serve as welfare polls even if physician surveys or the PTOs are not.

118 *See* Adler, *supra* note 25, at 11–13.

happiness on a scale from one to seven might answer purely as a matter of psychological intensity (in some sense), rather than in terms of the contribution that the state makes to *P*'s quality of life.

By contrast, QALY and happiness surveys that ask the respondent to rate her own health or psychological state (actual or hypothetical) on a numerical scale—like CV studies that ask about the subject's WTP/WTA for her own health or psychological state, or about her use of environmental goods—would seem naturally to invite a self-interested perspective. More research on the issue is certainly needed. It should be noted, however, that the critics of QALY and happiness studies have not identified extreme value or scope anomalies analogous to those that affect CV studies for nonuse values. The possibility of moral preferences is not a theme in the scholarly literature that is critical of QALY and happiness studies, in contrast with the literature that is critical of CV studies. This is some (admittedly preliminary) evidence that moral or otherwise disinterested preferences do not in fact substantially affect QALY and happiness studies that ask respondents about their own health or happiness states—the main variants currently in use.

B. Information

Welfare surveys will provide valuable data about well-being only if survey respondents are sufficiently informed. As the prestigious NOAA Panel on Contingent Valuation notes: "Adequate information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment."[119] But the Panel's bland advice is not very helpful here. What specific steps can welfare pollsters take to inform respondents? And how successful can these techniques be expected to be in overcoming respondents' informational deficits?

Start with CV surveys. These surveys do typically provide some information to the respondent, at least about certain attributes of the good at stake and the change in those attributes for which WTP/WTA is being elicited, and sometimes about other facts, such as substitute and complementary goods.[120]

---

[119] Report of the NOAA Panel on Contingent Valuation, 58 Fed. Reg. 4602, 4608 (Jan. 15, 1993).

[120] For discussions about the provision of information in CV surveys, see BATEMAN ET AL., *supra* note 16, at 308–10, 331–32; Boyle, *supra* note 16, at 123–33; Boyle & Bergstrom, *supra* note 113, at 193–95; Alistair Munro & Nick D. Hanley, *Information, Uncertainty, and Contingent Valuation*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 258; and Venkatachalam, *supra* note 16, at 103–05. On information provision in preference surveys generally, see John W. Payne et al., *Measuring Constructed Preferences: Towards a Building Code*, 19 J. RISK & UNCERTAINTY 243, 254–56 (1999).

> Provision of information on the item being valued is the fundamental component of a contingent valuation survey. Personal interviews have the highest ability because visual information is provided and an interviewer is available to explain the information and answer questions. A mail survey is more limited because no interviewer is present to explain the visual information. Ability to provide information in a telephone survey is much more limited because no visual information is available. Mixed mode surveys using a telephone interview after respondents have received written and visual information in the mail . . . is one way to overcome the informational deficiencies of telephone interviews.[121]

There is a literature that examines the effect of information on contingent valuation, which tends to find that information provision—or at least new information, not already known to the respondents—does shift WTP/WTA amounts.[122]

Even in the face-to-face format, the information provided in CV surveys is nothing close to the full information that an idealized-preference account of well-being requires. Outcome O is better than outcome O^* for person P only if P, under ideal conditions that include something like complete information, or at least the total amount of information that P can comprehend, prefers O to O^*. However these idealizing conditions are specified,[123] they presumably require a much richer description of the world than CV surveys actually provide.

Why the shortfall? To begin, there is a tradeoff between the amount of data provided and other desiderata, such as respondent's motivation and her success at processing the data. P, packed to the gills with information, might be bored or overwhelmed.[124]

The CV literature does not systematically discuss this important problem, namely how to optimize the amount of information given the cognitive and motivational costs of total information.[125] Part of the solution, presumably, is to use information-provision devices (such as helpful visual aids) that facilitate comprehension and processing.[126]

---

121 Boyle, *supra* note 16, at 121.

122 Munro & Hanley, *supra* note 120, at 259–61; Venkatachalam, *supra* note 16, at 103–05. For a parallel finding about the effect of information on the policy judgments of respondents to deliberative polls, see Luskin et al., *supra* note 7.

123 For a discussion, see THOMAS L. CARSON, VALUE AND THE GOOD LIFE 219–39 (2000).

124 *See, e.g.*, BATEMAN ET AL., *supra* note 16, at 309; Payne et al., *supra* note 120, at 255.

125 *See generally* Boyle & Bergstrom, *supra* note 113, at 195 (arguing for more research on information effects and fewer ad hoc practices).

126 Payne et al., *supra* note 120, at 255. *Cf.* Report of the NOAA Panel on Contingent Valuation, 58 Fed. Reg. 4602, 4612 (Jan. 15, 1993) (stating that pictures can be helpful in providing information but can also generate unwanted effects); Boyle, *supra* note 16, at 127 (same).

Surveys can also omit irrelevant information, which can interfere with the processing of relevant data. In practice, at least implicitly, agencies often do this by using CVs that redescribe the goods at stake, focusing on the attributes that (the agency believes) are directly welfare-relevant rather than merely having some causal connection with the respondents' well-being. For example, health or safety agencies that perform cost-benefit analyses of policies to reduce ambient, food, or workplace toxins do not employ CV studies that ask respondents about different concentrations of the toxins or different mitigation technologies. Rather, these studies ask about WTP/WTA for a change in fatality risk. Similarly, an environmental agency doing a cost-benefit analysis of improved hunting and fishing opportunities would probably not describe at great length the variety of ecological changes producing larger game or fish stocks, but would ask about WTP/WTA for the relevant end-result of these changes, namely increased numbers of game or fish or increased catch rate.[127]

A yet more systematic version of this idea—that the informational base for surveys should include only directly welfare-relevant information—is exemplified by the so-called "conjoint analysis" variant of contingent valuation. In this format, various dimensions along which options can vary are defined, including both money costs to the respondents and other dimensions. The respondent is then asked to choose among options described in terms of their locations on the dimensions.[128] For example, recreational users of a lake might be asked to choose between the status quo and a clean-up measure characterized in terms of the tax burden, the size of the fish population, water clarity, and whether or not the water is potable.

A very similar approach is often used in QALY valuation. In the QALY context, the problem of incomplete information is often discussed with reference to the choice between patient and general population surveys. Various noninformational considerations arguably weigh in favor of surveying the general population—for example, the fact that patients may be more prone to certain preference distortions or more likely to behave strategically.[129] On the other hand, patients—by virtue of their direct experience of the health state—will be better informed about it.

---

127 *See* Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 39, at 283; Adler, *supra* note 17, at 1009–10, 1012–13.

128 On conjoint analysis in the CV context, see generally BATEMAN ET AL., *supra* note 16, at 248–95, and Thomas P. Holmes & Wiktor L. Adamowicz, *Attribute-Based Methods*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 171.

129 COST-EFFECTIVENESS IN HEALTH AND MEDICINE 100 (Marthe R. Gold et al. eds., 1996); Dolan, *supra* note 28, at 1739.

> In assessing health states, a [general population] sample may be asked to consider a not-heretofore-experienced health state as well as to perform the unfamiliar task of comparing and rating health states against one another. The level of understanding of the nature of particular health states by members of the general public or by others who are not experiencing the health state is not always accurate . . . . Although efforts can be made to provide in-depth descriptions of the health state, lengthy descriptions can result in cognitive overload . . . .[130]

The use of conjoint analysis in general-population QALY surveys is one response to this dilemma. Consider: A citizen asked to assign a health state to a 0–1 scale might be (1) simply told the name of the state ("pancreatitis," "diabetes"); (2) given detailed information about the bodily changes that constitute the health state; (3) given some of that information, plus some information about the effect of the state on the subject's life (how painful it is, how much mobility is restricted); or (4) be provided this welfare-relevant information in a systematic way. Many QALY surveys take this last approach, using "health classification systems" to characterize health states—a direct analogue of the conjoint analysis approach to CVs.[131] For example, the Health Utilities Index, one of the most widely used health classification systems, characterizes health states as a combination of locations along eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain.[132]

Conjoint analysis is an important tool for QALY and CV surveys, but it should be stressed that the technique is no panacea. First, bringing into play all the welfare-relevant dimensions may be cognitively overwhelming for survey respondents. Second, and more fundamentally, part of the function of welfare polling formats is to help determine what the dimensions of welfare are, not merely to quantify the tradeoffs among dimensions. Conjoint analysis is no help in the former task. This observation suggests that the practice of welfare polling, ideally, should have a bifurcated structure. Many QALY or CV surveys surely should take for granted a set of welfare dimensions, using conjoint analysis or less formal techniques to focus respondents on the relative contributions of those dimensions to well-being; but other surveys should be undertaken to identify the dimensions themselves. Indeed, some survey work of this latter sort has occurred.[133]

---

130 COST-EFFECTIVENESS IN HEALTH AND MEDICINE, *supra* note 129, at 100.

131 *See* Brazier et al., *supra* note 29, at 57–81; Dolan, *supra* note 28, at 1731–32, 1744–45.

132 David H. Feeny et al., *Health Utilities Index*, *in* QUALITY OF LIFE AND PHARMACOECONOMICS IN CLINICAL TRIALS 239, 241–43 (Bert Spilker ed., 2d ed. 1996).

133 *See infra* Part V (describing WHOQOL survey).

I have focused on the tradeoff between the epistemic costs and benefits of information provision. Surveys, ideally, should evidence what people in an idealized informational, cognitive, and motivational state prefer; but providing more information may impair respondents' cognition and motivation. A different kind of cost of information provision is more prosaic.[134] Providing fuller information is expensive: In-person surveys, which do that best, are more expensive than mail or telephone surveys. This problem—the resource cost of securing or transmitting information—is of course a general one in policy analysis and is not limited to surveys.[135] Part of a solution has been already discussed: Rather than do a series of quick CV or QALY studies for particular decisions, agencies might perform a few very high-quality CV or QALY studies and incorporate their results in a multitude of decisions, for example, via schedules. The resource costs of information provision can be spread over multiple decisions.

Standard happiness surveys do not typically concern themselves with information provision: The respondent is presumed to know about her own life and is just asked to rate it. One central thrust of the critical literature is to challenge this assumption. The subject may have forgotten facts about her own life (even her experiential states), or those facts may not be present to her mind.

> When asked, "Taking all things together, how would you say things are these days?" respondents are ideally assumed to review the myriad of relevant aspects of their lives and to integrate them into a mental representation of their life as a whole. In reality, however, individuals rarely retrieve all information that may be relevant to a judgment. Instead, they truncate the search process as soon as enough information has come to mind to form a judgment with sufficient subjective certainty. Hence, the judgment is based on the information that is most *accessible* at that point in time.[136]

Kahneman's competitor proposal to the standard happiness-survey methodology[137] is an attempt to survey individuals about aspects of their experiential life about which they can be assumed to be well (indeed, perfectly) informed—namely, what a *current* experience feels like—and to circumvent their fallible memories about past experiences. This eases the informational demands of surveys and may well be an improvement on the standard format.

---

[134] *See* Munro & Hanley, *supra* note 120, at 272–78 (discussing socially and privately optimal level of information for CV surveys, given that securing information can be costly).

[135] *See, e.g.*, Maxine E. Dakins, *The Value of the Value of Information*, 5 HUM. & ECOLOGICAL RISK ASSESSMENT 281, 287–88 (1999).

[136] Schwarz & Strack, *supra* note 31, at 63 (citation omitted).

[137] *See supra* notes 67–68 and accompanying text.

C. *Preference Distortions*

The idealized preferences that constitute well-being are *rational* preferences. They must satisfy certain structural conditions. Those conditions plausibly include the axioms of expected utility theory. To be sure, the correctness of that particular theory of rationality is open to debate.[138] But *however* "preference distortions" are defined, it seems clear that respondents to welfare polls are often in their grip. And this fact about respondents (like the other features discussed in this Part, for example, that respondents may be disinterested, lack sufficient information, and so on) potentially undercuts the utility of welfare polls.

What, then, are the plausible preference distortions that characterize the respondents to CV, QALY, and happiness surveys, and how might they be addressed?

The CV literature identifies a number of major distortions. We have known since the seminal work of Tversky and Kahneman that ordinary individuals deviate from expected utility theory in processing probabilities,[139] so it is not surprising that survey respondents in particular do. Expected utility theory implies that WTP/WTA for small increments in risks should be roughly proportional to the size of the change; but many studies have found that stated WTP/WTA amounts tend to fall far short of proportionality, often changing very little in response to risk changes (a kind of scope-insensitivity).[140] Second, CV surveys regularly show large disparities between WTP and WTA, even where income effects are not in play (for example, in the well-known study where respondents endowed with a coffee mug had WTA values much higher than the WTP values of respondents not

---

138 Expected utility theory's failures as a *descriptive* theory of human action are clear. *See, e.g.*, sources cited *infra* notes 139, 142. Whether it fails as the best *normative* account is less clear. On that issue, see Matthew D. Adler, *The Puzzle of "Ex Ante Efficiency": Does Rational Approvability Have Moral Weight?*, 151 U. PA. L. REV. 1255, 1257 n.2 (2003) (citing sources presenting expected-utility account of rationality); Matthew D. Adler, *Rational Choice, Rational Agenda-Setting, and Constitutional Law: Does the Constitution Require Basic or Strengthened Public Rationality?*, *in* LINKING POLITICS AND LAW 109, 112 (Christoph Engel & Adrienne Héritier eds., 2003) (same); and *id.* at 112–13, 127–31 (discussing challenges to expected utility account).

139 *See* SCOTT PLOUS, THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 84–188 (1993).

140 *E.g.*, Baron, *supra* note 110, at 74; Jane Beattie et al., *On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 1*—Caveat Investigator, 17 J. RISK & UNCERTAINTY 5, 20–21 (1998); James K. Hammitt & John D. Graham, *Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?*, 18 J. RISK & UNCERTAINTY 33 (1999).

thus endowed).[141] These disparities are, to a substantial extent, a product of "loss aversion": Individuals frame effects as losses or gains relative to some arbitrary reference point, weighting losses more heavily than gains.[142] Even if one rejects expected utility theory as the correct account of rational choice, a strong normative case can be made that loss aversion is a kind of preference distortion.

Third, respondents evince "tradeoff resistance"—in particular, a resistance to trading off "priceless" goods such as health, life, or friendship for money. Tradeoff effects (like moral preferences) are evidenced by protest votes or scope-insensitivity for the priceless goods.[143] Whether tradeoff resistance is really a preference distortion depends on large issues about the incommensurability of welfare dimensions that I cannot pursue here, but there is a plausible case that certain aspects of tradeoff resistance are irrational.[144] Finally, a variety of other distortions have been observed in CV research—for example, a tendency to anchor on the initial bid in the auction format, or a "range bias" in the case of the payment card format.[145]

QALY surveys have parallel distortions. The "standard gamble" format assigns numbers to health states by asking respondents for a probability that makes them indifferent between a lottery over death and life, and the health state. Respondents may fail to value such lotteries in accordance with expected utility theory.[146] The "time tradeoff" format asks respondents for the time span spent in a per-

---

[141] The empirical literature on the WTP/WTA disparity is reviewed in John K. Horowitz & Kenneth E. McConnell, *A Review of WTP/WTA Studies*, 44 J. ENVTL. ECON. & MGMT. 426 (2006). There is a large theoretical literature about the sources of this disparity. Two good discussions are FREEMAN, *supra* note 16, at 43–94, and Robert Sugden, *Alternatives to the Neo-Classical Theory of Choice*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 152.

[142] Amos Tversky & Daniel Kahneman, *Loss Aversion in Riskless Choice: A Reference-Dependent Model*, 106 Q.J. ECON. 1039 (1991). It should be noted that the first two categories of preference distortion mentioned here may overlap. Loss aversion, along with probability weighting, may explain departures from expected utility theory in processing probabilities. *See, e.g.*, REID HASTIE & ROBYN M. DAWES, RATIONAL CHOICE IN AN UNCERTAIN WORLD 289–99 (2001). Still, the distortions are partly distinct, in that loss aversion does not wholly explain probability distortions (probability weighting is also part of the picture), and it affects choice under certainty.

[143] Adler, *supra* note 25, at 37–38; Baron, *supra* note 110, at 77; Payne et al., *supra* note 120, at 257–58.

[144] *See generally* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 158–66 (discussing incommensurability); Matthew Adler, *Incommensurability and Cost-Benefit Analysis*, 146 U. PA. L. REV. 1371 (1998) (same).

[145] VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 138–39; Ian J. Bateman et al., *Willingness-to-Pay Question Format Effects in Contingent Valuation Studies*, *in* BATEMAN ET AL., *supra* note 16, at 511, 512–16; Boyle, *supra* note 16, at 137–43; Venkatachalam, *supra* note 16, at 105–10.

[146] Brazier et al., *supra* note 29, at 30–34.

fectly healthy state that is equivalent to the time in a diseased state. This format demands a tradeoff between life and health, one that respondents sometimes resist.[147] The simple rating format avoids both tradeoffs and probabilities, but may be characterized by "response spreading," a kind of range bias: Respondents feel impelled to use the whole 0–1 scale, even where the health states being valued are quite similar.[148]

Finally, there appear to be substantial preference distortions in standard happiness surveys. Kahneman's work has emphasized a particular kind of distortion, the so-called "peak-end rule," which apparently determines memories of temporally extended experiential episodes: "[T]he remembered utility of pleasant or unpleasant episodes is accurately predicted by averaging the Peak (most intense value) of instant utility . . . recorded during an episode and the instant utility recorded near the end of the experience."[149] The duration of the experience is ignored. The peak-end rule might be seen as a kind of availability heuristic, and indeed Kahneman suggests that the answers to questions such as "How satisfied are you with your life now?" are generally driven by the facts about their lives that are most available to respondents.[150]

Schwartz and Strack, in a comprehensive critical review of happiness surveys, identify a number of recurrent distortions.[151] They emphasize framing effects: For example, an individual who remembers a particular positive life event and construes it as part of her current life tends to give higher answers to questions about current happiness or life-satisfaction than one who views it as part of her past life (and thus as a standard for comparison). University freshmen told to remember something good that happened "two years ago" reported greater happiness than those told to remember something "two years ago, that is, before you came to the university."[152]

How should preference distortions in CV, QALY, and happiness surveys be handled? One possibility is to employ debiasing measures. A number of CV studies have sought to redress probability distortions by employing devices to help respondents better grasp what probabilities mean, such as verbal analogies, pie charts, risk ladders, or graph

147 *See id.* at 36–39. A recent suggestion in the QALY literature is that both standard gamble and time tradeoff values are also affected by loss aversion. *See, e.g.*, Han Bleichrodt, *A New Explanation for the Difference Between Time Trade-Off Utilities and Standard Gamble Utilities*, 11 HEALTH ECON. 447 (2002).

148 Brazier et al., *supra* note 29, at 34–35.

149 Kahneman et al., *Back to Bentham*, *supra* note 68, at 381.

150 Kahneman, *Objective Happiness*, *supra* note 68, at 21.

151 *See* Schwarz & Strack, *supra* note 31, at 62–74.

152 *Id.* at 66 (emphasis omitted).

paper with blacked-out squares.[153] Probability aids have also been used in eliciting QALY values when the standard gamble format is employed.[154] Research on the efficacy of these aids in reducing probability distortions is mixed.[155]

QALY and CV researchers should consider experimenting with more intensive probability debiasing techniques, for example familiarizing respondents with the axioms of expected utility theory (or whichever competing account of rational choice under uncertainty is taken to be correct).

As for loss aversion: CV researchers have tried to reduce the WTP/WTA disparity, a product (in part) of loss aversion, by using formats in which each subject's valuation is repeatedly elicited, on the theory that familiarity with the valuation task will reduce the extent of loss aversion. Again, results are mixed.[156] Finally, it has been suggested that so-called "multi-attribute utility theory" techniques, in which respondents are prompted to think through their internal tradeoff rates for the different attributes of choices that they care about, could reduce tradeoff biases in the CV context.[157] These sorts of techniques could also be used in QALY surveys. Multi-attribute utility theory debiasing techniques would be most naturally paired with the conjoint-analysis approach to eliciting QALYs and CVs.[158]

A second possible approach to preference distortions is to change the elicitation method. For example, the anchoring bias characteristic of the auction format widely used in CV studies in the 1970s and 1980s can be eliminated by shifting to the payment-card or the open-ended question formats, and can be reduced, if not eliminated, by shifting to other formats.[159] Carthy finds that a novel "chained" method for determining WTP/WTA for the risk of death reduces probability distortions;[160] and Hammitt and Graham find the same for a novel

---

153 Phaedra S. Corso et al., *Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation*, 23 J. RISK & UNCERTAINTY 165, 169–70, 177–79 (2001).

154 *See* Brazier et al., *supra* note 29, at 25.

155 Corso et al., *supra* note 153, at 169–70, 177–79.

156 Horowitz & McConnell, *supra* note 141, at 440–42.

157 *See* Payne et al., *supra* note 120, at 257–58.

158 *See supra* text accompanying notes 128–32.

159 *See* sources cited *supra* note 145. To be sure, the open-ended and payment-card formats may trigger biases that are avoided by the auction format, such as cognitive load bias (in both cases) and range bias (in the latter case). The point here is only that certain biases are more strongly associated with certain formats; if those biases produce a particularly large degree of preference distortion, shifting to a different format may be justified.

160 Trevor Carthy et al., *On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 2-The CV/SG "Chained" Approach*, 17 J. RISK & UNCERTAINTY 187 (1999).

"indifference-risk" elicitation method that holds constant the price of a safety device and varies the risk reduction.[161]

The QALY literature offers a striking illustration of the point that a given bias may be differentially problematic for different elicitation approaches. The standard gamble technique assigns QALY values to particular health states by asking respondents for their indifference probabilities; deviations from expected utility theory will therefore directly affect standard gamble valuations. By contrast, the time tradeoff method eschews talk of lotteries, asking respondents to trade off certain health states that vary in their duration. Probability distortions are thereby circumvented.[162]

A third, more radical debiasing possibility is to change the kind of welfare polling format entirely. I have argued at some length elsewhere that QALYs may, on balance, be a better basis for measuring the health and fatality impacts of policy choices than CV surveys.[163] Part of the advantage is that QALY surveys never ask respondents to value health or life in money—a particularly demanding and emotionally laden tradeoff.

A different advantage of QALYs is that policymakers' valuations are constrained—by virtue of the additive formula used to value impacts—to be proportional to the amount of the impact. Scope-insensitivity is therefore automatically circumvented. Remember that the role of citizen responses to QALY surveys is to place health states on a 0–1 scale. Those numbers are then incorporated into policy analysis by multiplying the change in health by the duration of the change. For example, if a policy abates the effects of a toxin that causes a population of 10,000 individuals to suffer an uncomfortable respiratory condition, and does so for one month, and the value of the condition is 0.85 on a 0–1 scale, then the policymaker's valuation of the change would be $(0.15)\times10{,}000\times(1/12)=125$ QALYs. If the policy

161 Hammitt & Graham, *supra* note 140, at 47, 57–58.

162 Adler, *supra* note 25, at 41–42; Sylvie M.C. van Osch et al., *Correcting Biases in Standard Gamble and Time Tradeoff Utilities*, 24 MED. DECISION MAKING 511, 515 (2004) ("The epithet of the [standard gamble technique] as gold standard has faded during years of practice. [Time tradeoff] seems to have been accepted as a practical gold standard."). To be sure, if probability distortions are a product of loss aversion plus probability weighting, and loss aversion also affects time tradeoff values, then shifting to the time tradeoff technique will not wholly eliminate the underlying biases. But probability weighting, at least, should go away. Bleichrodt, *supra* note 147, at 450, 453–54 & tbl.1.

Some recent QALY research also experiments with novel variants of standard gamble or time tradeoff elicitation formats that may help reduce distortions. *See, e.g.*, Anne Spencer, *The Implications of Linking Questions Within the SG and TTO Methods*, 13 HEALTH ECON. 807, 807–08 (2004) (discussing two-stage, "chained" approach to eliciting values).

163 *See* Adler, *supra* note 25, at 24–42, 69–83.

abates the toxin for a year rather than a month, then the QALY formula would automatically value that impact as twelve times the month-long impact, i.e., as 1500 QALYs. By contrast, individuals asked for WTP to avoid a year rather than a month of the condition might well be in the grip of preference distortions that cause scope-insensitivity, and therefore fail to express valuations that are twelve times larger.

Kahneman's proposal similarly aims to circumvent preference distortion by shifting welfare polling formats. As discussed above, he criticizes traditional happiness surveys and proposes to use hedonic surveys to value moments of experience and then to have policy-makers (not citizens) value temporally extended experiential episodes by aggregating momentary values.[164]

D. Mental Effort

Economic theory traditionally assumes that mental operations are costless and instantaneous. Survey respondents, on this model, have preexisting preference orderings over complete outcomes; they can costlessly process information, using it to ascribe probabilities to outcomes; and they can costlessly derive preferences over goods, choices, and so on from their underlying outcome preferences and probabilities. If respondents were indeed costless computers, they might still lie about their preferences—truthtelling would still be a problem for welfare polls—but respondents would never have a reason to shirk in exerting mental effort.

But the assumption of zero-cost computation is wildly unrealistic for actual humans.[165] It is therefore quite possible that respondents to welfare surveys will economize on mental effort. The potential for

---

[164] This Section has focused on the use of debiasing techniques, the choice of elicitation technique within a general welfare polling format, and the choice of overall format, as methods for reducing preference distortions. Two other possibilities should be mentioned: (1) screening out respondents with especially distorted preferences; or (2) using calibration factors. On the first approach, see, for example, Nancy J. Devlin et al., *Logical Inconsistencies in Survey Respondents' Health State Valuations—A Methodological Challenge for Estimating Social Tariffs*, 12 HEALTH ECON. 529 (2003); Hammitt & Graham, *supra* note 140, at 48–52 (distinguishing in one survey between respondents who correctly or incorrectly answered probability comparison question, and in another between respondents who expressed high confidence or less confidence in their answers); and Murphy & Stevens, *supra* note 113, at 186 (discussing surveys that ask respondents about their degree of certainty regarding WTP/WTA). On the second, see, for example, van Osch et al., *supra* note 162 (discussing and employing methods for estimating QALY values that attempt to correct for various biases).

[165] The *locus classicus* for discussions of bounded rationality is, of course, Herbert A. Simon's scholarship. *See generally* 1 HERBERT A. SIMON, MODELS OF BOUNDED RATIONALITY (1982); 2 HERBERT A. SIMON, MODELS OF BOUNDED RATIONALITY (1982); 3 HERBERT A. SIMON, MODELS OF BOUNDED RATIONALITY (1997). For more recent work

rational apathy, exhaustively discussed by economists in contexts such as voting,[166] readily carries over to the construction of preferences and valuations, given realism about human mental abilities. Why should the participant in a QALY, happiness, or CV survey take much trouble to figure out how she ranks the health state on a 0–1 scale, how happy she is, or what her WTP/WTA for some good is, once she realizes that her particular response (truthful or not) has a vanishingly small chance of changing the governmental policy to which the survey will be an input?

The problem of insufficient mental effort—strategic laziness, as it were—is little discussed by the welfare polling literatures.[167] Scholarship about CVs almost completely ignores it, focusing a bright spotlight instead on the sister problem of strategic deception. Strategic laziness surfaces to a limited extent in some of the critical literature on happiness studies—for example, in the observation that the availability bias, which skews answers to standard happiness questions, is a heuristic device for answering such questions quickly and easily.[168]

Nor is mental effort a key focus of the general literature on survey design. There is a large subliterature on survey nonresponse.[169] But survey or individual item nonresponse—an outright refusal to answer a question—is only one manifestation of the conservation of mental effort. The psychologist Jon A. Krosnick, in an unusually good treatment of the problem of mental effort, discusses different ways in which survey respondents might "satisfice" rather than exert the mental energy required for a high-quality response, including "selecting the first response alternative that seems to constitute a reasonable answer," "agreeing with any assertion the interviewer makes," "endorsing the *status quo* instead of endorsing social change," "failing to differentiate among a set of diverse objects in rat-

---

in this area, see, for example, GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999), and ARIEL RUBINSTEIN, MODELING BOUNDED RATIONALITY (1998).

166 *See, e.g.*, DENNIS C. MUELLER, PUBLIC CHOICE III 303–32 (2003).

167 *See* Adamowicz, *supra* note 21, at 435. By contrast, the issue has been more fully discussed in the literature on policy deliberation formats. *See, e.g.*, Luskin et al., *supra* note 7, at 456–61. Deliberation, after all, involves making a substantial effort to think about some issue.

168 *E.g.*, Kahneman, *Objective Happiness*, *supra* note 68, at 21–22; Schwarz & Strack, *supra* note 31, at 63.

169 *See, e.g.*, BATEMAN ET AL., *supra* note 16, at 115–16 (discussing use of tailored design method to increase response rate and survey quality for CV surveys); DON A. DILLMAN, MAIL AND INTERNET SURVEYS: THE TAILORED DESIGN METHOD (2d ed. 2000); ROBERT M. GROVES & MICK P. COUPER, NONRESPONSE IN HOUSEHOLD INTERVIEW SURVEYS (1998); SURVEY NONRESPONSE (Robert M. Groves et al. eds., 2002).

ings," "saying 'don't know' instead of reporting an opinion," and "randomly choosing among the response alternatives offered."[170]

If (1) lazily constructed valuations were normatively on par with effortful valuations, or if (2) lazy valuations were statistically unbiased estimates of effortful valuations, then welfare pollsters would not need to worry about laziness. But neither of these premises holds true. Mental effort, like good information, is normative for welfare: Outcome O_1 is better than O_2 for P only if P, under sufficiently ideal conditions (including sufficient effort), prefers O_1. Indeed, it would be incoherent to contemplate an idealized subject who must be presented with lots of information but is free to apathetically ignore it. And (2) seems quite counterfactual. Information and measures to reduce preference distortion can surely systematically change preferences, at least for some groups and goods.[171] In other words, distorted or poorly informed preferences will not be unbiased estimates of idealized preferences, at least for some groups or goods. Therefore, preferences constructed without efforts to absorb information or to participate in debiasing will not (at least for some goods and groups) be unbiased estimates of idealized preferences.

A leading manual on CV studies, discussing the problem of nonresponse and protest responses (zero or very high WTP/WTA amounts), states:

> [A]nalysts usually make the assumption that the true WTP of non-responders [and protest responders] will be similar to that quoted by households with comparable characteristics.
>
> Following the removal of non-respondents from the sample, therefore, analysts should ensure that the characteristics of the sample have not been systematically biased. Analysts should examine the distribution of key characteristics of households in the sample (for example, household income, age profiles and access to the non-market good) and ensure that it does not differ significantly from the distribution of these characteristics in the population.[172]

This purely statistical technique may or may not be adequate to the problem of outright nonresponse, but it offers no solution to the problem of lazy responses. A solution to that problem, presumably, will mean designing surveys to trigger the motivators of mental effort. Krosnick suggests that these motivators include: the extent to which respondents "get intrinsic rewards from effortful mental exercises"; "the degree to which the topic of a question is personally important to

---

[170] Jon A. Krosnick, *Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys*, 5 APPLIED COGNITIVE PSYCHOL. 213, 215–20 (1991).

[171] *See* sources cited *supra* notes 122, 155, 156.

[172] BATEMAN ET AL., *supra* note 16, at 178–79.

the respondent"; the extent to which respondents "think that the survey in which they are participating is important and/or useful to some segment of society"; "interviewer behaviour"; the respondent's sense of "accountability" to the interviewer; and the length of the interview.[173] The substantial survey literature on reducing outright nonresponse rates will be helpful here, since the norms, emotions, interests, etc., that motivate respondents to take the initial step of participating in a survey can also presumably be deployed to motivate more effortful participation.[174]

E. Truthtelling

Welfare surveys will have epistemic value for regulators only if respondents are sufficiently rational and informed and exert enough mental effort. But these are necessary, not sufficient, conditions. After all, well-informed and rational respondents who make a mental effort might still lie to their interviewers. The problem of eliciting truthful valuations has been discussed at some length in the CV literature,[175] less so in the other literatures. Consider the simplest CV elicitation format: the open-ended question, which asks "How much are you willing to pay?" for some good. If the respondent prefers having the good together with its predicted cost to her (for example, the predicted increase in her taxes) to not having the good, then she has an incentive to overstate her true WTP—because a larger stated WTP increases the chance that the good will be provided. More precisely:

> Faced with an open-ended question [about WTP for a public good], a very large WTP response does turn out to be the optimal strategy for an agent who believes (a) the cost of the public good to the agent is fixed, (b) her true willingness to pay for the good is larger than the cost if provided, and (c) the good is more likely to be supplied the larger the sum of the willingness to pay responses given by agents.[176]

---

173 Krosnick, *supra* note 170, at 223–25.

174 For example, Dillman, an influential expert on nonresponse, argues that increasing survey response is a matter of inducing "social exchange," specifically by (1) providing rewards, (2) lowering costs, and (3) establishing trust, and offers a number of concrete recommendations in each category. DILLMAN, *supra* note 169, at 14–21.

175 For two particularly important discussions, see Richard T. Carson et al., Incentive and Informational Properties of Preference Questions (Feb. 2000) (unpublished manuscript, available at http://weber.ucsd.edu/~rcarson/cgm.pdf) [hereinafter Carson et al., Incentive and Informational Properties]; and Robert Sugden, *Public Goods and Contingent Valuation*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 131. See Carson et al., *supra* note 16, at 189–93, for a summary of the literature.

176 Carson et al., Incentive and Informational Properties, *supra* note 175, at 28.

Reacting to this problem, some scholars have sought to identify "incentive-compatible" elicitation formats for CV surveys. A format is incentive-compatible if respondents maximize their preference-satisfaction by truthfully stating their valuations. It turns out that, under some conditions, a dichotomous choice question will be incentive-compatible. This is intuitively clear: If an individual is asked to vote in a referendum between the status quo and a single alternative, and she believes that her vote increases the chance of the selected option being implemented, then clearly she best advances her preferences by voting for the option she actually prefers. Similarly, if an individual (1) is asked, as per the dichotomous choice format, whether she prefers the status quo or, instead, having a good provided by a government agency at a cost to her of $\$X$; (2) believes that the agency will pick one of these two options; and (3) believes that her statement to the interviewer raises the probability of the agency picking whichever option she claims to prefer; then her rational response is to truthfully articulate her preference.[177]

Not only is it the case that dichotomous choice questions *can* be incentive-compatible. It also turns out, quite strikingly, that *only* dichotomous choice questions can be incentive-compatible.[178] Finally, the research on incentive compatibility shows that even dichotomous choice questions may not always prompt truthful answers.[179] Imagine the respondent is told and believes that the good will be costless to her unless she voluntarily contributes to its provision, and is asked whether she would contribute $\$X$ for the good. In that case, she has an incentive to answer the question affirmatively, even if $\$X$ exceeds her true WTP.

The apparent conclusion from this line of research is that CV survey designers can address strategic bias by using the incentive-compatible variants of dichotomous choice questions. Indeed, the prestigious NOAA panel on CVs made precisely this recommendation:

> [T]he referendum format, especially when cast in the willingness to pay mode—"Would you be willing to . . . be taxed . . . D dollars to cover the cost of avoiding or repairing environmental damage X?"—has many advantages. It is realistic: referenda on the provision of public goods are not uncommon in real life. There is no

177 *See id.* at 10–20.

178 More precisely, no other format can be incentive-compatible unless the shape of preferences is restricted. *Id.* at 11, 31–32.

179 *Id.* at 11.

> strategic reason for the respondent to do other than answer truthfully . . . .[180]

But there is a problem. Consider the fully rational, self-interested individual whose valuation of, say, cleaner air is elicited by means of this supposedly incentive-compatible question: "Would you vote in favor of a measure to produce cleaner air, to be funded through taxes that will increase your tax bill by $\$X$?" Imagine that the respondent believes the actual policy alternative on the table to be regulation of polluters, not taxation-and-spending for cleaner air, and believes that the cost of regulation to her (in higher product prices, say) will be $\$Y$ instead of $\$X$. She will then answer the question affirmatively or negatively depending on how her valuation of cleaner air compares to $\$Y$, not $\$X$.

Supposedly incentive-compatible dichotomous choice CV questions pair policy measures on the government's agenda with hypothetical cost figures that are picked by the interviewer (and, usually, varied among respondents) so as to elicit valuations in a statistically efficient way.[181] But this technique will elicit truthful valuations from a self-interested respondent only if the respondent misunderstands how the CV technique works—only if she believes the "cost" figure to be the interviewer's prediction of the measure's cost, rather than a hypothetical number—and, further, only if the respondent is gullible enough to believe this cost "prediction."[182]

Robert Sugden, mindful of this difficulty, writes: "I can see no escape from the conclusion that, *if* survey respondents are motivated solely by rational self-interest, the CV method is fatally flawed."[183] Presumably the conclusion carries over to happiness and QALY surveys, which use open-ended questions that seem even less likely than CV "referenda" to be incentive-compatible.

Does this conclusion sound the death knell for welfare polls? I suggest not. Incentive-compatibility research about surveys has asked whether fully rational individuals maximizing their preferences (specifically, in accordance with expected utility theory) and lacking a preference for truthtelling (or the prospect of sanctions for lying) would answer truthfully. To begin, individuals may not be fully rational. Preference distortions—failures of full rationality—may

---

[180] Report of the NOAA Panel on Contingent Valuation, 58 Fed. Reg. 4602, 4606 (Jan. 15, 1993).

[181] *See, e.g.*, Anna Alberini, *Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-Bound, Double-Bound, and Bivariate Models*, 28 J. ENVTL. ECON. & MGMT. 287, 287–88 (1995) (discussing optimal design of dichotomous choice surveys).

[182] *See* Sugden, *supra* note 175, at 136–37.

[183] *Id.* at 137.

actually help the polling enterprise elicit truthful valuations. Actual respondents may not realize that lying is in their interest or may not be up to the cognitive strain of keeping track of their lies.[184] As for sanctions, while formal sanctions for lying are of course unavailable in the survey context, social norms may come into play. "[T]he social setting of interviewer and interviewee [may] evoke[ ] norms of honesty."[185] Relatedly, respondents may have some preference not to lie, and the interview format can be designed to take advantage of this preference—for example, by paying the respondent a token amount (which might strengthen guilt feelings about lying).[186]

Further, even if some (perhaps large) fraction of survey respondents does lie, that does not imply that survey responses are epistemically worthless. It is a large fallacy to leap from the premise that respondents are strategically misstating their preferences to the conclusion that those misstatements have zero informational value for policymakers. For example, if the respondent has an incentive to overstate his valuations, the policymaker can infer that the respondent's true value is no higher than the stated value. Nonzero WTP responses to open-ended contingent valuation questions could be seen by policymakers as upper bounds to true WTP amounts. More generally, if stated valuation amounts are correlated with true valuations, rather than being random, then the statements will be at least somewhat useful to policymakers in updating their estimates of true valuations.

Indeed, much evidence suggests that statements about valuations in welfare polls *are* correlated with respondents' true valuations (how they truly value the good at the time of the statement). One body of research looks to correlations between stated preferences and behavior.[187] In one particularly extensive study of this sort, Carson and his co-authors performed a meta-analysis of studies that provided both CV and revealed-preference (RP) estimates of the same good. They found that the average CV/RP ratio was 0.89, 0.77, or 0.92,

---

184 *Id.* at 137; BATEMAN ET AL., *supra* note 16, at 380–81. *See generally* Chris William Sanchirico, *Evidence, Procedure, and the Upside of Cognitive Error*, 57 STAN. L. REV. 291 (2004) (analyzing value of cognitive imperfection for evidentiary processes).

185 Sugden, *supra* note 175, at 137.

186 On the use of money or gifts as incentives in surveys, see, for example, DILLMAN, *supra* note 169, at 14–21. Dillman suggests providing a "token of appreciation"—not full monetary compensation—as a step to establish the respondent's trust. *Id.* at 19–20.

187 *E.g.*, Carson et al., *supra* note 16, at 194–95; James J. Murphy et al., *A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation*, 30 ENVTL. & RESOURCE ECON. 313 (2005); Venkatachalam, *supra* note 16, at 110–12; Christian A. Vossler & Joe Kerkvliet, *A Criterion Validity Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior for a Public Referendum*, 45 J. ENVTL. ECON. & MGMT. 631 (2003).

depending on whether a complete, weighted, or trimmed sample was used, and that CV and revealed-preference measures were substantially correlated using two standard measures of correlation.[188] Similar research has been done for happiness surveys, finding correlation between survey answers and non-self-report evidence of the respondent's happiness, such as assessments by spouses, family, or friends; the duration of authentic smiles; heart rate, blood pressure, and skin-resistance measures of stress; psychosomatic illnesses; and EEG measures of brain activity.[189] There is less work in this vein on QALYs,[190] but there is "internal" evidence that responses to QALY surveys track underlying valuations to some extent. For example, intrarater reliability is high (respondents provide the same valuations over time), respondents give lower values to health states that are unambiguously more serious, and the standard gamble and time tradeoff formats correlate reasonably well.[191] All of this would be puzzling if QALY responses were simply random.

Given this evidence of the correlation between stated and true valuations, one might naturally think that stated valuations should be adjusted by a calibration factor. For example, if welfare polls on average produce CV values that are twice those evidenced by counterpart revealed-preference studies, then policymakers could apply a fifty percent discount to CV values. Indeed, a substantial body of scholarship seeks to estimate such calibration factors.[192] A note of caution

---

188 Carson et al., *supra* note 17, at 86, 91–92. A different body of scholarship focuses on comparing CV valuations for particular goods with the amounts that are actually paid for these goods in experiments. This work tends to find that the CV value is higher than the value actually paid. *E.g.*, Murphy et al., *supra* note 187, at 314. For my purposes in this Section, the crucial point evidenced by the Carson meta-analysis is the fact of correlation between CV and revealed-preference measures, not the size of the multiplier.

189 Blanchflower & Oswald, *supra* note 32, at 1360–61 (citing James Konow & Joseph Earley, The Hedonistic Paradox: Is Homo-Economicus Happier? 6 n.3 (1999) (unpublished manuscript, available at http://myweb.lmu.edu/jkonow/Hedonistic%20Paradox.pdf)). For other discussions of the correlation between survey and nonsurvey evidence of individual happiness, see, for example, Diener et al., *supra* note 31, at 278; Rafael Di Tella et al., *The Macroeconomics of Happiness*, 85 REV. ECON. & STAT. 809, 811–812 (2003); and Ed Sandvik et al., *Subjective Well-Being: The Convergence and Stability of Self-Report and Non-Self-Report Measures*, 61 J. PERSONALITY 317 (1993).

190 *See* Brazier et al., *supra* note 29, at 18.

191 Adler, *supra* note 25, at 41–42; Brazier et al., *supra* note 29, at 30–46. There is analogous "internal" evidence that CV and happiness surveys correlate with respondents' true valuations, in addition to the behavioral evidence summarized in the text. *See, e.g.*, Richard C. Bishop et al., *Contingent Valuation*, *in* HANDBOOK OF ENVIRONMENTAL ECONOMICS 629, 629–46 (Daniel W. Bromley ed., 1995); Carson et al., *supra* note 16, at 193–95; Sandvik et al., *supra* note 189, at 319–21.

192 *E.g.*, Murphy et al., *supra* note 187. NOAA once considered requiring that a fifty percent calibration factor be applied to CV values in calculating natural resource damages. *See* Navrud & Pruckner, *supra* note 20, at 13.

needs to be sounded: Calibration factors derived from the correlation between stated preferences and behavioral evidence need to be used with care, because (as elaborated below, in Part IV) behavioral evidence is no gold standard for valuation. Among other things, the actors whose behaviors undergird revealed-preference work may be poorly informed, may have distorted preferences, and may economize on mental effort. The best solution to problems of deception would be to use norms and incentives to encourage truthtelling; but if that seems unavailing, calibration factors offer a second-best approach to deriving information about underlying valuations from welfare polls.[193]

On the topic of truthtelling, it also bears note that welfare polls are just a small part of a vast survey literature that encompasses political opinion polls, censuses, psychological surveys, and consumer product research. By one estimate, "[a]bout 20 million interviews are conducted each year in the United States."[194] Many of these surveys are not incentive-compatible—either because they are "inconsequential" (the response will not change what government or other actors do) or, if "consequential," they give the respondent an incentive to lie.[195] Still, there remains great demand for these surveys by politicians, psychologists, and others.[196]

Consider, in particular, consumer product research—the closest parallel to welfare polls. Incentive-compatibility problems afflict these surveys: For example, the respondent who is asked whether she would purchase a new widget at some price might as well say "yes" even if her WTP for the widget is lower, since the widget's introduction into the market gives her a free option to buy it if her preferences

193 This Section has focused on the main source of concern about truthtelling discussed in the welfare polling literature—namely, that respondents will strategically lie. A different set of worries about the truthfulness of survey responses, grounded in psychology rather than economics, is that the respondent will have a "compliance bias": She will provide a socially acceptable answer, or the one that she thinks the interviewer wants to hear, rather than the true answer. *E.g.*, Colin Green & Sylvia Turnstall, *A Psychological Perspective*, *in* VALUING ENVIRONMENTAL PREFERENCES, *supra* note 16, at 207, 237–38. Here, as with strategic deception, the best response is to use survey design to mitigate the bias. *See id.* at 237–38. Calibration is second-best.

194 FISHKIN, THE VOICE OF THE PEOPLE, *supra* note 7, at 80.

195 Carson et al. define an "inconsequential" preference survey question as one where "the survey responses are not seen as having any influence on agency decisions or the agent is indifferent to all possible outcomes of the agency decision." Carson et al., Incentive and Informational Properties, *supra* note 175, at 3 (emphasis omitted). They argue that economic theory makes no prediction how respondents will answer inconsequential questions—in particular, economic theory does not predict truthful responses—and then go on to analyze when consequential surveys will be incentive-compatible. *See id.* at 3–5.

196 *See* Di Tella et al., *supra* note 189, at 811–12; Sugden, *supra* note 175, at 137–38.

change.[197] But firms continue to conduct these surveys at substantial cost; this would be surprising if the surveys had little informational value about consumers' preferences.[198] In fact, the informational value of consumer product research has been confirmed by studies showing a correlation between the degree of interest in a new product expressed in surveys and actual purchases of the product.[199] Correlational data of this sort is now regularly used to derive calibration factors for consumer surveys[200]—a close analogue to the use of calibration factors for welfare polls.

F. Question Formulation

Welfare polls will be informative to policymakers only if respondents answer the question posed in the survey, or (more precisely) answer the question that policymakers believe to be posed, or (more precisely yet) answer a question sufficiently close to that which policymakers believe to be posed. Consider an extreme case: effortful, sincere, and well-informed respondents whose answers to valuation question I, which they take the survey to pose, are uncorrelated with their answers to valuation question II, which the pollster reads the survey as posing.

> Survey questions can be misunderstood for various reasons:
>
> [What follows are] the major classes of interpretive difficulty that survey designers encounter. The question's grammatical structure (its *syntax*) may be ambiguous or too complicated for respondents to take in. Lengthy or complex questions can exceed respondents' capacity to process them, resulting in misinterpretations . . . . The question's meaning (or *semantics*) may elude respondents if they misunderstand vague, unfamiliar, or ambiguous terms or if they are misled by inapplicable presuppositions. Finally, the intended use of the question (its *pragmatics*) may create difficulties . . . .[201]

It is trivial to see that syntactic problems could affect welfare polls, like all other surveys, and easy to see that semantic difficulties could as well. The standard gamble and time tradeoff variants of the QALY technique ask respondents to use an esoteric method that they may

---

197 Carson et al., Incentive and Informational Properties, *supra* note 175, at 13–14.

198 *See, e.g.*, GILBERT A. CHURCHILL, JR. & DAWN IACOBUCCI, MARKETING RESEARCH: METHODOLOGICAL FOUNDATIONS 12–16, 212–30 (2005) (describing amount of marketing research, and discussing use of surveys as one of main mechanisms for collecting primary data, along with observation of behaviors).

199 *E.g.*, *id.* at 210; William J. Infosino, *Forecasting New Product Sales from Likelihood of Purchase Ratings*, 5 MARKETING SCI. 372, 375 (1986).

200 CHURCHILL & IACOBUCCI, *supra* note 198, at 209–11; Diamond & Hausman, *Is Some Number Better than No Number?*, *supra* note 106, at 54.

201 ROGER TOURANGEAU ET AL., THE PSYCHOLOGY OF SURVEY RESPONSE 25 (2000).

not grasp, while the rating scale asks respondents to locate health states on a 0–1 scale whose cardinal properties they may not understand. That latter problem affects happiness surveys too. And while respondents presumably *do* understand what dollars are, they may misconstrue the precise CV question posed—for example, the frequency of the payment (annual/monthly/lifetime WTP), maximum versus minimum, and so on.

As for the "pragmatics" of meaning, the problem here is that respondents may grasp the syntax and literal semantics of the survey question, but may interpret it nonliterally by virtue of communicative norms. For example, a general question about life-satisfaction that follows a specific question about life-satisfaction in some domain ("How happy are you with your marriage?") is naturally read to exclude that specific domain. Synonymous questions about happiness will tend to receive the same answers if separated in a survey; but if such questions are asked in succession, they are likely to receive different answers, since respondents—avoiding an interpretation that creates redundancy—will try to read them differently.[202]

The problem of misunderstood questions is, of course, a very general one for survey research, and a wide range of responses to the problem has been deployed. Traditionally, pollsters designed surveys using informal techniques such as "pretesting": giving the survey to a small group and developing an informal sense of the survey's problems.[203] Focus groups are a more elaborate way to do this. "Cognitive interviewing" is yet more elaborate.

> Ordinary interviews focus on producing codable responses to the questions. Cognitive interviews, by contrast, focus on providing a view of the processes elicited by the questions. Concurrent or retrospective *think-alouds* and/or probes are used to produce reports of the thoughts that respondents have either as they answer the survey questions or immediately after.[204]

Think of pretesting, focus groups, and cognitive interviews as second-order polling techniques—not techniques for eliciting valuations, but techniques for designing the first-order techniques. Second-order techniques also include second-order experiments: administering different trial surveys to different groups. There is now a large scholarly

---

202 Schwarz & Strack, *supra* note 31, at 64.

203 *See* Stanley Presser et al., *Introduction to* METHODS FOR TESTING AND EVALUATING SURVEY QUESTIONNAIRES 1, 2 (Stanley Presser et al. eds., 2004).

204 *Id.* at 4.

literature on second-order techniques,[205] and some of the more sophisticated approaches have percolated into welfare polls.[206]

Second-order techniques allow survey designers to identify and then reformulate misunderstood questions. More generally, they have diagnostic value with respect to most of the problems surveyed in this Part. Cognitive interviews, for example, can help reveal whether moral preferences are driving valuations; whether respondents lack relevant information; whether they are confused about probabilities, loss-averse, or otherwise irrational; and whether they are making a mental effort.

G. Representativeness

There is a straightforward answer to worries about whether the respondents to welfare polls are representative of the relevant population: Use random sampling techniques. These techniques, like those for ensuring question comprehension, are part of the general armamentarium of survey design.[207] Indeed, the origins of polling in the United States are bound up with random sampling: George Gallup became famous because his random sample of a few thousand people accurately predicted the outcome of the 1936 presidential election while the *Literary Digest*'s sample, a large but self-selected sample consisting of millions of postcards sent in by subscribers, did not.[208]

Random samples can be expensive. Nonrandom samples cost less—for example, the notorious convenience samples of college students used in many psychological surveys; or a "quota sample" of shoppers intercepted at a mall, which is sometimes the sampling format for CV studies.[209] But the expense of random sampling is presumably justified for a welfare poll that is meant to inform major governmental decisions, let alone a general schedule of valuations—at

---

[205] *E.g.*, PAUL P. BIEMER & LARS E. LYBERG, INTRODUCTION TO SURVEY QUALITY 258–304 (2003); COGNITION AND SURVEY RESEARCH (Monroe G. Sirken et al. eds., 1999); ROBERT M. GROVES ET AL., SURVEY METHODOLOGY 241–53 (2004); METHODS FOR TESTING AND EVALUATING SURVEY QUESTIONNAIRES, *supra* note 203; GORDON B. WILLIS, COGNITIVE INTERVIEWING: A TOOL FOR IMPROVING QUESTIONNAIRE DESIGN (2005).

[206] *See, e.g.*, BATEMAN ET AL., *supra* note 16, at 151–56; Champ, *supra* note 19, at 85–87.

[207] *See, e.g.*, BIEMER & LYBERG, *supra* note 205, at 305–50; GROVES ET AL., *supra* note 205, at 93–135.

[208] FISHKIN, THE VOICE OF THE PEOPLE, *supra* note 7, at 76–78.

[209] *See* BIEMER & LYBERG, *supra* note 205, at 309–12 (distinguishing between random sampling and different kinds of nonrandom sampling, such as convenience, purposive, and quota sampling).

least to the extent that it is important for the sample to be representative.

Why this last caveat? Variation in valuations may reflect (1) variation in objective circumstances, or (2) variation in preferences. In the first case, having a representative sample is crucial. Imagine, for example, a CV survey to value some policy that will clean up a park. The population of park users will vary in their objective circumstances—how they interact with the park. A sample skewed toward intensive users or users who are especially sensitive to aesthetics will tend to overstate the average WTP/WTA of the overall population of park users. By way of contrast, consider a QALY survey where all the participants are told about a particular hypothetical health state and asked to value that (not their own health). Here, there is variation in preferences but not objective circumstances, and it may be more important to have a high quality sample (respondents who are well informed, nondistorted, and so on) rather than a representative one.[210] In any event, to the extent that welfare polls ought to reflect the valuations of the U.S. citizenry as a whole, or some geographically or functionally defined subset thereof, random sampling techniques are available—and indeed regularly employed by CV, QALY, and happiness researchers alike.

H. Deliberative Welfare Polls as a Solution?

The previous sections surveyed specific obstacles to the enterprise of welfare polling and the matching techniques responsive to each obstacle. This section considers whether a change in the format of welfare polling from individual to group surveys might be generically useful in overcoming these obstacles.

Current practice is to administer welfare polls individually. Focus groups may be used to fine-tune the questionnaires, but the ultimate valuations and other data are derived from subjects responding solo, each separated from the other respondents. This is, of course, the general practice for policy surveys too, such as political opinion polls that ask for a stance about an issue before the government. The central thrust of the literature on "policy deliberation formats," as I term them—citizen juries, deliberative polls, citizen advisory boards, planning cells—is that group deliberation about policy questions can

210 *Cf.* Paul Dolan, *Aggregating Health State Valuations*, 2 J. HEALTH SERVICES RES. & POL'Y 160, 160 (1997) (noting that QALY surveys might be aggregated using median rather than mean values).

improve on solitary policy polling.[211] Might not the same be true for welfare polls?

Consider that the scholars who favor policy deliberation formats do so because they believe that structured group discussion can overcome informational, cognitive, motivational, and interpretive problems that afflict individual polls. If the solo-to-group shift has this benefit in the *policy* context, would it not also in the *welfare* context? A single presentation by an expert to the respondents assembled en masse is a cheap way to provide them with information. Debiasing techniques can also be thus cheaply presented; further, and more profoundly, group discussion itself is (or may be) a kind of debiasing. Lazy or deceptive types may find these postures harder to sustain in the face of group monitoring or collective enthusiasm for the valuation task. Misunderstandings about the meanings of questions can be sorted out in conversation.

In short, we should consider the possibility of *deliberative welfare polls*: survey techniques that involve collective discussions about the numerical value of some welfare impacts as measured using QALYs, CVs, a happiness scale, or some other scale, with ancillary informational, debiasing, and question-clarification techniques administered to the group; and that culminate in a collective verdict or in individual responses informed by the group deliberation.

In fact, this is not a new idea. CV scholars have toyed with this very idea, calling it "deliberative monetary valuation."

> DMV [deliberative monetary valuation] is the use of formal deliberation concerning an environmental impact in order to express value in monetary terms for policy purposes, and more specifically as an input to CBA. For example, consider a proposal to build a new road through a wilderness area . . . . A group of citizens would be selected and meet to discuss information about the environmental damages associated with the development. . . . The citizens would form a jury aiming to provide a monetary value for environmental damages which might be in terms of an individual willingness to accept compensation to allow the project to proceed.[212]

---

211 *See* sources cited *supra* note 7.

212 Simon Niemeyer & Clive L. Spash, *Environmental Valuation Analysis, Public Deliberation, and Their Pragmatic Syntheses: A Critical Appraisal*, 19 ENV'T & PLAN. C: GOV'T & POL'Y 567, 576–77 (2001). For other scholarship on deliberative money valuation, see sources cited *id.* at 576, and M. Sagoff, *Aggregation and Deliberation in Valuing Environmental Public Goods: A Look Beyond Contingent Pricing*, 24 ECOLOGICAL ECON. 213, 223–27 (1998).

A few of these group-deliberative CV studies have actually been conducted, although they remain very unusual.[213]

It has been objected, by authors sympathetic to policy deliberation formats, that "the DMV approach . . . is restricted to producing a monetary value. . . . The environment is still regarded as a commodity under DMV which crowds out civic virtues."[214] But this is really a generic objection to the CV format, equally applicable to both traditional solo CV surveys and group-based formats such as DMV. It does not show why welfare polls, if justifiably undertaken, are best conducted without inter-respondent deliberation.

A different objection, more to the point here, is that the valuations produced by deliberative welfare polls are, on balance, lower quality than traditional valuations—because groups will be too small to produce statistically representative results; because groups can work together to figure out strategic responses; because groups "go to extremes."[215] This objection may be apt, but it is hard to see why the objection would apply *differentially* to policy and welfare polling formats. Plausibly, citizen juries trump opinion polls if and only if deliberative welfare polls trump solo surveys. Perhaps that is strong. In any event, the literature on deliberative polling provides a rich set of group-based techniques that might be incorporated into the practice of welfare polling without abandoning its basic focus on well-being.

IV
WELFARE POLLS: A DEFENSE

The Article, up to this point, has been largely descriptive. Parts I and II described the existing trio of welfare polls and the various roles that they currently play, or might plausibly play, in administrative governance—specifically, in guiding policy analysis, informing government communications, and calibrating individual entitlements and

213 *See* Robin S. Gregory, *Valuing Environmental Policy Options: A Case Study Comparison of Multiattribute and Contingent Valuation Survey Methods*, 76 LAND ECON. 151 (2000); Lorna J. Philip & Douglas C. Macmillan, *Exploring Values, Context and Perceptions in Contingent Valuation Studies: The* CV Market Stall *Technique and Willingness to Pay for Wildlife Conservation*, 48 J. ENVTL. PLAN. & MGM'T 257, 259–60 (2005) (study using deliberative money valuation). Focus groups do seem to be fairly common in CV research, but these are group-based techniques for designing the questionnaire, not for collecting the data itself. *See* Michael D. Kaplowitz & John P. Hoehn, *Do Focus Groups and Individual Valuations Reveal the Same Information for Natural Resource Valuation?*, 36 ECOLOGICAL ECON. 237, 237 (2001); *see also* Alan Shiell et al., *Reliability of Health Utility Measures and a Test of Values Clarification*, 56 SOC. SCI. & MED. 1531, 1533–34 (2003) (employing deliberative, although not group-based, approach to eliciting QALYs).

214 Niemeyer & Spash, *supra* note 212, at 579.

215 *See id.* at 578–79; Cass R. Sunstein, *Group Judgments: Statistical Means, Deliberation, and Information Markets*, 80 N.Y.U. L. REV. 962, 1004–06 (2005).

obligations. Part III comprehensively reviewed the informational, cognitive, motivational, communicative, and strategic obstacles to using surveys in the elicitation of welfare valuations, and the possible solutions to these difficulties.

This Part is normative. First, it provides a moderate, nonutilitarian defense of the various uses of well-being information in administrative governance described in Part II—a defense grounded in the moral framework of "weak welfarism." Then (here drawing together the material presented in Part III), it responds to two objections to welfare polls: the "revealed-preference" tradition in economics, which is generally skeptical of surveys; and the deliberative-democratic tradition in political theory, which is skeptical of questions about preference, interest, or welfare, rather than the public good.

A. Weak Welfarism and the Need for Welfare Information

Let us distinguish, to begin, between the *moral* relevance of well-being to administrative choice, and its *legal* relevance. Eric Posner and I have elsewhere argued at length for the moral view we term "weak welfarism."[216] Weak welfarism says that overall well-being is one of the moral considerations that bears on governmental choice, but may not be the only such consideration. Formally, morality has the structure $\{W^*, F_1, \ldots, F_M\}$, where W^* is overall well-being, and $M = 0$. The F_i are possible moral considerations other than the maximization of aggregate well-being: for example, the protection of moral rights, the promotion of intrinsic environmental values, or the equitable distribution of well-being.

Weak welfarism, unlike utilitarianism or stronger variants of welfarism, eschews a monomaniacal focus on welfare. Utilitarianism insists that overall well-being is the sole morally relevant consideration. It has the structure $\{W^*\}$. The strong kind of welfarism popular among economists,[217] and defended by Louis Kaplow and Steven Shavell in a recent, high-profile book,[218] allows for distributive considerations but insists that only information about well-being is relevant to moral evaluation. According to Kaplow and Shavell, morality has

---

[216] ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 39–61; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39, at 204–16, 243–45; Adler, *supra* note 104, at 288–319.

[217] *See, e.g.*, Philippe Mongin & Claude d'Aspremont, *Utility Theory and Ethics*, *in* 1 HANDBOOK OF UTILITY THEORY 371, 394–95 (Salvador Barbera et al. eds., 1998); Andrew Moore & Roger Crisp, *Welfarism in Moral Theory*, 74 AUSTRALASIAN J. PHIL. 598 (1996).

[218] LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE (2002).

the structure $\{W_1, \ldots, W_N\}$, where each W_i is sensitive only to facts about welfare.

Weak welfarism, by contrast, allows that morality overall may well be sensitive to non-welfare facts—in the form of factors $F_1, \ldots, F_M$, which may well focus on aspects of individual lives or outcomes other than well-being. Moral rights and intrinsic environmental values would be the obvious candidates for such nonwelfarist moral factors.[219] But weak welfarism insists that well-being is an integral *part* of moral evaluation, in virtue of factor W^*.

To be sure, this *moral* discussion does not directly address the questions of *legal* obligation and authority that primarily concern government officials and legal scholars. Law and morality can come apart. Only a pure natural-law view—generally rejected by modern jurisprudents—would say otherwise. Weak welfarism is an account of the structure of morality, not a legal framework. Weak welfarism, if true, establishes that government officials are morally required to be sensitive to well-being. It does not establish that they are legally required or even legally permitted to do so.

In practice, however, administrative agencies *are* legally required or at least permitted to take account of well-being. First, agency organic statutes frequently use open-ended language that legally directs (or at least permits) agencies to pick the policy that maximizes overall welfare.[220] Second, statutes that do not take this open-ended, balancing form still might focus agencies on some aspect of welfare—for example, health and safety—and indeed frequently do.[221] Third, although statutes sometimes fit neither the first template nor the second—for example, statutes that take the form of rules rather than standards, directing agencies' attention to features of the world more readily ascertainable than welfare impacts[222]—*all* statutes have some degree of open texture, some area where agencies have legal discre-

[219] Moral rights prohibit certain kinds of infringements (for example, intentional physical harms), and are nonwelfarist in that the degree of prohibition is not calibrated to the welfare impact of the infringement. Intrinsic environmental values protect certain aspects of the environment (for example, the continued existence of a plant species) independent of the benefit of that aspect for humans or other entities (certain animals) that possess well-being.

[220] *See, e.g.*, Cass R. Sunstein, *Cost-Benefit Default Principles*, 99 MICH. L. REV. 1651, 1666–67 (2001). Sunstein discusses statutes that instruct agencies to eliminate "unreasonable" risks or balance costs and benefits. Given my view of cost-benefit analysis as a proxy for overall well-being, *see* sources cited *supra* note 58, I would argue that this sort of language requires or at least permits agencies to pick the policy that maximizes overall welfare.

[221] *See supra* text accompanying notes 55–57.

[222] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 64–65, 73–80.

tion.[223] It is both morally and legally appropriate for agencies to take account of well-being in resolving the discretionary choices that inevitably present themselves. Fourth, and a bit more concretely, the presidential cost-benefit orders, in place now for twenty-five years, have imposed a legal obligation on executive agencies—flowing from the President's legal powers to oversee executive agency decision-making—to consider overall welfare where that is statutorily permissible.[224] Fifth, although administrative officials' legal and moral obligations are distinct, Congress can always merge them. For example, Congress can amend particular organic statutes, converting them to the open-ended balancing form or the form that requires agencies to focus on some aspect of welfare. Or it can pass (and indeed has considered passing) a welfare supermandate that would give statutory teeth to the general legal obligation to consider overall welfare now embodied in the presidential cost-benefit orders.[225]

In short, questions about human well-being have substantial relevance, both moral and legal, to administrative governance. This basic observation synthesizes the different functions for welfare surveys discussed in Part II. To begin, where legally permitted, agencies should use policy-analytic techniques that help them ascertain which policy maximizes overall well-being or, alternatively, which one maximizes the particular aspect of well-being that is statutorily salient.[226] Cost-benefit analysis is the most obvious such technique and currently the one most widely employed by agencies. But there are others: Agencies might maximize QALYs. Or they might adjust longevity for happiness rather than health, and maximize happy life expectancy. Further, cost-benefit analysis is really a family of techniques, rather than a single rigid formula: Thus a variety of welfare polling formats, not just CV studies, can inform cost-benefit analysis—for example, through QALY-to-dollar conversions (now a regular practice at the FDA) or happiness-to-dollar conversions (currently a topic of scholarly work).

It *may* also be legally and morally appropriate for agencies to take into consideration the distribution of welfare. This suggestion raises large issues, which lie beyond the scope of this Article: What is the currency for distributive justice—welfare, resources, or something

---

223 *See* H.L.A. HART, THE CONCEPT OF LAW 124–36 (2d ed. 1994).

224 *See supra* text accompanying note 39.

225 *See* Fred Anderson et al., *Regulatory Improvement Legislation: Risk Assessment, Cost-Benefit Analysis, and Judicial Review*, 11 DUKE ENVTL. L. & POL'Y F. 89, 89–108 (2000) (discussing various proposed statutes requiring agencies to employ cost-benefit analysis).

226 *See supra* text accompanying notes 39–68.

else? And what is the optimal institutional structure for redistribution?[227] Still, it is at least plausible that (1) fair distribution means the fair distribution of well-being and (2) agencies in general, not just legislatures or the specialized agencies involved with the tax-and-transfer system, should concern themselves with the distribution of well-being. Just as welfare polls can inform administrative policy-analytic techniques that seek to maximize well-being or some of its aspects, so they can inform distributive analysis by agencies or other governmental bodies.[228]

Agencies do more than analyze and implement policy choices. They must often inform the public about the choices at hand, or about the current state of the world. Choices and outcomes can be characterized in various ways. Weak welfarism helps on this score, suggesting that government communications to the public should, morally, include welfare information; and such communications may also be legally required or permitted. Concretely, the policy impact statements such as those required by NEPA in the case of agency decisions that affect the environment, or the general "statement of basis and purpose" required by the Administrative Procedure Act whenever an agency proposes a legally binding rule, can and should describe welfare impacts. And the periodic statistics about the polity that government offices announce might include data about welfare, for example in the form of national well-being accounts.[229]

Nothing said to this point hinges on the specific content of administrative regulations. Those regulations might be opaque to welfare: In some contexts, an agency maximizes overall welfare by promulgating a rule for private actors that does not itself make reference to welfare—instead, for example, instructing those actors to use a specified technology, or to achieve a particular performance specified in natural rather than well-being units. But, sometimes, regulations are (partly or wholly) transparent to welfare. In other words, a private actor's obligations or entitlements may be defined (at least in part) in terms of the welfare effect of her actions. The clearest examples are regulations that impose a monetary exaction on some private activity, where the amount of the exaction (variously called a "tax," "fine," "fee," or "damages") depends on the welfare cost of the activity to third parties. A related example is suggested by environmental

---

227 *See, e.g.*, Richard J. Arneson, *Welfare Should Be the Currency of Justice*, 30 CAN. J. PHIL. 497 (2000); Chris William Sanchirico, *Deconstructing the New Efficiency Rationale*, 86 CORNELL L. REV. 1003 (2001).

228 *Cf. supra* text accompanying notes 62–64 (discussing possible use of welfare polls for tax policy).

229 *See supra* text accompanying notes 90–103.

trading markets. Private actors might be allocated tradeable credits for beneficial activities (for example, refraining from pollution), with the amount of the credit dependent on the welfare impact of the activity. In both of these contexts, CV, QALY, or happiness surveys can be useful in ascertaining what the welfare effects of private activities are.[230]

To be sure, welfare polls are not costless undertakings. The elements of a well-designed survey can be expensive: Academics or other individuals with expertise in the relevant good, in valuation, and in survey design must spend time in drafting the poll; focus groups or other second-order techniques must be undertaken to test it; interviewers may need to be hired to conduct the survey; respondents may need to be compensated, and in any event the use of their time is an economic cost; and econometricians will need to glean valuations from completed surveys. But in some cases—for example, the design of a major policy with large anticipated impacts—the expected welfare benefits of conducting a specific survey will outweigh the expected costs. And, more generally, the costs of surveys can be spread over multiple decisions by applying a single survey to a multiplicity of similar policies, or by using a few very high-quality surveys to estimate either a general schedule of valuations for different types of impacts or a general "benefits function."[231]

B. The Revealed-Preference Objection to Welfare Polling

The argument thus far, synthesizing the material in Part II, shows that government officials are (morally and legally) justified in securing information about well-being, but it does not yet establish that the government is justified in securing that information through welfare polls. This brings us to the various obstacles to welfare polling surveyed in Part III. Why welfare polls rather than other techniques—in particular, deriving valuations from behavioral evidence?

A population's valuations for some good might be inferred in two generic ways: through "stated-preference" techniques, which ask members of the population to say what their valuations are; or through "revealed-preference" techniques, which infer valuations from the population's nonverbal behaviors, such as their transactional activities or their locational, occupational, or recreational choices. The most important revealed-preference technique in economics is, of course, to use observed demand and supply curves in markets for private goods to infer consumer and supplier valuations for these mar-

230 *See supra* text accompanying notes 69–89.

231 *See supra* text accompanying notes 44, 82.

keted goods.[232] This technique is so pervasive that scholarship about the choice between revealed-preference and stated-preference approaches to valuation often does not include it in the first category. But it sits squarely there: Inferring *P*'s value for *X* from the amount that *P* actually pays or receives for *X* is the quintessential measurement tool of applied welfare economists.

The other commonly employed revealed-preference techniques seek to infer valuations of public goods, or of private goods that are not separately marketed, from observed activities. The leading examples, here, are *travel cost* techniques (which infer the recreational and other use values of sites such as wilderness areas from the costs in time and travel expense that individuals are willing to incur to visit the sites); *property value* approaches (which use the correlation between housing prices at different sites and environmental quality at those sites to infer valuations of environmental quality); *hedonic wage* techniques (which infer worker valuations of job characteristics, typically fatality risks, from the correlation between those characteristics and wages); and *defensive behavior* techniques (which infer valuations of health states from behavior to avoid or mitigate illness, such as seeking medical care, purchasing safety devices such as bicycle helmets or smoke detectors, or using bottled water to avoid contaminated water supplies).[233]

Many economists, particularly outside the areas of environmental and health economics, reflexively prefer these sorts of revealed-preference techniques to stated-preference techniques. Unlike mainstream psychologists, who are perfectly comfortable with probing mental states by asking patients to talk, much of the economics profession remains suspicious of surveys, at least surveys about subjective states such as preferences.

> Many surveys contain a wealth of subjective questions that are at first glance rather exciting. Examples include: . . . "How satisfied are you with yourself?"; or "How satisfied are you with your work?" Yet despite easy availability, this is one data source that economists rarely use. In fact, the unwillingness to rely on such

---

232 On these techniques, see RICHARD E. JUST ET AL., THE WELFARE ECONOMICS OF PUBLIC POLICY (2004).

233 *See* FREEMAN, *supra* note 16, at 95–136, 353–452; Nancy E. Bockstael, *Travel Cost Models*, *in* THE HANDBOOK OF ENVIRONMENTAL ECONOMICS, *supra* note 191, at 655; Mark Dickie, *Defensive Behavior and Damage Cost Methods*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 395; A. Myrick Freeman III, *Hedonic Pricing Methods*, *in* THE HANDBOOK OF ENVIRONMENTAL ECONOMICS, *supra* note 191, at 672; George R. Parsons, *The Travel Cost Method*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 269; Laura O. Taylor, *The Hedonic Method*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 331.

> questions marks an important divide between economists and other social scientists.[234]

This generic suspicion of subjective surveys has fueled much of the opposition to the CV method.[235] And it surfaces in the current OMB guidance concerning cost-benefit analysis, which allows agencies to use surveys but places them lower in the hierarchy of sources than behavioral data.

> Other things equal, you should prefer revealed preference data over stated preference data because revealed preference data are based on actual decisions, where market participants enjoy or suffer the consequences of their decisions. This is not generally the case for respondents in stated preference surveys, where respondents . . . may be inclined to bias their responses for one reason or another.[236]

The revealed-preference objection to welfare polling might be framed in two forms: strong (noncomparative) or weak (comparative). The strong, noncomparative objection is that welfare surveys provide essentially no information about a population's valuations. They are no more informative than responses to gibberish questions. To put the noncomparative objection formally: The rational official's beliefs about the population's valuations are no different after the survey than before.

Although the noncomparative objection to welfare polls is a bit of a straw man, some of the critical literature comes close to making this strong claim.[237] And the claim seems less extreme once it is recognized that valuations are a matter of idealized preferences, a point I have already stressed. *P*'s statements about what he currently wants might be little better than noise as evidence of his hypothetical, fully informed, fully rational preferences.

Still, the strong objection to welfare polls seems overstated. A crucial point is that the consumers of the polls—government officials—are themselves imperfectly informed. An already omniscient social planner's estimates of a given population's valuations would not be altered by welfare surveys, but the head of the EPA's office of policy analysis is not God. Imagine a policy analyst who wants to estimate a population's valuations of some good and has not yet examined any specific valuation studies, either revealed-preference or

---

234 Marianne Bertrand & Sendhil Mullainathan, *Do People Mean What They Say? Implications for Subjective Survey Data*, 91 AM. ECON. REV. 67, 67 (2001).

235 *See* Carson et al., *supra* note 16, at 176.

236 OFFICE OF MGMT. & BUDGET, EXECUTIVE OFFICE OF THE PRESIDENT, OMB CIRCULAR NO. A-4, REGULATORY ANALYSIS 24 (2003), *available at* http://www.whitehouse.gov/omb/circulars/a004/a-4.pdf [hereinafter OMB CIRCULAR A-4].

237 *E.g.*, Diamond & Hausman, *Is Some Number Better than No Number?*, *supra* note 106, at 46.

stated-preference studies. Her current estimates are shaped by introspection (a sense of what *she* would prefer); by the unsystematic observation of others' value-revealing behaviors and utterances that will occur over the course of any normal human life; and perhaps by the general theoretical literatures, in economics, philosophy, and psychology, on rationality, preference-formation, and well-being. Relative to *this* kind of "prior" information about a population's valuations, welfare polls surely provide substantial new information.[238]

But what of the comparative objection that welfare polls are dominated by revealed-preference studies? Think of the objection this way: For the policy analyst whose estimates of a population's valuations are shaped both by the general background data just described (introspection/experience/general theory) *plus* specific revealed-preference studies, welfare polls do not typically furnish substantial new data.

One traditional answer to this sort of question is that welfare polls can elicit valuations that will not be reflected in behavior. Specifically, proponents of CV studies often argue that these can be employed to measure environmental nonuse values, which may be difficult or even impossible to estimate with revealed-preference methods.

> CV surveys measure the total value of the described good while revealed preference techniques, which are based on observed behavior in private markets related to the environmental good, measure only direct use value. Revealed preference techniques are usually only capable of capturing . . . the direct use portion of total value, because they rely on the availability of an implicit private market for a characteristic of the good in question. . . . In contrast, passive use value can be seen as simply a special case of a *pure* public good.[239]

The problem with this traditional defense of welfare polls is twofold. First, it does not apply to welfare polls that inquire about the multifold aspects of welfare that *are* evidenced by behavior, for example CV studies that focus on health, recreation, visibility, and

---

238 My discussion, here and below, is Bayesian in spirit. The policy analyst has subjective beliefs about valuations, which are updated through surveys. For a fuller description of Bayesian approaches to the estimation of valuations, see Adler, *supra* note 25, at 54–56.

239 Carson et al., *supra* note 16, at 176. For similar defenses of CV surveys, see, for example, Report of the NOAA Panel on Contingent Valuation, 58 Fed. Reg. 4602, 4602–03 (Jan. 15, 1993); BATEMAN ET AL., *supra* note 16, at 21–22; FREEMAN, *supra* note 16, at 154; and Nicholas E. Flores, *Conceptual Framework for Nonmarket Valuation*, *in* A PRIMER ON NONMARKET VALUATION, *supra* note 16, at 27, 47–50. *See also* Boyle & Bergstrom, *supra* note 113, at 191 (noting "conventional wisdom that CV is the only game in town when it comes to measuring nonuse values").

other use values, or QALY and happiness studies generally. Second, as argued earlier, nonuse values often arise from moral or disinterested preferences. Shocking as this may sound to the environmental economists who have done much to develop the CV technique, a well-designed study to estimate welfare valuations should try to screen out disinterested preferences, and ignoring nonuse values is a practicable way to do that.

What we need, in short, is an argument for why welfare polls have substantial evidentiary value, as compared to revealed-preference techniques, in estimating *use* values—preferences for features of outcomes that derive from the subject's own self-interest, quintessentially preferences that entail a physical interaction (a "use"). The existing literature, focused on the comparative advantages of welfare polls for estimating nonuse values, does not furnish this argument.

I believe such an argument can be furnished, as follows.[240] One set of problems with using revealed-preference techniques to estimate use values involves the idealizing conditions for valuation discussed in Part III: The actors may be disinterested, uninformed, irrational, or apathetic. A distinct set of problems involves the measurement of the actors' valuations: That is to say, even if the observer is satisfied that the actor's motivational, cognitive, and informational state is sufficiently idealized, it may be difficult for an observer constrained not to *talk* to the actor to estimate the actor's valuations from her behavior. The ultimate objective of any valuation study is to estimate the strength of some individual's subjective preference for some good, on some numerical scale (QALY, dollar, happiness, or other). Welfare polls inquire directly about the strength of preference and intervene to shape the subject's perception of the good (by providing information, by characterizing the good as a package of attributes, and so on). Revealed-preference techniques eschew these questions and interventions, and therefore run up against distinctive measurement obstacles that welfare polls can avoid—at least if subjects are truthful and at least if they absorb the information provided by the pollster, points to be returned to in a moment.

1. Measurement Problems

I will briefly survey these two sets of problems for revealed-preference studies, beginning with the measurement problems. One large measurement problem involves the gap between preferences (an

240 *See also* CHURCHILL & IACOBUCCI, *supra* note 198, at 212–30 (discussing pros and cons of surveys versus behavioral data in marketing research, with particular focus on versatility as main informational advantage of surveys).

unobservable mental state) and action. Consider the very simplest example: estimating average WTP for a marketed good among a population currently consuming the good in a competitive market. All the consumers are observed paying the same price P for the good; but of course it is *not* the case that each consumer is willing to pay P for each unit of the good she purchases, or that P is each consumer's average WTP for the goods she purchases. If we assume consumers are rational, we might infer that each consumer values each unit of the good at P or greater, and values her least valued unit at P; but we still have neither observed, nor inferred, average WTP. In other words, P reveals the consumers' marginal valuations (if they are rational), but not their inframarginal valuations.

A slightly more complicated example, again from the heartland of revealed-preference work: market behavior. Imagine that the price of a marketed good X decreases from P to P'; consumer incomes, and the prices of all other goods, remain the same. What is aggregate consumer WTP for this change in price? It is tempting to say that, at least if we can observe the demand curve for X, the aggregate WTP is simply the change in area under the demand curve—the change in ordinary consumer surplus. It turns out (for fairly fundamental reasons in demand theory) that this is not true. Rather, aggregate WTP is the change in area under the "Hicksian" or "income compensated" demand curve, an unobservable entity that separates the substitution and income effects of a price change. A very substantial literature in applied economics discusses the conditions under which the change in consumer surplus is a good approximation for aggregate WTP for a price change and, if not, what other techniques (usually involving assumptions about the shape of utility functions) can be used.[241]

The problem of inferring the subject's beliefs about the good being valued is a different, and equally pervasive, measurement problem for revealed-preference techniques. The ultimate objective of a revealed-preference study or welfare poll is to determine how some good, with characteristics $C_1^*, \ldots, C_n^*$, or a change in some good from characteristics $C_1^*, \ldots, C_n^*$ to $C_1', \ldots, C_n'$, is valued by some group of subjects. But observing how a subject behaves in the vicinity of a good that actually has those characteristics, or in response to an actual change of that sort, does not directly evidence the valuation, because the subject might misperceive the good or change. If we eschew discursive measures (asking the subject how he sees the good),

241 A good review of this literature, and the basic theoretical flaws of ordinary consumer surplus, is FREEMAN, *supra* note 16, at 49–72.

then we must assume that the perception is accurate (an heroic and unwarranted assumption) or somehow (again without discursive evidence) posit a function correlating perceived and actual characteristics.

This problem recurs in the literature on revealed-preference techniques. Consider property value models. To quote a leading textbook on applied economics, a crucial question to be asked in using these methods is the following: "Is there sufficiently close correspondence between [individuals'] perceptions of amenity levels (which presumably govern the choices reflected in property prices) and the objective measures of amenity levels that are available to the researcher?"[242] The same question arises with respect to travel cost models.

> Objective measures of [site] quality are reproducible . . . . However, people might make choices about recreation on the basis of their perceptions of quality rather than the objective measure. If individuals' perceptions are functions of objective measures and personal characteristics, then it may be possible to estimate a "perception function" and to use this function to model choices and measure welfare values.[243]

Or consider the standard use of hedonic wage models or defensive behavior models to estimate WTP to avoid risk. A critical problem here is that the objective risk associated with a job or a defensive behavior may differ from the risk perceived by the worker or actor. "Valuation methods [for risk] based on revealed-preference have the virtue of relying on actual behavior but can be applied only when the analyst knows (or can reasonably infer) what decision alternatives and consequences (including their pecuniary, health, and other attributes) were perceived by the decision maker."[244]

A third measurement problem for revealed-preference techniques is that of determining valuations for counterfactual goods. Imagine that a policy will change the attributes of some marketed or nonmarketed good from $C_1, \ldots, C_n$ to $C_1', C_2, \ldots, C_n$. If existing goods of this sort never attain C_1', or if C_1' is attained but typically bundled with changes in the good's other characteristics, using revealed-preference techniques to measure the value of the hypothesized good may be very difficult.

> [Revealed-preference] methods . . . suffer on the grounds that the new situation (after the environmental quality change) may be outside the current set of experiences (or outside the data range) . . . . [They also] may suffer from colinearity among attrib-

242 *Id.* at 363.
243 *Id.* at 428.
244 Hammitt & Graham, *supra* note 140, at 34.

> utes. Colinearity precludes the isolation of factors affecting choice. . . . For example, water quality attributes (BOD, turbidity, etc) may be correlated but the economic valuation may only be interested in valuing an improvement in one of the attributes.[245]

A similar point holds about consumer preferences for currently nonexistent products that a firm might introduce, and indeed is a major reason for the use of surveys in market research. "[I]f Doritos were to create a new spicy salsa-flavor chip as a line extension, by definition no purchase data would exist because the snack food would not have been available yet for purchase."[246]

A final measurement problem involves the cost of goods. Revealed-preference methods use information about the cost of a good to some subject as a step in estimating the subject's valuations. If the good is marketed and the consumer pays for it out of pocket, the cost is pretty straightforward: its price. But if the good is not marketed, estimating cost may be more complicated. This is a key problem for travel cost studies, where cost equals the sum of the direct and opportunity costs of traveling to the site. The solution has been to *survey* site users about crucial travel details (for example, where the trip originated and what the user's income is). And if the good is marketed but the consumer is insured, the price will exceed her cost. In particular, the wide existence of health insurance makes inferring WTP for health from the prices of health care services or products very tricky.

2. *Idealization Problems*

Let us turn, now, to the other set of reasons why welfare polls have informational value even with revealed-preference studies in the picture—namely, the idealizing conditions for valuation. Those conditions, again, are that the subjects be self-interested; be well-informed; have undistorted preferences; and exert sufficient mental effort. Real world actors can, of course, fall short on all of these dimensions. And while revealed-preference techniques offer some opportunities, by a judicious selection of goods or subjects, to screen out problematic preferences, they eschew the full range of interventions that welfare polls can utilize.

---

[245] W. Adamowicz et al., *Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities*, 26 J. ENVTL. ECON. & MGMT. 271, 272 (1994). The problem of colinearity is a major one for defensive-behavior methods, often referred to in that context as the problem of "joint production." *See* Cropper, *supra* note 17, at 335; Dickie, *supra* note 233, at 412–13; F. Reed Johnson et al., *Valuing Morbidity: An Integration of the Willingness-to-Pay and Health-Status Index Literatures*, 16 J. HEALTH ECON. 641, 644 (1997).

[246] CHURCHILL & IACOBUCCI, *supra* note 198, at 209.

Consider, to begin, the problem of disinterested preferences. Revealed-preference techniques can partly address this problem by ignoring goods for which preferences are mainly moral or otherwise disinterested. The classic example is nonuse values for environmental goods. But preferences for experienced goods can also be partly disinterested. A good example is risk: Hedonic wage studies that infer WTP from wage premiums for riskier jobs are problematic, in part, because workers may be incurring risk out of a sense of responsibility to family.[247] The same goes for individuals who purchase safety devices (particularly if those devices directly benefit family members and not just the individual, as in the case of safer cars or appliances).[248]

Admittedly, as discussed in Part III, welfare pollsters do not seem to have experimented much with discursive devices for screening out disinterested preferences.[249] Currently, therefore, with respect to the problem of self-interest, the choice between revealed- and stated-preference techniques is pretty much a wash. If successful discursive screening devices are developed, these would be one comparative advantage (among the others herein discussed) of welfare polls.

Turn, now, to the problem of information. Clearly, the actors in revealed-preference studies may lack an idealized stock of information[250]—a problem already touched upon in my discussion of the gap between the true and perceived characteristics of goods.[251] Focusing the study on well-informed actors is a possibility, but the relevant facts may not be widely known, or known only by a nonrepresentative segment of the population, creating potential problems of sample size or statistical bias. Here, current welfare polling techniques, as described in Part III, do offer important advantages. Those include telling respondents more about the good at stake; deleting welfare-irrelevant information; and characterizing the good as a bundle of locations along different dimensions of welfare, as in conjoint analysis. A more experimental possibility is switching from solo to group surveys, so as to disseminate information cheaply.[252]

A parallel analysis applies to the problem of preference distortion. To begin, it is clear that preferences measured in behavioral

---

247 ELIZABETH ANDERSON, VALUE IN ETHICS AND ECONOMICS 197–99 (1993).

248 *Cf.* James R. Bettman et al., *Constructive Consumer Choice Processes*, 25 J. CONSUMER RES. 187, 197–99 (1998) (suggesting that consumers often must justify their choices to others and use choice procedures that make justification easier).

249 *See supra* text accompanying notes 113–14.

250 *See* OMB CIRCULAR A-4, *supra* note 236, at 20; Bettman et al., *supra* note 248, at 201–02.

251 *See supra* text accompanying notes 241–44.

252 *See supra* text accompanying notes 126–32, 211–15.

studies are distorted. Consider the major distortions that affect welfare polls, discussed in Part III: deviations from the expected utility model for processing probabilities, loss aversion, and resistance to trading off different dimensions of well-being. All of these are pervasive features of human decisionmaking[253] and are hardly confined to welfare polls. For example, the literature on consumer behavior shows that each of the distortions characterizes individuals purchasing marketed goods. Consumers often make probabilistic decisions in accordance with prospect theory, a model of choice that seems descriptively much more accurate than expected utility theory.[254] "[D]ecision makers choose more optional features in a consumer choice (e.g., air conditioning in an automobile) when they are given a fully loaded model and asked to delete options they do not want than when they are given a base model and asked to add options at additional cost" (loss aversion).[255] James Bettman, Mary Luce, and John Payne have extensively documented tradeoff aversion in consumer choice. If a few simple axioms are satisfied, rationality would require that consumers resolve purchase decisions involving products with attributes along multiple dimensions by employing a "weighted additive" method: weighting each dimension, scoring the products on the different dimensions, determining an overall score for each product, and choosing the product with the highest score.[256] But consumers regularly fail to engage in weighted additive decisionmaking and instead use methods such as satisficing or lexicographic choice—either to economize on mental effort or to avoid thinking about tradeoff rates for dimensions (such as health or life) that are particularly important.[257]

Some of the other distortions that have raised concerns about welfare polls—such as framing, anchoring, and availability effects—

---

253 *See, e.g.*, HASTIE & DAWES, *supra* note 142, at 289–312; PLOUS, *supra* note 139, at 84–188; Baron, *supra* note 110, at 82–84.

254 *See* Rong Chen & Jianmin Jia, *Consumer Choices Under Small Probabilities: Overweighting or Underweighting?*, 16 MARKETING LETTERS 5, 5 (2005) (citing sources).

255 MARY FRANCES LUCE ET AL., EMOTIONAL DECISIONS: TRADEOFF DIFFICULTY AND COPING IN CONSUMER CHOICE 43 (2001); *see also* Nathan Novemsky & Daniel Kahneman, *The Boundaries of Loss Aversion*, 42 J. MARKETING RES. 119, 119–20, 125–26 (2005) (describing research showing loss aversion in consumer choice and suggesting ways that marketers may mitigate it).

256 *See* Bettman et al., *supra* note 248, at 190. For a discussion of the "independence" axioms that, together with the basic axioms of rational choice theory, entail weighted-additive decisionmaking, see DETLOF VON WINTERFELDT & WARD EDWARDS, DECISION ANALYSIS AND BEHAVIORAL RESEARCH 331–41 (1986).

257 *See* Bettman et al., *supra* note 248, at 189–99. The analysis also includes ease of justification as a choice goal. On tradeoff avoidance in particular, see *id.* at 205–06 and generally LUCE ET AL., *supra* note 255.

have been observed in consumer behavior as well.[258] Further, all of the distortions that characterize consumers also presumably characterize visitors to parks, purchasers of properties with environmental amenities or disamenities, workers considering risky jobs, and individuals mitigating or averting disease states—the individuals whose behaviors are studied by travel-cost, property-value, hedonic-wage, and defensive-behavior techniques, respectively.

As with information, revealed-preference methods can try to circumvent distortions by focusing on the behavior of unusually rational types—but that creates potential sample size and bias issues. By contrast, welfare polling currently offers a range of techniques for reducing distortion, discussed in Part III. To recap the discussion there: Direct debiasing techniques have had some mixed success. A more successful response to preference distortions may be to change the elicitation method within a given welfare polling format, or indeed to change the format—for example, to switch from CV to QALY surveys to overcome tradeoff biases in valuing health states or from standard happiness studies to Kahneman's moment-based surveys to overcome availability biases. Switching from solo to group surveys may help to facilitate debiasing, just as it may help to disseminate information cheaply.[259]

Finally, real-world actors do economize on mental effort. In particular, as just mentioned, an important focus of research into consumer choice has been to document how the fact of bounded rationality—the fact that humans cannot instantly and effortlessly retrieve items from memory, process new information, and perform computations—produces heuristics.[260] Admittedly, respondents to welfare polls have an additional, strategic incentive to eschew effort that consumers and other actors considering choices with significant personal effects lack: Consumers will internalize the benefits of their incremental mental efforts, while survey respondents will not. On the other hand, as discussed in Part III, the polling format can presumably intensify the internal and social pressures that induce additional effort—by remunerating respondents for their answers, by using face-to-face interviews rather than mail or Internet surveys, and (once more) by shifting from solo to group formats.[261]

---

[258] *See* Bettman et al., *supra* note 248, at 201–02 (availability); *id.* at 208 (framing); Joseph C. Nunes & Peter Boatwright, *Incidental Prices and Their Effect on Willingness to Pay*, 41 J. MARKETING RES. 457 (2004) (anchoring).

[259] *See supra* text accompanying notes 153–64, 211–15.

[260] *See* JOHN W. PAYNE ET AL., THE ADAPTIVE DECISION MAKER 9–15 (1993); Bettman et al., *supra* note 248, at 189–99.

[261] *See supra* text accompanying notes 173–74, 211–15.

* * * *

In sum, the comparative objection to welfare polling fails. Welfare polls can provide substantial new information to the (nonomniscient!) policy analyst, given the limitations of revealed-preference methods with respect to measurement and the idealizing conditions for valuation (information, nondistortion, self-interest, and mental effort).[262]

My response to the revealed-preference objection assumes, to be sure, that the communicative difficulties that may affect welfare polls—respondent misunderstanding of the question asked, on the one hand, and a lack of truthfulness in answering the question, on the other—do not negate their informational value. *If* the respondents to these polls assigned semantic content to the questions asked that did not correlate with the semantic content intended by the pollsters, or if the respondents understood the questions but provided untruthful answers that did not correlate with their real preferences, the revealed-preference objection to welfare polling (indeed, in its strong, noncomparative form) would be persuasive.

To begin, it should be underscored that the possibility of respondent misunderstanding and strategic deception, along with that of strategic laziness, is hardly unique to welfare polls, but generalizes to all sorts of surveys—political polls, psychological surveys, consumer research, and so on. Presumably the various "consumers" of these surveys, often quite savvy (e.g., politicians, firms), would not use them if deception, shirking, and misunderstanding by respondents seriously undercut their informational value.

Why does this not occur? Shirking has already been discussed. Avoiding misleading questions has been intensively studied by polling scholars, and a large set of second-order techniques (pretesting, focus groups, cognitive interviews) are available.[263]

---

[262] Two other categories of generic problems that affect revealed-preference techniques should also be noted. The first category includes problems of bias and self-selection: For example, workers with a greater appetite for risk tend to take risky jobs, which means that the observed wage premium for a given risk will tend to be lower than population WTA for that risk. *See, e.g.*, OFFICE OF RESEARCH AND DEV., U.S. ENVTL. PROT. AGENCY, HUMAN HEALTH METRICS FOR ENVIRONMENTAL DECISION SUPPORT TOOLS: LESSONS FROM HEALTH ECONOMICS AND DECISION ANALYSIS 18–19 (2001), *available at* http://www.epa.gov/nrmrl/pubs/600r01104/600R01104.pdf; OMB CIRCULAR A-4, *supra* note 236, at 20 ("[In revealed-preference studies,] the specific market participants being studied should be representative of the target populations to be affected by the rulemaking under consideration . . . ."). The second category involves assumptions about market structure—for example, the assumption in property value models that relocation costs are low. *See supra* text accompanying note 53; OMB CIRCULAR A-4, *supra* note 236, at 20 (stating that, in revealed-preference studies, market should be competitive).

[263] *See supra* Part III.F.

As for strategic deception, the discussion in Part III suggested that the attempt to design incentive-compatible formats is probably a dead end. A better answer to worries about deception points to bounded rationality and norms. Boundedly rational individuals may find it too difficult to maintain a consistent pattern of lies (particularly in the face of questions designed to test for the consistency of preferences), and anyone may feel internal or social pressure to tell the truth. Surveys can be designed to intensify this pressure. Further, we have seen that stated valuations correlate with behaviorally inferred valuations, which underscores the informational value of surveys. If respondents were not constrained (by norms, internal pressure, incentive-compatible design, or whatever) to state their actual preferences—or at least valuations that (true or not) are systematically related to their actual preferences—then this observed correlation would be very puzzling. Finally, and reciprocally, stated valuations need not be accurate to be informative. Inaccurate but nonrandom statements also have evidentiary value. Ideally, survey design would induce truthtelling, but a cruder, second-best solution is to use calibration factors to adjust for various factors that drive a wedge between stated and actual valuations.[264]

A final thought: Might it not be possible to combine the advantages of surveys and revealed-preference techniques by placing actors in a favorable informational, cognitive, and motivational state, and then inferring their valuations from their behavior (rather than their statements)? Indeed, there is a small but growing body of scholarship, in applied economics, that undertakes valuation "experiments."[265] In these experiments, subjects in a controlled setup are given the opportunity to buy and sell goods. Thus far, valuation experiments have been mainly used to learn about individual behavior or to test the valuations emerging from CV studies. But one can imagine experiments themselves becoming a primary source of valuation data.

An analysis of the potential value of controlled experiments (including one kind of quasi-experimental setup that has recently become popular, the "information market")[266] in producing information about welfare is beyond the scope of this Article. It seems clear, however, that controlled experiments would be applicable to a narrower range of welfare impacts than welfare polls. This is because an

---

264 *See supra* Part III.E.

265 *See* Murphy & Stevens, *supra* note 113; Laura O. Taylor, *Experimental Methods for the Testing and Design of Contingent Valuation*, *in* HANDBOOK ON CONTINGENT VALUATION, *supra* note 97, at 177.

266 *See* Michael Abramowicz, *Information Markets, Administrative Decisionmaking, and Predictive Cost-Benefit Analysis*, 71 U. CHI. L. REV. 933 (2004).

experimental setup is one in which the welfare effect being valued is actually allowed to occur. Where the effect involves a substantial reduction in either an individual's well-being or in overall well-being, the running of the experiment may itself be precluded by weak welfarism: Consider the use of experiments to value serious disease, physical injury, or the environmental degradation of residential or recreational areas. Whatever the potential informational advantages of experiments vis-à-vis both traditional revealed-preference techniques and welfare polls, there are presumably moral limits—cognizable within the framework of weak welfarism—to the kinds of welfare effects that they should justifiably be used to value. Thus, the possibility of experiments as a source of valuation data certainly bears additional research but does not undercut the case for welfare polls presented here.

C. *The Deliberative Democracy Objection*

Let us turn, then, to the second general objection to welfare polling: Call it the "deliberative democracy" or "civic republican" objection. This objection, like the revealed-preference objection, has emerged most clearly in the critical literature about CV surveys. One set of critics has been traditional economists who agree that policymakers need information about preferences, but are suspicious about surveys. A different set of critics has been environmentalists or political theorists who have no aversion to surveys, polls, and discussions, but deny that environmental or other policy issues should be resolved through the monetary measures of preferences that CV studies yield. They argue that policy should be sensitive to *citizen deliberation*: processes where citizens adopt a public-regarding (rather than self-interested) perspective and reach judgments about what policy best serves the public good. And these critics claim that CV surveys, which ask about exogenous preferences rather than judgments endogenous to the process of reasoning about policy—and which use WTP questions that require or at least invite a self-interested perspective on the part of respondents—have a structure inconsistent with citizen deliberation.

The philosopher Mark Sagoff is probably the leading civic republican critic of CV studies (and of cost-benefit analysis more generally). In a number of books and articles, Sagoff has argued along the following lines:

> When individuals participate in the political process to determine the common values and purposes that hold them together as a community or as a nation, they regard themselves as judges of public policy, not merely as channels or locations at which wants can be

> found. Debates in which individuals or their representatives discuss and decide upon public values need have no analogy, then, with markets where individuals determine and pursue personal preferences. In a democracy the application of a cost-benefit formula cannot replace the public discussion of ideas; it is not just what the person *wants* but what he or she *thinks* that counts.[267]

Similar criticisms of the CV technique are offered by other critics of cost-benefit analysis, most recently Lisa Heinzerling and Frank Ackerman in a high-profile book, *Priceless*:

> Asking people in a shopping mall about hypothetical scenarios involving bronchitis, or talking to people who answer the phone about how much they would pay to protect the bald eagle, amounts to elevating the consumer over the citizen. It also turns the very idea of republican government on its head, suggesting that elected representatives should no longer try, through deliberation, reasoning, and debate, to shape the mass of public opinion into a sensible and lasting set of ideas, but should instead take their marching orders from a small sample of nameless individuals who answer a survey.[268]

These quotations exemplify a standard line of critical scholarship about CV studies,[269] but it should be noted that the criticisms would also seemingly apply to welfare polls that seek to elicit nonmonetary measures of respondents' self-interested preferences, such as QALY or happiness surveys. Indeed, Heinzerling and Ackerman explicitly criticize QALY measures.[270]

The civic republican view of appropriate policymaking that infuses Sagoff's and similar criticisms of welfare polls also motivates the proposals for citizen juries, citizen advisory committees, and deliberative polling—what I have termed policy deliberation formats. For example, James Fishkin, the leading proponent of deliberative polling, sees deliberation as one of the central desiderata for political choice that this survey approach instantiates.[271] By deliberation, Fishkin

---

[267] MARK SAGOFF, THE ECONOMY OF THE EARTH 100 (1988). This book draws on a number of Sagoff's prior articles. *Id.* at x. For more recent statements drawing a similar distinction between consumer preferences and citizen preferences or judgments, see Sagoff, *supra* note 212, at 213–14, and MARK SAGOFF, PRICE, PRINCIPLE, AND THE ENVIRONMENT 2–3 (2004).

[268] FRANK ACKERMAN & LISA HEINZERLING, PRICELESS: ON KNOWING THE PRICE OF EVERYTHING AND THE VALUE OF NOTHING 213 (2004).

[269] *See, e.g.*, ANDERSON, *supra* note 247, at 203–10; Steven Kelman, *Cost-Benefit Analysis: An Ethical Critique*, REGULATION, Jan.–Feb. 1981, at 33, 38; Lester B. Lave, *Benefit-Cost Analysis: Do the Benefits Exceed the Costs?*, *in* RISKS, COSTS, AND LIVES SAVED 104, 116–17 (Robert W. Hahn ed., 1996); Laurence H. Tribe, *Policy Science: Analysis or Ideology?*, 2 PHIL. & PUB. AFF. 66, 95–97 (1972).

[270] ACKERMAN & HEINZERLING, *supra* note 268, at 98–102.

[271] FISHKIN, THE VOICE OF THE PEOPLE, *supra* note 7, at 40–43, 161–63.

means public-regarding deliberation, or at least deliberation that is open to claims about the public good rather than individual interest. "Deliberation" approximates Habermas's ideal speech situation: "All arguments deemed relevant by anyone in the discussion are given as extensive a hearing as anyone wants and people are willing to consider all the arguments offered on their merits."[272] Peter Dienel and Ortwin Renn, who have been active in promoting "Planning Cells" (a kind of citizen jury) in Germany, write:

> Participants of Planning Cells have no defined constituents to whom they are obliged. They are selected to embody and represent the interests of all citizens rather than a specific group. It is interesting to note that citizens occupy the role of advocates of the common good almost from the beginning of the sessions.[273]

Ned Crosby, who has spearheaded the development of the citizen jury format in the United States, argues that it embodies a social contract approach to policy choice, as opposed to a utility calculation or a political power approach.[274] This roughly tracks the familiar distinction between civic republicanism, cost-benefit analysis, and interest-group pluralism.

Wendy Kenyon and co-authors, reflecting on the critical literature on CV surveys and on the proposals for policy deliberation formats, write:

> Economists and others have suggested that a CV questionnaire asks respondents the wrong question, assuming that consumers think about environmental goods (public goods) in the same way as they do about private goods. . . . [T]he use of [citizen juries] as a method of preference revelation allows . . . deliberation on the environmental issue in terms of what is best for society.[275]

In short, the civic republican critique can be framed as a contrastive claim: that citizen juries, deliberative polls, planning cells, and other policy deliberation formats have the appropriate public-regarding structure for citizen involvement in policymaking, and welfare polls do not.

The response to this objection is, I believe, straightforward. *Policy deliberation formats and welfare polls are complementary, not mutually exclusive.* Policy deliberation formats ask what an appro-

---

[272] *Id.* at 40.

[273] Peter C. Dienel & Ortwin Renn, *Planning Cells: A Gate to "Fractal" Mediation*, *in* FAIRNESS AND COMPETENCE IN CITIZEN PARTICIPATION, *supra* note 7, at 117, 126. On the use of Planning Cells in Germany, elsewhere in Europe, and (in one instance) the United States, and the authors' involvement, see *id.* at 130–36, and Kenyon et al., *supra* note 7, at 558.

[274] Crosby, *Using the Citizens Jury® Process*, *supra* note 7, at 401–02.

[275] Kenyon et al., *supra* note 7, at 559.

priate policy would be; if the answer is that an appropriate policy would be sensitive to considerations of well-being, in one or another way, welfare polls can then be brought to bear (by the citizens themselves, or by the administrators implementing the citizens' policy judgment).

To begin, it should be stressed that weak welfarism—the moral framework I have relied upon in arguing for welfare polls—is perfectly consistent with the use of policy deliberation formats. There are a range of possible arguments, intrinsicalist and instrumentalist, for why some governmental decisions[276] should be informed by policy deliberation formats. Intrinsicalist arguments say that public-spirited citizen participation is an intrinsic good. Instrumentalist arguments say that policy deliberation formats advance important purposes that are conceptually distinct from participation itself—in particular, producing better governmental decisions.[277]

Are these arguments persuasive? If so, what specific kinds of decisions should policy deliberation formats be convened to address? These are interesting and difficult questions that are well beyond the scope of this Article. The point I wish to emphasize here is that the intrinsicalist and instrumentalist arguments for policy deliberation formats are arguments that the weak welfarist can *accept*. The arguments do not entail nonwelfarism.

To begin, it is plausible that participation in government is an intrinsic *welfare* good—in other words, that an individual's well-being is enhanced by her participation, or, more precisely, that individuals with full information and nondistorted preferences would self-interestedly prefer participation for its own sake.[278] And in any event the weak welfarist will surely find very plausible the suggestion that, in some contexts, policy deliberation formats are *instrumentally* valu-

---

276 I say "some" because, clearly, policy deliberation formats cannot be used to address all governmental decisions. These are too numerous and, in some cases (most obviously, adjudicative decisions involving particular individuals), legal or moral constraints might well preclude the use of such formats.

277 *See* Thomas Christiano, *The Significance of Public Deliberation*, *in* DELIBERATIVE DEMOCRACY, *supra* note 6, at 243, 244–46 (describing different kinds of instrumental and intrinsic values that public deliberation might have).

278 I have elsewhere argued that participation in government does not, in fact, have intrinsic welfare significance. *See* Adler, *supra* note 104, at 283–88. But that position is not entailed by weak welfarism. Weak welfarism, itself, does not necessitate the conclusion that participation lacks intrinsic welfare value. *That* depends on whether individuals with full information and so forth prefer participation—and one can adopt weak welfarism without taking a position on that.

Therefore, the weak welfarist can accept the intrinsicalist case for policy deliberation formats. And in any event, as discussed immediately below, she can certainly accept the instrumental case for policy deliberation formats.

able—instrumentally valuable as measured by the very criteria of weak welfarism. In particular, someone who holds a moral view with the structure $\{W^*, F_1, \ldots, F_M\}$, where W^* is overall welfare and $M=0$, can certainly believe that the use of policy deliberation formats, in some area, will lead to policies that are better in light of $\{W^*, F_1, \ldots, F_M\}$. This is true even if $M=0$—that is, even if the weak welfarist is a utilitarian. Utilitarians might be confident, for example, that policy deliberation formats will often (if not always) come to recognize the moral significance of overall welfare and will help control bureaucracies that too often promote their own interests rather than overall welfare. But if utilitarians can support policy deliberation formats on instrumental grounds—and I suggest they can—then, a fortiori, the various nonutilitarian theories that are also subsumed within the framework of weak welfarism certainly can.

All of this is a little abstract. Someone who holds a moral view with the structure $\{W^*, F_1, \ldots, F_M\}$ can believe that policy deliberation formats are morally required in some contexts. If so—given the interplay, if not identity, between morality and law—she can also believe that policy deliberation formats are legally required in some contexts. But how, concretely, can she believe all of this and *also* support welfare polls?

Consider that the subset of governmental decisions appropriately informed by "deliberative polls," "citizen juries," or other policy deliberation formats might be first-order or second-order decisions.[279] First-order decisions are decisions to adopt particular policies—to issue particular regulations, build particular projects, and so forth. Second-order decisions are decisions about which procedures agencies should employ in making their first-order decisions. Congress faces a second-order choice in deciding whether to enact a statute that instructs a health and safety agency to employ cost-benefit analysis or, instead, some competing procedure for choosing regulatory measures. Congress also faced a second-order choice when it debated the enactment of a generic cost-benefit supermandate.[280] An agency faces a second-order choice in deciding what general approach to take in implementing its existing, open-ended, statutory mandate.

Clearly, *second-order* policy deliberation formats and welfare polls are complementary, not mutually exclusive.[281] For example, a citizen jury might be convened on the question of whether an agency

279 This should not be confused with my earlier distinction between second- and first-order techniques for designing surveys. *Supra* text accompanying notes 203–06.

280 *See supra* note 225 and accompanying text.

281 *See, e.g.*, Crosby, *Using the Citizens Jury® Process*, *supra* note 7, at 413 (noting that citizen juries can help design structure of policy analysis).

should follow a procedure of QALY-maximization; if the jury and the agency decide affirmatively, the agency will then (down the line) use QALY surveys. In other words, the citizens participating in the second-order policy deliberation format may come to the conclusion that the agency should employ an administrative decision procedure that attends to human welfare or aspects of welfare. The agency, in line with the citizens' recommendation, will implement the welfare-regarding decision procedure, and in so doing may well employ welfare polls.

First-order policy deliberation formats and welfare polls are also compatible, although this is less obvious. Imagine that a policy deliberation format has been convened to determine the content of an agency regulation. For example, an environmental agency drafting a pollution-control regulation for pollutant *X* might ask a citizen jury, "What should be the permissible level of pollutant *X*?" The citizen jury, if truly deliberative, would not decide that question blindly, without mediating concepts. Nor, if truly informed, would it decide the question as a matter of common knowledge, without outside data. Instead—civic republicans would surely agree—the permissible level ought to be picked in light of public goals and values, and the jury should gather and discuss information so as to determine what pollution level best advances those goals and values. For example, the jury might determine that optimal pollution policy should promote overall well-being, should avoid large skews in the distribution of well-being, and should protect moral rights.[282] If so, the jury would then need to consult welfare polls as well as revealed-preference data, to help determine the extent to which lowered pollution levels (via lowered fatality risks and fewer diseases) increase overall welfare and (if the effects of pollution are disproportionately borne by the poor) diminish distributive skews. CV or QALY valuations of fatality risk and disease would surely be helpful to the jury at this stage in its reasoning process.

In sum: Welfare polls are only *part of* a broader process of administrative decisionmaking. Administrative decisionmaking is a kind of legal reasoning that should begin with applicable legal sources (statutes, the Constitution, executive orders). At some point in the reasoning process, the administrator may conclude that she is legally required to invite citizen participation, or that doing so is legally per-

---

[282] *See supra* Part IV.A (discussing "weak welfarism," which allows for moral relevance of rights or distributive considerations as well as overall well-being); ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 59–60, 155–56 (noting that well-being might be "currency" of distributive justice).

missible and morally required.[283] So, potentially, citizens as well as the administrator will become involved in deliberations about what the agency ought to do. But these deliberators, citizen or official, may well conclude that facts about human well-being are relevant to their decision—that they are legally required to attend to well-being, or legally permitted and morally required to do so. If weak welfarism is indeed the correct moral view then the official or citizen deliberators *should* and (if not too misguided) *will* often reach this conclusion. And once the official or citizen deliberators conclude that well-being information is relevant, they may well then decide (for just the sorts of reasons presented earlier in this Article) that welfare surveys are a useful way to secure that information.

In other words, the civic republican critics of welfare polls are attacking a caricature. The critics seem to assume that the polls are self-bottoming—that they come into play from the beginning, displacing public-spirited deliberation. But that is not how welfare polls work, or at least not how they should work. The self-interested perspective that the respondents to welfare polls are asked to adopt is not the starting point for policy. Rather, deliberating citizens or officials should start from an impartial starting point. From that starting point, however, they may deliberate their way to the proposition that the effect of policies on the interests of some group of individuals is a legally or morally relevant concern. Reasoning as public-regarding citizens, and bracketing our narrow interests, we might conclude that government should, inter alia, be sensitive to the effect of its policies on our narrow interests. Welfare polls ask respondents to take a self-interested perspective because the informational value of these polls is in elucidating what individual interest consists in—information that a fully moral and impartial deliberation process might well (indeed, surely should) take account of.

V

NEW DIRECTIONS FOR WELFARE POLLS

The overwhelming majority of welfare polls to date have employed three general formats: the CV, QALY, and happiness formats. This Article has therefore focused on these formats—discussing the role that CV, QALY, and happiness surveys currently play in agency decisionmaking and other aspects of administrative govern-

[283] To repeat, as argued above, *see supra* text accompanying notes 276–82, the proposition that citizen participation will be morally or legally required in some contexts is *consistent* with weak welfarism.

ance;[284] canvassing other possible roles;[285] describing a variety of possible improvements to the survey enterprise within this trio of formats;[286] and defending welfare polling from objections from two quarters, the revealed-preference camp within economics and the deliberative democracy camp within political theory.[287]

But should the trio of CV, QALY, and happiness surveys be changed—perhaps supplemented by different welfare polling formats, or perhaps even pared back? The dominance of these three approaches is, to a large extent, an historical accident. CVs originate in the Kaldor-Hicks view of policy analysis long dominant, despites its flaws, in applied economics.[288] QALYs became popular, in substantial part, because public health researchers were disinclined to monetize health and longevity and wanted a scale that was different from the WTP/WTA scale.[289] And the rise of happiness surveys is largely a matter of scholarly cycles within psychology, shifting from the traditional focus on negative states such as anxiety and depression to include positive states as well.[290] Only recently have economists and others with an interest in policymaking latched on to happiness surveys.

Despite their accidental origins, CV, QALY, and happiness surveys turn out to be vital governmental tools, as discussed above. QALY and happiness surveys capture important aspects of well-being; CVs are yet more inclusive. But, at a minimum, there is much room to supplement these standard welfare polling formats with new ones, in at least four different ways.

First, more survey work should be undertaken to characterize the multidimensional structure of welfare. Such a characterization is both directly useful in policymaking—for example, in clarifying the full range of well-being impacts that, ideally, a cost-benefit analysis or annual well-being report should cover—and indirectly useful in guiding the ongoing enterprise of welfare polling.

---

284 *See supra* Part II.

285 *See id.*

286 *See supra* Part III.

287 *See supra* Part IV.

288 For critical discussion of Kaldor-Hicks efficiency, with citations to other critical literature, see, for example, Adler, *supra* note 104, at 249. My defense of cost-benefit analysis rests on the criterion of overall well-being, which is different from the Kaldor-Hicks criterion. *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 21–23, 39–61; Adler & Posner, *Implementing Cost-Benefit Analysis*, *supra* note 39, at 272–76; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39, at 190–91, 204–38. On the continuing dominance of the Kaldor-Hicks criterion in applied economics, see, for example, JUST ET AL., *supra* note 232, at 646.

289 *See* Adler, *supra* note 25, at 14.

290 *See, e.g.*, Ryan & Deci, *supra* note 31, at 142.

There is a long philosophical tradition, going back to Aristotle, of drawing up lists of the different aspects of well-being.[291] Martha Nussbaum, John Finnis, and James Griffin are prominent contemporary philosophers who have continued this enterprise. Nussbaum's list is: life, bodily health, bodily integrity, use of the "senses, imagination, and thought," emotions, practical reason, affiliation, interaction with other species, play, and control over one's political and material environment.[292] Finnis's is: life, knowledge, play, aesthetic experience, sociability, practical reasonableness, and religion.[293] Griffin's is: accomplishment, autonomy, understanding, enjoyment, and deep personal relations.[294] Philosophers tend to work through reflection and discussion with each other, not surveys, and their efforts to describe the multiplicity of well-being dimensions can usefully be supplemented through systematic survey work.

For a rare example of such work, consider the World Health Organization's efforts to develop a questionnaire—the so-called "WHOQOL instrument"—designed to capture all aspects of quality of life, not just health as traditionally conceived.[295] Research groups in fifteen different countries were involved in the effort. Each research group conducted focus groups with the general population to develop a preliminary list of the "aspects of life that they considered contributed to its quality."[296] Based on these focus groups, a preliminary questionnaire was developed, consisting of numerous facets of well-being and matching questions designed to determine the respondent's achievement with respect to each facet. The preliminary questionnaire was administered to at least 300 respondents in each of the fifteen countries. The WHO researchers then performed a statistical analysis of this data—for example, looking at correlations within and between facets—to arrive at a final facet structure and matching ques-

---

291 *See, e.g.*, SABINA ALKIRE, VALUING FREEDOMS 78–84 tbl.2.12 (2002); L.W. SUMNER, WELFARE, HAPPINESS, AND ETHICS 45–80 (1996). This philosophical work grows out of the objectivist approach to well-being, but the difference between that approach and one that looks to fully informed preferences may be slight. *See supra* text accompanying note 105.

292 MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT 78–80 (2000).

293 JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 85–90 (1980).

294 JAMES GRIFFIN, VALUE JUDGMENT 29–30 (1996).

295 For discussions of the structure, development, use, and psychometric properties of the WHOQOL, see Amy E. Bonomi et al., *Validation of the United States' Version of the World Health Organization Quality of Life (WHOQOL) Instrument*, 53 J. CLINICAL EPIDEMIOLOGY 1 (2000); Silvija Szabo, *The World Health Organization Quality of Life (WHOQOL) Assessment Instrument*, *in* QUALITY OF LIFE AND PHARMACOECONOMICS IN CLINICAL TRIALS 355 (Bert Spilker ed., 2d ed. 1996); and The WHOQOL Group, *The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties*, 46 SOC. SCI. & MED. 1569 (1998).

296 The WHOQOL Group, *supra* note 295, at 1570.

tionnaire. The final WHOQOL structure consists of twenty-four facets or dimensions of well-being, grouped into six domains.[297]

TABLE 1
THE WHOQOL QUALITY OF LIFE DOMAINS (6) AND FACETS (24)

Physical Domain	Psychological Domain	Independence Domain	Social Domain	Environment Domain	Spiritual Domain
1 Pain and Discomfort	4 Positive Feelings	9 Mobility	13 Personal Relationships	16 Physical Safety and Security	24 Spirituality
2 Energy and Fatigue	5 Thinking, Learning, Memory, and Concentration	10 Activities of Daily Living	14 Social Support	17 Home Environment	
3 Sleep and Rest	6 Self-Esteem	11 Dependence on Medication or Treatments	15 Sexual Activity	18 Financial Resources	
	7 Body Image and Appearance	12 Working Capacity		19 Health and Social Care: Availability and Quality	
	8 Negative Feelings			20 Opportunities for Acquiring New Information and Skills	
				21 Participation in and New Opportunities for Recreation/ Leisure	
				22 Physical Environment	
				23 Transport	

Survey work such as the WHOQOL, together with more traditional philosophical efforts by philosophers such as Nussbaum, Griffin, and Finnis, gives policymakers and welfare pollsters an overarching framework for categorizing welfare impacts and attendant surveys and metrics. A second, new direction for welfare polling consists of survey work designed to characterize the fine structure of particular well-being dimensions. Consider the WHOQOL domains:

297 The facet and domain structure of the current WHOQOL is summarized in Bonomi, *supra* note 295, at 4. The full names of the twenty-four facets are provided in The WHOQOL Group, *supra* note 295, at 1576–78.

QALY surveys typically cover at most the first and third domains and part of the second (omitting positive feelings, self-esteem, and body image), and wholly ignore the last three.[298] Happiness surveys, if focused on the respondent's positive and negative affects, cover only parts of the first and second domains. If focused on the respondent's sense of satisfaction with his life, they cover the whole WHOQOL map, but only in an indirect way.[299] CVs cover the whole map, but tradeoff biases may interfere with monetary valuation of goods such as friendship, self-esteem, and spirituality.

Nonmonetary quantitative measures, such as QALYs, that focus on a subset of the WHOQOL dimensions are useful (1) in reducing tradeoff biases and other cognitive distortions that may especially affect CVs, and (2) in circumventing the variable marginal utility of money.[300] In theory, both of these benefits could also be realized through an inclusive nonmonetary measure of well-being—an idea I will return to in a moment—but dimension-specific measures such as QALYs have the further benefit of providing fine-grained information about well-being. A QALY survey asks the respondent to imagine changes in her health state, holding nonhealth characteristics fixed. To put this in terms of the WHOQOL, imagine a QALY survey that focuses on facets 1, 2, 3, 5, 8, 9, 10, 11, and 12, which would be reasonably typical. These are the health dimensions; the other WHOQOL facets are background dimensions. Then the QALY survey will instruct the respondent that the number one means the best possible state *with respect to all of the health dimensions*, together with some set of background characteristics (most straightforwardly, the respondent's actual characteristics), and zero means the worst possible state *with respect to all of the health dimensions*, together with that same set of background characteristics.[301] The respondent will then be asked

298 Adler, *supra* note 25, at 50–51.

299 *See supra* note 60 (describing standard conception of happiness as mix of positive affect, negative affect, and satisfaction with life). Some happiness surveys focus more on affects, others on satisfaction with life. *See* Andrews & Robinson, *supra* note 34, at 62–63, 70–76. The respondent to a standard, global question about life-satisfaction (for example, that used in the Eurobarometer series, *see supra* text accompanying note 33) could reasonably answer by thinking about his achievements with respect to the totality of WHOQOL dimensions. However, the question does not ask him to do that, nor does it otherwise draw his attention to those dimensions. Further, an answer to a question such as "How satisfied are you with your life?" is not direct evidence of the respondent's overall evaluation of his actual life circumstances, because the answer is mediated by the respondent's perception of those circumstances. This is in contrast with the QALY method, which specifies a state for the respondent to evaluate.

300 *See supra* text accompanying notes 46–47, 163.

301 This is not quite accurate. Zero on the QALY scale is usually used to mean death or a state no better than death, as opposed to the worst possible health state. But the basic point that the QALY approach asks the respondent to vary health characteristics, rather

to place an actual or hypothetical health state of hers on this 0–1 scale, using some technique such as the standard gamble or time tradeoff. Contrast that with an inclusive exercise that tells the respondent that one is an ideal state *with respect to all of the WHOQOL dimensions*, and that zero is the worst possible state *with respect to all of the WHOQOL dimensions*. Presumably the QALY scale is more sensitive to relatively small differences in health states than the more inclusive scale would be. To give one illustrative example: Respondents to QALY surveys are able to distinguish angina and pancreatitis (a recent collection of QALY scores includes a score of 0.75 for angina and 0.81 for pancreatitis).[302] Would they be able to distinguish between an ideal life marred only by angina and one marred only by pancreatitis?

This is a long-winded way of arguing for the benefits of nonmonetary, dimension-specific well-being measures such as QALYs. But presumably those benefits are not unique to the first three WHOQOL domains (where we have at least partial coverage, with QALY and happiness surveys) as opposed to the last three. Researchers should experiment with social quality-of-life metrics and surveys, which would correspond to the fourth WHOQOL domain and would quantify on a nonmonetary scale (e.g., a 0–1 scale analogous to QALYs) the contribution that different sorts of social interaction make to well-being. The same goes for "personal environment" quality-of-life metrics, corresponding to the fifth WHOQOL domain, and maybe even spirituality metrics.

A third avenue for new work in welfare polling involves altering the respondent's perspective. QALYs, happiness surveys, and CV polls all ask the respondent about *her own life*. QALYs (when administered to laypeople rather than doctors) ask the respondent to compare outcomes in which *she* experiences a given health state with outcomes in which *she* experiences death or perfect health. Happiness surveys ask the respondent how happy or satisfied she is with her life—or, in Kahneman's framework, how good her current experiences are.[303] CV surveys (if appropriately restricted to screen out disinterested preferences) will focus on the respondent's WTP/WTA for changes in the world that affect her.

---

than both health *and* background characteristics, and is therefore presumably more sensitive to health changes, remains true. *See* Adler, *supra* note 25, at 47–52.

302 Tufts-New Eng. Med. Ctr. Inst. for Clinical Research & Health Policy Studies, Cost Effectiveness Analysis Registry, Preference Weights 1976–1997, http://www.tufts-nemc.org/cearegistry/data/phaseIpreferenceweights.pdf (last visited Oct. 23, 2006).

303 *See supra* text accompanying notes 67–68 (discussing Kahneman's approach to measuring experience).

Welfare polls need not adopt this respondent-centered focus. In principle, a poll might ask the respondent to consider someone else's life—the life of a hypothetical person, or the actual life of some other person—and to express that life's well-being value on some quantitative scale. Questions of this sort would seem to have both advantages and disadvantages vis-à-vis the current formats. They might be particularly good at focusing respondents on the particular aspects of well-being that the surveyor wants quantified. (Current formats, when asking the respondent about a hypothetical state substantially at variance with her actual condition, risk triggering a protest reaction wherein the respondent outright refuses to value the hypothetical state or, more subtly, fails to fully consider it.) Relatedly, non-respondent-centered formats might be particularly effective at encouraging respondents to refine their general judgments and views about the nature of well-being (since the task of evaluating their own particular lives has been excised), and would be easier to mesh with group-based well-being surveys than the current surveys.[304]

On the other hand, non-respondent-centered formats risk inviting answers that are not welfare-focused. For example, a respondent asked to rank various health histories of other persons might express his judgments about how "healthy" these individuals are in some sense detached from well-being, or his moral judgment about a fair allocation of health care resources among them, rather than a judgment about the welfare-goodness of the different lives.[305] Whether, on balance, non-respondent-centered formats provide substantial new well-being information beyond that afforded by QALYs, CVs, and happiness surveys is something we can only determine by experimenting with these novel formats.

A fourth and final area for experimentation with new welfare polling formats involves developing a genuine interpersonal utility scale. This scale would be inclusive, covering all of the dimensions of well-being. And, ideally, it would be a cardinal scale, capable of representing well-being levels (crucial for questions about the distribution of well-being) and well-being differences (crucial for determining overall well-being).[306] Imagine that the interpersonal utility scale ranges from zero to one. Then, if person A is at level 0.3 and person B is at level 0.4, this means that B is better off than A. If a policy changes A's level to 0.31 and B's to 0.37, this means that the policy

304 *See supra* text accompanying notes 211–15 (discussing possibility of group-based welfare polls).

305 *See supra* text accompanying note 118.

306 *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 39–43 (discussing how different moral theories require measurement of well-being levels and differences).

decreases overall well-being (because the 0.01-unit positive difference it makes to *A*'s well-being is less than the 0.03-unit negative difference it makes to *B*'s).

Designing surveys to develop an interpersonal utility scale would mean taking a view about the meaningfulness and content of interpersonal welfare comparisons—something I have written about extensively elsewhere but lack space to discuss at any length here.[307] The short answer is that interpersonal comparisons *are* meaningful and (according to one plausible account deriving from work by Harsanyi) reduce to convergent well-informed preferences regarding lotteries over possible lives. Such preferences could, in principle, be elicited through a standard gamble question analogous to the QALY standard gamble: The respondent is told that the number one corresponds to the best possible life and zero to the worst possible life. A particular possible life is described, and the respondent is asked to place it on the 0–1 scale by expressing the probability p that makes her indifferent as between that life and a lottery with a p probability of the best possible one and a $1 - p$ probability of the worst possible one.[308] Different respondents might express different indifference probabilities for the same life, which raises a problem of aggregation. But that problem is no different from the aggregation problem currently faced by QALY surveys.[309]

A more troubling objection, already alluded to, is that the survey protocol just described might be cognitively overwhelming for most respondents. QALYs limit the cognitive demand by holding nonhealth dimensions fixed. In contrast, a parallel but more inclusive format that defined one as the best state with respect to all twenty-four of the WHOQOL dimensions, and zero as the worst state with respect to all twenty-four, might overtax the imaginative abilities of many humans. One answer might be to use internal consistency checks to screen out responses from those who are overwhelmed; another, to use visual aids and other cognitive aids to help respondents grasp the protocol and provide meaningful answers; a third, to change elicitation technique.[310] What the change of technique would be is itself a matter for experimentation. But just as the time tradeoff

307 *See id.*; Adler, *supra* note 25, at 17–24; Adler, *supra* note 104, at 289–302; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 39, at 204–09.

308 *Cf.* Andrews & Robinson, *supra* note 34, at 66, 74 (describing Cantril's Ladder Scale, which asked respondents to locate themselves on ladder with top representing "the best possible life for you" and bottom "the worst possible life for you").

309 *See, e.g.*, Adler, *supra* note 25, at 41; Dolan, *supra* note 210, at 160.

310 *See supra* text accompanying notes 153–62; *supra* note 164 (discussing these devices within context of existing welfare polls).

method has emerged, within QALY research, as a better technique for eliciting QALY values than the QALY standard gamble—once thought to be the gold standard[311]—so the standard gamble method for eliciting interpersonal utility numbers that I have described might not be the most practicable way to do so. If interpersonal comparisons are indeed a matter (as per Harsanyi) of preferences over lotteries of possible lives, then the standard gamble is theoretically most compelling—but that does not preclude other utility-elicitation methods that might be easier to use and might produce values that approximate those obtained by the standard gamble method.[312]

Why an interpersonal utility scale? We already possess, in dollars, an inclusive scale—one that covers all twenty-four of the WHOQOL dimensions. Why an inclusive *nonmonetary* scale? The answer, above all, has to do with the variable marginal utility of money. WTP/WTA values are only a rough measure of changes in well-being.[313] If P is willing to pay $100 for a policy, and Q is willing to accept $50 in exchange for the policy, it does not necessarily follow that the policy increases overall well-being—that the positive impact on P's welfare outweighs the negative impact on Q's. P's WTP might exceed Q's WTA because P is wealthier, or an ascetic, or for some other reason reaps a relatively small welfare improvement from incremental dollars.

Despite the inaccuracy of WTP/WTA amounts in tracking well-being, cost-benefit analysis is probably the best currently available welfarist tool for policy analysis—either the traditional form of cost-benefit analysis that values all impacts with WTP/WTA amounts, or a hybrid form that values some using WTP/WTA and others using QALY-to-dollar or similar conversions.[314] But survey data valuing welfare impacts on an interpersonal utility scale could be very helpful in structuring cost-benefit analysis—for example, by helping set dis-

311 *See supra* note 162 and accompanying text.

312 One possibility may be to attempt to identify a plurality of well-being dimensions that interact in a simple (additive or multiplicative) way to determine overall well-being; to use surveys to establish weights for the different dimensions, for example by asking about compensating changes in one dimension for changes in another; and then to calculate the interpersonal utility number for a given state as a function of the dimension-specific measures for that state. *See, e.g.*, Payne et al., *supra* note 120, at 257–58 (discussing multi-attribute utility theory techniques, such as eliciting "swing weights" for different dimensions).

313 Similarly, money measures are only a rough index of welfare levels. Consider wealth, the most obvious monetary proxy for an individual's welfare level: P may have more wealth than Q but be worse off, given his physical condition or lack of access to public goods.

314 *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 39, at 62–123, 142–46; Adler, *supra* note 25.

tributive weights, or by guiding the choice of QALY-to-dollar or happiness-to-dollar conversion factors.[315] More generally, in any context where money values might be skewed by unusually high or low marginal utilities, information from interpersonal utility surveys could help recalibrate those values.

To summarize, the trio of QALYs, CVs, and happiness surveys can be usefully supplemented by (1) surveys such as the WHOQOL that attempt to characterize the multidimensional structure of well-being; (2) dimension-specific analogues to QALY and happiness surveys, covering dimensions such as social life; work life; housing, neighborhood quality, and other aspects of an individual's physical environment; or recreational opportunities; (3) surveys that ask respondents to evaluate others' lives, not their own; and (4) survey work to measure welfare impacts on an inclusive, nonmonetary, interpersonal utility scale. And what of the thought that these novel formats might displace QALYs, happiness surveys, or CVs? For now, that thought is premature. In two or three generations, perhaps, welfarist policy analysis might dispense with money as its commensurating device and express costs and benefits in terms of interpersonal utility units. But—given the huge amount of information about money values provided by behavioral data as well as existing CV studies, and the absence of a comparable body of interpersonal utility information—that prospect seems distant. The enterprise of welfare polling needs to be expanded, in the ways suggested in this Part, rather than redirected away from the current three formats that have proven so popular.

CONCLUSION

Scholarship about law and government sometimes leads, sometimes lags behind, real advances in governance. The latter is the case, I want to suggest, for CV, QALY, and happiness surveys. CV research now comprises a whole subfield of applied economics, with dedicated practitioners, lots of primary surveys, a large secondary literature, and a real role in governmental decisionmaking at a number of federal agencies. QALYs are equally important in the fields of public health and health economics, and the results of QALY surveys now frequently figure in cost-benefit analysis at the FDA. Happiness surveys have long been an area of interest for psychologists, and are now a hot topic for economists.

315 *See* Adler, *supra* note 25, at 57–74 (discussing use of interpersonal utility numbers in setting QALY-to-dollar conversion factor).

And yet, at a somewhat higher level of academic generality—at the level of public-law scholarship and political theory, where general normative questions of governmental design are pursued—these survey enterprises have been largely ignored. The contrast with normative work on citizen juries, deliberative polling, and other policy deliberation formats is striking. Here, the quantity of high theory vastly outstrips the actual amount of polling work undertaken or its actual impact to date on governmental decisionmaking.

This Article has sought to redress the theory-practice imbalance. I have provided a new construct—the welfare poll—that, I hope, provides a unifying perspective on QALYs, CVs, and happiness research. The construct is useful both in generating recommendations about survey practices and governmental applications within this trio of survey formats, and in suggesting new formats.

Welfare polls can provide substantial information about the sources and nature of human well-being. This information is not fully provided by revealed-preference studies, and its legal and moral relevance is (I have argued) unimpeachable. The informational content of welfare polls does, of course, depend on whether survey respondents tell the truth, make a sufficient effort, have sufficient facts, have preferences that are not too distorted, understand the question asked, are focused on well-being, and so on. I have systematically surveyed these sorts of conditions and have argued that they can be satisfied sufficiently well.

I have also stressed that welfare polls are complementary with, not opposed to, policy deliberation formats. The old duality of "citizen" versus "consumer" needs to be transcended. The Article, emphatically, is *not* an attack on citizen juries, citizen advisory commissions, or deliberative polls. But, reciprocally, the theorists of policy deliberation ought to recognize that survey instruments that secure information about well-being by inviting respondents to take a self-interested perspective on policy *also* are morally and legally defensible. Why assume that civic republican deliberation would end up denying the normative significance of welfare? The institutions and decision procedures that incorporate welfare polls—those described in Part I of this Article—are justifiable on the basis of a view, "weak welfarism," which citizens impartially deliberating about the aims of government surely could endorse.

There is a second, deep link between the existing literature on policy deliberation formats and the defense of welfare polling presented in this Article. Both embrace the premise that good governance will, at some point, require *asking people what they think* (be it about policy or about well-being), and creating favorable motiva-

tional, epistemic, cognitive, and communicative conditions for this discursive exercise. Both reject the traditional aversion within economics to survey data, and both are committed to improving citizens' judgments or preferences (about policy or well-being) by providing fuller information and by creating discursive structures that will encourage rationality, mental effort, and truthfulness. The development of survey techniques that improve preferences or judgments, and the very exercise of questioning citizens—not just observing their behavior—are vital to good governance. These are key premises of my defense of welfare polls, and are just as central for the many scholars who have argued in favor of citizen juries, citizen advisory commissions, deliberative polls, and other citizen-involving formats for policy deliberation.