J O I N T C E N T E R
AEI-BROOKINGS JOINT CENTER FOR REGULATORY STUDIES

Welfare Polls: A Synthesis

Matthew D. Adler*

Related Publication 06-04
February 2006

* Professor of Law, University of Pennsylvania Law School. Many thanks to Abby Wright for superb research assistance.

WELFARE POLLS: A SYNTHESIS

Matthew D. Adler*, University of Pennsylvania Law School, February 2006.

Abstract: "Welfare polls" are survey instruments that seek to quantify the determinants of human well-being. Currently, three "welfare polling" formats are dominant: contingent-valuation surveys, 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 actual and potential uses of welfare polls in government 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.

* Professor of Law, University of Pennsylvania Law School. Many thanks to Abby Wright for superb research assistance.

INTRODUCTION

What are the avenues for citizen participation in the administrative state? The traditional answers 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. But 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] they can show up and talk 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, statically "biased" sample of the public; and they will tend to be uninformed and to 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

[1] *See* 1 RICHARD J. PIERCE, JR., ADMINISTRATIVE LAW TREATISE 529-43 (4th ed. 2002)
[2] *See id.* at 415-24.
[3] *See id.* at 424-63.
[4] On agency use of informal public meetings, see, e.g., Daniel J. Fiorino, *Citizen Participation and Environmental Risk: A Survey of Institutional Mechanisms*, 15 SCI., TECH. & HUMAN VALUES 226, 230-31 (1990).
[5] *See* Mariano-Florentino Cuellar, *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.

democracy" or civic republican tradition in political theory.[6] The tradition envisions a particular kind of citizen participation: public-spirited, concerned to advance the public good rather than personal preferences or interests. A variety of concrete formats have been proposed for civic-republican citizen deliberation: citizen advisory committees or review panels, citizen juries, "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 -- requiring a highly structured process, led by convenors, whereby representative citizens become informed, deliberate, and then vote or state their views. These formats have been used only very occasionally by U.S. governmental entities, state or federal.[9]

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.

"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

---

[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); DELIBERATIVE DEMOCRACY: ESSAYS ON REASON AND POLITICS (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: PUBLIC OPINION AND DEMOCRACY (1995); James S. Fishkin, *Toward Deliberative Democracy: Experimenting with an Ideal*, in CITIZEN COMPETENCE AND DEMOCRATIC INSTITUTIONS (Stephen L. Elkin & Karol E. Soltan eds., 1999); Robert Luskin et al., *Considered Opinions: Deliberative Polling in Britain*, 32 BRIT. J. POL. SCI. 455 (2002). 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);
Ned Crosby, *Citizens Juries: One Solution for Difficult Environmental Questions*, in FAIRNESS AND COMPETENCE: EVALUATING MODELS FOR ENVIRONMENTAL DISCOURSE 157 (Orwin Renn et al. eds., 1995); Ned Crosby et al., *Citizens Panels: A New Approach to Citizen Participation*, 46 PUBLIC ADMIN. REVIEW 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, *supra*; THE POLL WITH A HUMAN FACE: THE NATIONAL ISSUES CONVENTION EXPERIMENT IN POLITICAL COMMUNICATION (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 INDIANA L.J. 904 (1998); Thomas 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 ENVT. & PLANNING C: GOV'T AND POLICY 557 (2001).

[8] *See* Applegate, *supra* note 7, at 925. 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 COMEPTENCE, *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., *Considered Opinions*, *supra* note 7, at 461.

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. "Welfare polls" start from the premise that welfare matters to governmental choice; citizens 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; *QALY surveys*, which ask citizens[10] to rank health states on a nonmonetary scale with 1 representing perfect health and 0 death; and *happiness surveys*, which ask citizens to rank their own "happiness" or "life satisfaction" on various nonmonetary scales. [11]

Each of these techniques has generated vast scholarly literatures.[12] Further, the results of CV surveys and, increasingly, QALY 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 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). [13]

Happiness surveys, the third leg of the welfare-polling triad, have yet to play the role in U.S. governmental practice that CV surveys 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 hasn't anyone written about welfare polls? More precisely, why hasn't anyone written about welfare polling as such? There is plenty of writing about QALYs, about CVs, about 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 seeing QALY, CV, 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.

---

[10] 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 24-25.
[11] *See infra* Part I (describing these formats).
[12] *See infra* notes 18, 22, 28.
[13] *See infra* Part II (describing use of CVs and QALYs by federal agencies).

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 the novel forms of participation worth discussing should be modeled on the civic-republican ideal.

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 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 evaluation is undertaken in Part IV. Synthesizing the material from previous parts, I provide a normative case for welfare polls, grounded in the moral and legal relevance of well-being. Part IV argues, 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; and, 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 get at 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 (CVS, QALYS, 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 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 solo, or to deliberate together about well-being? What scale should be used?

These sorts of questions will be examined below, in Part III and again in Part 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 the leading examples.

CV surveys were invented by environmental economists in the 1960s.[14] They are now conducted, not just for ecological goods, but for virtually every aspect of well-being: recreation, noise, smell, visibility, fatality risks, health states, psychological states, cultural amenities, and aesthetic values. 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 – randomly or nonrandomly selected, depending on the study design.[15] Mail, telephone, and in-person surveys are all common,[16] and the rise of the internet has created yet another possible way to administer CV surveys. Surveys are typically undertaken by academic researchers, usually applied economists, or by governmental agencies or contractors working for agencies.

It is estimated that thousands of CV surveys have been undertaken.[17] The secondary literature is correspondingly large.[18] Whole journals are focused on publishing primary CV studies or discussing methodology.[19]

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

---

[14] Good reviews of the CV technique include: IAN BATEMAN ET AL., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES: A MANUAL (2002) [hereinafter ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES] ; ROBERT CAMERON MITCHELL & RICHARD T. CARSON, USING SURVEYS TO VALUE PUBLIC GOODS: THE CONTINGENT VALUATION METHOD (1989); A. MYRICK FREEMAN, THE MEASUREMENT OF ENVIRONMENTAL AND RESOURCE VALUES: THEORY AND METHODS 161-87 (2003); Kevin Boyle, *Contingent Valuation in Practice*, in PATRICIA CHAMP ET AL., A PRIMER ON NONMARKET VALUATION 111 (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 (J.A. Hausman ed., 1993). Helpful recent literature reviews are: L. Venkatachalam, *The Contingent Valuation Method: A Review*, 24 ENVTL. IMPACT ASS. REV. 89 (2004); and Richard T. Carson et al., *Contingent Valuation: Controversies and Evidence*, 19 ENVTL. & RESOURCE ECONOMICS 173 (2001).

[15] *See infra* text accompanying notes 195-198.

[16] *See* ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 89-111; Patricia Champ, *Collecting Survey Data for Nonmarket Valuation*, in CHAMP ET AL., *supra* note 14 , at 59, 69-80.

[17] *See* Stale Navrud & Gerald J. Pruckner, *Environmental Valuation-To Use or Not to Use?*, 10 ENVT'L & RESOURCE ECON. 1, 8 (1997).

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

[19] *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 non-market valuation, particularly contingent valuation).

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; 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.[20]

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 –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. Isn't it, therefore, more accurate to characterize CVs as a weird 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 (such as recreation, smell, noise, health, psychological states or fatality risks) where self-interested preferences predominate – by 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.[21]

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.[22] 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.[23]

QALY surveys ask respondents to place a given health state on a zero-one scale, with zero representing death and one perfect health. These surveys are sometimes given to experts (namely, doctors) but expert surveys are now viewed skeptically in the field,[24] and QALY surveys of

---

[20] For a discussion of CV elicitation techniques, see ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 135-45; Boyle, *supra* note 14, at 135-43. Variations on the straight dichotomous choice question have also been developed, for example the "one-and-a-half" and "double bounded" dichotomous choice formats. *See* ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 141.

[21] *See infra* text accompanying notes 98-108.

[22] 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 n.1 (2006) (citing sources). On the size of the QALY literature, see *id.* at 3.

[23] *See id.* at 48-50.

[24] *See* Paul Dolan, *Whose Preferences Count?*, 19 MED. DECISIONMAKING 482, 482 (1998); G. Ardine de Wit, *Sensitivity and Perspective in the Valuation of Health Status: Whose Values Count?*, 9 HEALTH ECON. 109, 110 (2000).

laypersons (either patients or members of the general public) are the preferred technique[25] – 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.[26] 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 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 0 to 100.[27]

Let us turn, finally, to happiness surveys.[28] The U.S. General Social Survey, conducted annually or biannually for more than 30 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: "Taken all together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?"[29] A parallel large-scale survey conducted several times a year in European Union member states, the Eurobarometer series, asks "On the whole are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?"[30] Similar questions have been asked in a host of nonperiodic surveys, conducted by academics, governments, or by other organizations, in the U.S. and elsewhere, often involving very large samples.[31] 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 1-3, 1-7, or 1-10; or to subsume their happiness or life-satisfaction in one of an ordered set of categories ("very satisfied," "fairly satisfied," "not at all satisfied"). [32]

---

[25] *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, 1738-39 (A.J. Culyer & J.P. Newhouse eds., 2000)

[26] *See, e.g.*, J. Brazier et al., *A Review of the Use of Health Status Measures in Economic Evaluation*, 3 HEALTH TECH. ASSESSMENT, no. 9, at 114-32 (1999) (listing numerous QALY surveys, including both interviewer- and self-administered studies)

[27] On QALY elicitation methods, see J. Brazier et al., *supra* note 26, at 23-56 (1999); Dolan, *supra* note 24, at 1732-37.

[28] For overviews of happiness surveys and the literature they have generated, see BRUNO S. FREY & ALOIS STUTZER, HAPPINESS AND ECONOMICS: HOW THE ECONOMY AND INSTITUTIONS AFFECT HUMAN WELL-BEING (2002); Ed Diener et al., *Subjective Well-Being: Three Decades of Progress*, 125 PSYCH. 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. PSYCH. 141 (2001); 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).

[29] *See* David G. Blanchflower & Andrew J. Oswald, *Well-Being Over Time in Britain and the USA*, 88 J. PUB. ECON. 1359, 1363-66 (2004); GENERAL SOCIAL SURVEYS, 1972-2002: CUMULATIVE CODEBOOK vi, 179 (Nat'l Opinion Research Center, University of Chicago, February 2003).

[30] *See* Blanchflower & Oswald, *supra* note 29, at 1367-69; http://europa.eu.int/comm/public_opinion/description_en.htm.

[31] *See* 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 28, at 353, 353.

[32] *See* Andrews & Robinson, *supra* note 31, at 70-73.

Psychologists pioneered happiness research, and have undertaken most of these surveys, as well as generating much of the secondary literature on happiness. But happiness has recently become a hot topic in economics, and there is now a large and growing body of work by economists that analyzes the surveys to identify and quantify the determinants of happiness, discusses the econometrics of these inferences, or makes policy recommendations for increasing happiness.[33]

II. WELFARE POLLS: CURRENT AND POTENTIAL GOVERNMENTAL USES

Some intellectual tools that are influential in the academy never make it into the public sector. Welfare polls are not like that. This Part surveys their current governmental uses and interweaves a discussion of potential uses. My focus here is federal agencies, and to that extent is underinclusive, not including state governments,[34] to say nothing of governments abroad.[35]

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 OMB, and to conform all regulations to a cost-benefit standard where statutorily permissible.[36] Traditionally, the money valuations of goods employed by cost-benefit analysis were derived using "revealed preference" techniques, which look to market prices or non-transactional behaviors. However, as elaborated in Part IV, these techniques are far from perfect,[37] and agencies now regularly incorporate the results of CV surveys into cost-benefit analysis.

In particular, EPA routinely relies upon CV surveys in conducting cost-benefit analysis as part of rulemaking -- and it does so not only to quantify "non-use" values, but also "use values," in particular mortality risk, health effects, visibility, and water quality.[38] Agencies also perform cost-

---

[33] This work is summarized in FREY & STUTZER, *supra* note 28.

[34] State governments do use CV studies. See John B. Loomis, *Contingent Valuation Methodology and the US Institutional Framework*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 618-20. Oregon relied on QALYs in a notorious episode 15 years ago, but it appears that state governments do not use QALYs much. *See* Adler, *supra* note 22, at 3-4. As far as I'm aware, happiness surveys have not been yet employed by either state or federal governments.

[35] "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." Francois Bonnieux & Pierre Rainelli, *Contingent Valuation Methodology and the EU Institutional Framework*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 593. By contrast, QALYs have a less central role in health policymaking in the U.S. government as compared to certain foreign governments. *See* Adler, *supra* note 22, at 4.

[36] *See* Exec. Order No. 12,291, 3 C.F.R. 127 (1982); Exec. Order No. 12,866, 3 C.F.R. 638 (1994). 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 (forthcoming 2006); Matthew D. Adler & Eric A. Posner, *Implementing Cost-Benefit Analysis When Preferences are Distorted*, in COST-BENEFIT ANALYSIS: LEGAL, ECONOMIC, AND PHILOSOPHICAL PERSPECTIVES 269 (Matthew D. Adler & Eric A. Posner eds., 2001); Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165 (1999).

[37] *See infra* text accompanying notes 227-248.

[38] 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 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

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, Fish and Wildlife Service, and National Parks Service.[39]

How, exactly, do CV studies percolate into these agencies' cost-benefit analyses? Who does 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.[40]

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

---

were employed to estimate values, the list here is surely not comprehensive. *See* Regulatory Impact Analysis for the Stationary Internal Combustion Engine (RICE) NESHAP, at 8-24 to -26 (February 2004) (mortality risk); Effluent Limitations Guidelines and New Source Performance Standards for the Concentrated Aquatic Animal Production Point Source Category, 67 Fed. Reg. 57872, 57913-14 (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. 38888, 38945-46 (June 22, 2000) (mortality risk); Asbestos Worker Protection, 65 Fed. Reg. 24806, 24817 (April 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, VII-47, VII-57 (Feb. 10, 2000) (mortality risk, chronic bronchitis, 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 to -39 (Aug. 29, 1996) (mortality risk); Effluent Limitations Guidelines, Pretreatment Standards, and New Source Performance Standards: Metal Products and Machinery, 60 Fed. Reg. 28210, 28261 (May 30, 1995) (mortality risk); Municipal Solid Waste Landfills: NSPS, at 12-10 (April, 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).

[39] *See* Loomis, *supra* note 34, at 613-18.

[40] The alternative techniques described in this paragraph are all forms of "benefits transfer," generally discussed in FREEMAN, *supra* note 14, at 453-56; Stale Navrud, *Value Transfer and Environmental Policy*, in INTERNATIONAL YEARBOOK OF ENVIRONMENTAL AND RESOURCE ECONOMICS 2004/2005: A SURVEY OF CURRENT ISSUES 189 (Tom Tietenberg & Henk Folmer eds.); and Randall S. Rosenberger & John B. Loomis, *Benefit Transfer*, in CHAMP ET AL., *supra* note 14, at 445.

costs or benefits relative to the status quo. The CV technique, too, 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 chief 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. The FDA then translates that QALY number into a dollar figure, to be incorporated in cost-benefit analysis, using a conversion factor such as $100,000 or $300,000 per QALY.[41]

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.[42] The answer, very briefly, is twofold. First, certain cognitive distortions that interfere with CV studies can be circumvented by QALY surveys.[43] 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.

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 non-traditional money measures may be less likely to go astray – to pick policies that actually reduce overall welfare – than cost-benefit analysis incorporating the traditional, WTP/WTA measures.

Potential Use: Happiness-Based Cost-Benefit Analysis

Happiness surveys, still confined to the academy in United States, have various potential roles in governmental decisionmaking. One such role parallels 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

[41] *See* Adler, *supra* note 22, at 57-60.
[42] *See id.* at 24-42.
[43] On this point, see also *infra* text accompanying notes 151-152.

Amsterdam airport from happiness surveys.[44] 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. An equation explaining happiness values (the dependent variable) as a function of these characteristics (the independent variables) was then estimated. Happiness was negatively correlated with noise, and positively correlated with income.

The ratio of the coefficients established a noise-to-dollar tradeoff ratio R. That ratio 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 money equivalent is RN. Since R is itself the product of the noise-to-happiness tradeoff rate H and the happiness-to-dollar tradeoff rate D[45], 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 amounts units into dollars. This approach for monetizing welfare 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 martial status. And they point out that the method, in principle, is applicable to any life change – the relevant characteristics need simply to be documented, along with happiness levels and income, in a happiness survey.[46]

The rationale for using happiness-to-dollar conversions, in lieu of or in addition to traditional WTP/WTA amounts,[47] 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 accurate only if relocation costs are low.[48] 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

---

[44] *See* BERNARD VAN PRAAG & ADA FERRER-I-CARBONELL, HAPPINESS QUANTIFIED: A SATISFACTION CALCULUS APPROACH 219-38 (2004).

[45] D=dollars/happiness. H=happiness/noise. R=dollars/noise=H×D. RN=(HN)×D.

[46] *See* 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. *See* Bruno S. Frey & Alois Stutzer, *Happiness Research: State and Prospects*, 62 REV. SOC. ECON. 207, 220-23 (2005).

[47] Happiness-to-dollar conversions based on happiness surveys that focus on the respondents' positive and negative affects are clearly distinct from valuations 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 valuations, although even here the valuations may be different given, for example, different biases that may affect the two formats. *See infra* text accompanying note 54 (discussing affective versus life-satisfaction conceptions of happiness surveys).

[48] *See* VAN PRAAG & FERRER-I-CARBONELL, *supra* note 44, at 220-24.

distortions and wealth effects aren't expected to be large, happiness-to-dollar conversions furnish useful, additional information for the cost-benefit analyst.

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 non-monetary 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 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).[49] Fatalities furnish a crude index of public health and safety. Air pollutants, workplace toxins, and dangerous foods or additives 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 those just mentioned.[50] This point – it should be stressed – applies not merely to the agencies that have traditionally interested public health researchers, including FDA, but to *any* agency implementing some statute requiring the maximization of health or safety – for example, EPA or OSHA.[51]

---

[49] *See* Matthew D. Adler, *Risk, Death, and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. 1293, 1414-17 (2003); ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___.

[50] 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, e.g., 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. American Trucking Ass'ns, 531 U.S. 457, 464-71 (2001) (interpreting Section 109 of the Clean Air Act to preclude a 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," often proposed by public health scholars. *See* Adler, *supra* note 22, at 8-10.

[51] *See* American Trucking Ass'ns v. EPA, 175 F.2d 1027, 1039-40 (D.C. Cir., 1999) (suggesting that EPA cure constitutional difficulties in the Clean Air Act by measuring the benefits of air pollution regulations using QALYs), *rev'd in part*, 531 U.S. 457 (2001); Rafael Ponce et al., *Quality Adjusted Life Years (QALYs) and Dose-Response Models in Environmental Health Policy Analysis: Methodological Considerations*, 274 SCIENCE OF THE TOTAL ENVIRONMENT 79 (2001) (discussing use of QALYs for risk assessment with heterogeneous health impacts).

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.[52] 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.[53] 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 costs. Total WTP of the impoverished persons who would benefit from the scheme is $100 million, and total WTA of the taxpayers who would fund the scheme is also $100 million, so the transfer itself is seen as a wash. But of course, if the marginal utility of money decreases with income, the transfer may increase overall welfare.

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, which I cannot consider in detail here. 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 merely in contexts where cost-benefit analysis is legally precluded, but even in contexts where it is not – 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).[54] A number of scholars have proposed that government evaluate policies by determining which one maximizes happiness. For example, Bruno Frey and Alois Stutzer write:

---

[52] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36; *Implementing Cost-Benefit Analysis*, *supra* note 36; *Rethinking Cost-Benefit Analysis*, *supra* note 36.

[53] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___; *Implementing Cost-Benefit Analysis*, *supra* note 36, at 286-87, 300-05; *Rethinking Cost-Benefit Analysis*, *supra* note 36, at 224.

[54] 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 SOCIAL INDICATORS RESEARCH 189, 200 (1997). Maximizing positive and negative affect is not the same as maximizing overall well-being, because well-being isn't just a matter of mental states. *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ch. 2. 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.

> 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 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.[55]

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" which maximizes the sum of individual utilities, in turn calculated as the logarithm of individual income.[56] This function is mathematically tractable and has the property of being increasing in income at a decreasing rate, thus justifying progressive taxation, but is not based in any systematic research into how income translates into well-being. Griffiths argues that the survey data on the correlation between income and happiness confirms the basic supposition of declining marginal income utility, but should be used to determine the specific shape of the social welfare function.[57]

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 surveys.[58] 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.[59]

The Nobel prize-wining psychologist Daniel Kahneman suggests a policy-analytic technique which is broadly similar to HLE analysis.[60] 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

---

[55] FREY & STUTZER, *supra* note 28, at 176. *See also Life Satisfaction: The State of Knowledge and Implications for Government* 35-36 (Strategy Unit, Cabinet Office, United Kingdom, Dec., 2002), available at http://www.e-democracy.gov.uk/knowledgepool/default.htm?mode=1&pk_document=28 (suggesting that happiness surveys might be used for policy analysis)

[56] *See* Thomas D. Griffith, *Progressive Taxation and Happiness*, 45 BOSTON COLLEGE L. REV. 1363, 1367-68 (2004).

[57] *See id.* 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 28, at 176-77; VAN PRAAG & FERRER-I-CARBONELL, *supra* note 44, at 239-317.

[58] Ruut Veenhoven, *Happy Life-Expectancy: A Comprehensive Measure of Quality-of-Life in Nations*, 39 SOCIAL INDICATORS RESEARCH 1, 29-31 (1996).

[59] 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.

[60] *See* 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)

individual statements of overall happiness or life-satisfaction.[61] 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.

Actual Use: Natural Resource Damage Assessment

To 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 can help answer the question.

The "Superfund" statute (CERCLA),[62] the Oil Pollution Act (OPA),[63] and the Clean Water Act (CWA)[64] create a federal liability regime for oil spills and other releases of hazardous substances that harm publicly owned natural resources.[65] The statutes define natural resources broadly as "land, fish, wildlife, biota, air, water, groundwater, drinking water supplies, and other such resources."[66] 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 Department of the Interior (DOI) guide assessments under CERCLA and CWA,[67] and the National Oceanic and Atmospheric Administration (NOAA) regulates assessments under OPA.[68]

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.[69] 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.[70] Current DOI regulations explicitly permit the use of CVs in Type B (those not using a standard computer model) assessments.[71] While NOAA regulations do not explicitly authorize the use of CVs, an appendix lists CVs as a potential tool.[72]

---

[61] *See id.*; Daniel Kahneman, *Objective Happiness*, in WELL-BEING, *supra* note 28, at 3; Daniel Kahneman et al., *Back to Bentham: Explorations of Experienced Utility*, 112 QUARTERLY J. ECON. 375 (1997).
[62] 42 U.S.C. § 9601 et seq. (2000).
[63] 33 U.S.C. § 2701 et seq. (2000).
[64] 33 U.S.C. § 1321 (2000) (creating liability for discharges into navigable waters).
[65] See Kevin R. Murray et al., *Natural Resource Damage Trustees: Whose Side Are They Really on?*, 5 ENVIRONMENTAL LAWYER 407, 413-18 (1999), for a description of the relationship between the three statutes.
[66] 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 TULANE L. REV. 417 (1997); and Dale B. Thompson, *Valuing the Environment: Courts' Struggles with Natural Resource Damages*, 32 ENVTL. L. 57 (2002).
[67] 43 C.F.R. Part 11 (2005).
[68] 15 C.F.R. Part 990 (2005).
[69] *See* State of Ohio v. United States Dep't of the Interior, 880 F.2d 432, 474-81 (D.C. Cir. 1989).
[70] *See* Kenneth Arrow et al., *Report of the NOAA Panel on Contingent Valuation* (Jan. 11, 1993).
[71] *See* 43 C.F.R. § 11.83 (c)(2)(vii) (2000).
[72] *See* 61 Fed. Reg. 498-99 (Jan. 5, 1996). The D.C. Circuit has upheld the use of CVs in OPA assessments. *See* General Electric v. United States Dep't of Commerce, 128 F.3d 767, 773-74 (1997).

CV surveys were in fact employed in the damage assessment for the 1989 Exxon Valdez oil spill in Prince William Sound, Alaska -- where the multi-billion dollar damage estimates generated by the studies helped induce Exxon's large settlement.[73] CVs have also been used in a number of less high-profile cases.[74]

Potential Use: Damages, Fines, Prices, Taxes

Using CVs in natural resource damage assessments is only the tip of the iceberg. Whenever monetary damages are meant to be compensation 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. The federal regime includes a schedule for smaller pollution spills, and a number of states also use schedules to calculate natural resources damages.[75]

"Damages," of course, need not be limited to natural resources. They might be damages 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.[76] More feasible is expert testimony at the damages phase that is grounded in CV surveys.[77] These surveys might also inform workers compensation schedules, which to a limited extent cover not merely lost wages and out-of-pocket medical costs (where welfare polls wouldn't be useful), but also pain and suffering.[78]

To be sure, tort law and workers compensation is 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: 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"

---

[73] *See* Loomis, *supra* note 34, at 622; Richard T. Carson et al., *Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill*, 25 ENVT'L AND RESOURCE ECON. 257, 278 (2003).

[74] *See* Loomis, *supra* note 34, at 620-22; Navrud & Pruckner, *supra* note 17, at 11-13. *Cf.* Thompson, *supra* note 66, at 70-87 (noting that CV studies have been prepared in a number of natural resource damages case, 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 non-use values, estimating non-use values has not been their exclusive use and certainly need not be. *See* Loomis, *supra* note 34, at 621. The DOI regulations, in fact, prefer CVs for use values as opposed to non-use values. *See* 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. *See* 61 Fed. Reg. at 499 .

[75] *See* Anderson, *supra* note 66, at 457-63; Murray B. Rutherford et al., *Assessing Environmental Losses: Judgments of Importance and Damage Schedules*, 22 HARV. ENVT'L L. REV. 51, 76-80 (1998). Indeed, CVs were, at least at one point, used as inputs to the federal schedule. *See* Rutherford, *supra*, at 78; Loomis, *supra* note 34, at 620.

[76] *See* Rutherford et al., *supra* note 75, at 75-76 (discussing personal injury scheduling abroad and proposals to do so in the U.S.)

[77] *See, e.g.*, Brendan I. Koerner, *What's Your Happiness Worth?*, LEGAL AFFAIRS (Jan./Feb. 2004) (describing an economist who testifies as an expert in tort cases regarding the size of hedonic damages, using a WTP measure).

[78] *See* Rutherford et al., *supra* note 75, at 72-73.

have the flavor of a sanction that is established ex post. But these are pretty 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.[79]

Environmental trading markets (ETMs) present a similar potential application for CV studies. The basic idea of such markets, as in "cap-and-trade" pollution regimes or wetlands banks, 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.[80] A crucial issue for any ETM is the "currency" for the market. Are actors allotted limits, and assigned credits for reductions in environmental 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 currently employed very much in ETMs, might seem unrealistic given the costs of CV studies.[81] 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. Concretely: 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 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 $1 credit for every dollar-equivalent its emissions are below the allotment. This approach, by 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's happiness-based monetization of noise.[82]

Actual Use: Environmental Impact Statements and Rulemaking Notices

---

[79] Denise Mauzerall et al., *Charging NOx Emitters for Health Damages: An Exploratory Analysis* (CESifo Working Paper No. 1442, April 2005), available at http://SSRN.com/abstract=706782

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

[81] *See id.* at 634.

[82] *See* Heinz Welsch, *Preferences over Prosperity and Pollution: Environmental Valuation Based on Happiness Surveys*, 55 KYLKOS 473 (2002). Note also that QALYs or happiness units could, in principle, be used as the ETM currency rather than dollars.

The National Environmental Policy Act (NEPA), enacted in 1970, 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."[83] This brief language has made environmental scoping a pervasive aspect of federal agency decisionmaking – as evidenced by the large body of federal case law about the impact-statement requirement[84] and by the sheer number of impact statements and preparatory documents. It is estimated that roughly 500 impact statements and 50,000 "environmental assessments" – preliminary documents which consider whether the federal action requires an impact statement – are issued by federal agencies every year.[85]

It is now clear that purely human impacts do not trigger NEPA. An action must have a effect on the physical environment – on "the air, land, [or] water"[86] -- to come within the scope of the statute.[87] 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.[88] 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.[89] 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.[90]

NEPA is the quintessential example of a "communication forcing" statute. It doesn't give substantive priority to environmental considerations, but simply requires agencies to publicly 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. A yet broader "communication forcing" statute is §553(c) of the Administrative Procedure Act, generally obliging an agency to provide a public "statement of basis and purpose" – typically in the Federal Register -- when it enacts legally binding rules.[91] Where

---

[83] 42 U.S.C. §4332 (2000).
[84] *See generally* DANIEL R. MANDELKER, NEPA LAW AND LITIGATION (Release 3, 2005) (reviewing NEPA case law).
[85] *See* Nicholas C. Yost, *NEPA Deskbook* 9 (3d ed. 2003).
[86] Douglas County v. Babbitt, 48 F. 3d 1495, 1505 (9th Cir. 1995).
[87] *See* Metropolitan Edison Co. v. People Against Nuclear Energy, 460 U.S. 766, 772-73 (1983); 40 C.F.R. §1508.14 (2005).
[88] *See* Balitmore Gas & Elec. Co. v. Natural Resources Defense Council, Inc., 462 U.S. 87, 106-07 (1983).
[89] *See, e.g.*, CHARLES H. ECCLESTON, ENVIRONMENTAL IMPACT STATEMENTS: A COMPREHENSIVE GUIDE TO PROJECT AND STRATEGIC PLANNING 175-81 (2000); R.K. JAIN ET AL., ENVIRONMENTAL ASSESSMENT 239-79 (2002).
[90] *See* Loomis, *supra* note 34, at 614-16; email from Charles Eccleston to Matthew D. Adler, 12/22/05. *See also* L. Venkatachalam, *supra* note 14, at 89 (claiming that "the CV method is a widely used nonmarket valuation method especially in the areas of [inter alia] environmental impact assessment"); DAVID JAMES, THE APPLICATION OF ECONOMIC TECHNIQUES IN ENVIRONMENTAL IMPACT ASSESSMENT (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.
[91] 5 U.S.C. §553(c) (2000). Rules of "agency organization, procedure, or practice" are exempted from the notice-and-comment requirement, as are the categories of legislative rules described in 5 U.S.C. §553(a) – although in the case of rules relating to "public property, loans, grants, benefits, or contracts," agencies often voluntarily choose to follow the §553(c) procedures.

welfare polls figure in internal agency deliberations preceding the enactment of a rule, the §553(c) statement accompanying the proposal may well discuss the polls.

The most important example involves cost-benefit analysis – because, to date, the leading function of welfare polls has been to inform cost-benefit analysis. 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.[92] Other agencies, relying on CVs in their analytic documents, will then publish these in the Federal Register. The communicative role of welfare polls, here, flows from their function in policy analysis, but is conceptually distinct. It's easy enough to imagine a nonpublic process of policy analysis, as indeed can occur for decisions that aren't §553(c) rules or covered by NEPA or some other communication-forcing mandate.

Potential Use: Other Governmental Communications

Governmental 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 governmental communications to include well-being facts as part of the transmitted information – potentially bringing welfare polls into play.

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 versus 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.[93]

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. [94] Finally, numerous scholars propose "environmental accounts" that would

---

[92] *See* Adler, *supra* note 22, at 58 n.195 (citing Federal Register statements where FDA has used QALY-to-dollar conversions).

[93] Daniel Kahneman et al., *Toward National Well-Being Accounts*, 94 AMER. ECON. ASS'N PAPERS AND PROCEEDINGS 429, 433 (2004).

[94] Ed Diener, *Subjective Well-Being: The Science of Happiness and a Proposal for a National Index*, 55 AMER. PSYCHOLOGIST 34 (2000).

track the state of the environment, and some have suggested that CVs could be used in preparing these.[95]

III. DO WELFARE POLLS PROVIDE SUBSTANTIAL INFORMATION ABOUT WELFARE?

The preceding part described a range of contexts in which governmental 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 that are self-interested (this rules out moral and other disinterested preferences) and that survive some degree of idealization.[96] 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 non-ideal in the sense of being (2) poorly informed,[97] (3) distorted by cognitive bias, or (4) not the result of sufficient mental effort. Further, 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 which 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 the 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.

Moral and Other Disinterested Preferences

---

[95] *See, e.g.*, Navrud & Pruckner, *supra* note 17, at 15-16.

[96] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___; *Implementing Cost-Benefit Analysis*, *supra* note 36; *Rethinking Cost-Benefit Analysis*, *supra* note 36, at 197-204; Matthew D. Adler, *Beyond Efficiency and Procedure: A Welfarist Theory of Regulation*, 28 F.S.U. L. REV. 241, 264-67 (2000).

[97] There is some dispute about whether "ideal" preferences are fully-informed preferences, objectively good preferences, or nonadapative preferences. *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___. The first two approaches may well be quite close, *see id.*, and since I find them more persuasive than the nonadaptive view of idealization, I will focus here on how to use welfare polls to provide relatively full information to respondents.

Much of the scholarly criticism of CV surveys concerns moral preferences.[98] This is a particular worry 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, for example 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 – for example, degradation in some wilderness area that she never plans to visit, or the extinction of an endangered species that she has never seen.

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.[99] For example, one well-known study by Desvouges 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 2,000 birds, the second 20,000, the third 200,000. Mean WTP values for the different surveys were virtually identical, despite the 10-fold differences in the size of the bird population saved: $80, $78, and $88.[100] 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.[101]

Moral preferences plausibly explain, or help explain, both of these anomalies.[102] 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

---

[98] 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, e.g., Peter A. Diamond & Jerry A. Hausman, *On Contingent Valuation Measurement of Nonuse Values*, in CONTINGENT VALUATION, *supra* note 14, at 3; Peter A. Diamond & Jerry A. Hausman, *Contingent Valuation: Is Some Number Better than No Number?*, 8 J. ECON. PERSP. 45 (1994); Daniel Kahneman et al., *Economic Preferences or Attitude Expression? An Analysis of Dollar Responses to Public Issues*, 19 J. RISK AND UNCERTAINTY 1 (1999). This scholarship is surveyed in Carson et al., *supra* note 14, at 177, 181-84.

[99] *See* Carson et al., *supra* note 14, at 181-84; Venkatchalam, *supra* note 14, at 95-102.

[100] *See* William H. Desvousges et al. , *Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliabiltiy*, in CONTINGENT VALUATION, *supra* note 14, at 91, 100.

[101] *See* Kahneman et al., *supra* note 98, at 213 (discussing this study).

[102] *See* sources *supra* note 98; Jonathan Baron, *Biases in the Quantitative Measurement of Values for Public Decisions*, 122 PSYCH. 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 ECOL. ECON. 35 (2005); Clive L. Spash & Nick Hanley, *Preferences, Information and Biodiversity Protection*, 12 ECOL. 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).

enterprise of contingent valuation and thus "protest votes": refusals to answer, or zero WTPs, or (once again) infinite WTAs.

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" – generically useful for well-being – this is a plausible, if rough, way to capture the impact of the policies on her well-being.[103] 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 citizen 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 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?[104] 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

[103] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ch. 3 (defending WTP/WTA as a rough proxy for interpersonally comparable well-being); Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 36, at 214-38.
[104] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___.

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[105] – because of the mistaken orthodoxy in economics that simply denies a distinction between disinterested and self-interested preferences.[106]

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 non-use works pretty well.[107] Those who do not physically interact with some environmental amenity should not be asked about 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.[108] The Department of Agriculture doesn't ask animal rights activists for their WTP/WTA to have slaughterhouses closed; the Postal Service doesn't ask religious activists for their WTP/WTA not to have pornography shipped through the mail; the FDA doesn't ask libertarians for their WTP/WTA not to have paternalistic regulations imposed on others.

The discussion to this point has focused on the CV instrument. What about QALYs 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).[109] In principle, the same is true of

---

[105] 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 14, 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 ENVIRONMENTAL PREFERENCES, *supra* note 14, at 196-99. But 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 – the tendency of respondents to overstate what they actually would pay. *See, e.g.*, James T. Murphy & Thomas H. Stevens, *Contigent Valuation, Hypothetical Bias, and Experimental Economics*, 33 AGR. & 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?* (U. Mass., Working Paper No. 2003-2), at 1 (describing "cheap talk script" that enjoined respondents not to articulate the fair price for a good).

[106] *See, e.g.*, W. Michael Hanemann, *Valuing the Environment through Contingent Valuation*, 8 J. ECON. PERSP. 19, 33 (1994).

[107] I say "pretty well," not perfectly, because nonuse values subsume not merely existence values but also option values – self-interested preferences to preserve some environmental good that the respondent doesn't 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.

[108] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at __; *Implementing Cost-Benefit Analysis*, *supra* note 36, at 282; Diamond & Hausman, *Is Some Number Better than No Number?*, *supra* note 98, at 59.

[109] 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, *supra* note __, at 26-27, 39-41. Unlike CVs, PTOs may well be an appropriate way to elicit citizen

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 0-1, 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[110]; 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 happiness on a 1-7 scale 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 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 states, or about her use of environmental goods – would seem to naturally invite a self-interested perspective. More research on the issue is certainly needed. But it should be noted that the critics of QALYs and happiness studies have not identified extreme value or scope anomalies analogous to those that affect CV studies for non-use values. The possibility of moral preferences is not a theme in the critical literature, here. 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.

Information

The prestigious NOAA Panel on Contingent Valuation recommends: "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."[111] This bland advice conceals a host of complexities.

CV 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, for example about substitute and complementary goods.[112]

> 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

---

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 being mutually exclusive. *See infra* text accompanying notes 259-264. 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 PTO are not.

[110] *See* Adler, *supra* note 22, at 11-13.

[111] *See* Arrow et al., *supra* note 70, at 32-33.

[112] For discussions about the provision of information in CV surveys, see, e.g., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 308-10, 331-2; Boyle, *supra* note 14, at 123-33; Boyle & Bergstrom, *supra* note 105, at 193-95; Alistair Munro & Nick D. Hanley, *Information, Uncertainty, and Contingent Valuation*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 258; Venkatachalam, *supra* note 14, 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).

> 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.[113]

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.[114]

Even in the face-to-face format, the information provided in CV surveys is nothing close to the "full information" that is normative for welfare. 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,[115] 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.[116]

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.[117] Part of the solution, presumably, is to use information-provision devices (such as helpful visual aids) that facilitate comprehension and processing.[118]

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 welfare relevant rather than attributes over which respondents have merely instrumental preferences – as means to some further end. For example, health or safety agencies that perform cost-benefit analyses of policies to reduce ambient, food, or workplace toxins don't 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

---

[113] Boyle, *supra* note 14, at 121.
[114] See Munro & Hanley, *supra* note 112, at 259-60; Venkatachalam, *supra* note 14, 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.
[115] For a discussion, see THOMAS L. CARSON, VALUE AND THE GOOD LIFE 219-39 (2000).
[116] *See, e.g.*, ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 309; Payne et al., *supra* note 112, at 255.
[117] *See generally* Boyle and Bergstrom, *supra* note 105, at 195 (arguing for more research on information effects and fewer ad hoc practices).
[118] *See* Payne et al., *supra* note 112, at 255. *Cf.* Arrow et al., *supra* note 70, at 55 (stating that pictures and other visual aids can be helpful in providing information, but can also generate unwanted effects); Boyle, *supra* note 14, at 127 (same).

about WTP/WTA for the relevant end-result of these changes, namely increased numbers of game or fish or increased catch rate.[119]

A yet more systematic version of this idea – that the informational base for surveys should include only 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.[120] 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. The problem of incomplete information is often discussed, here, with reference to the choice between patient and general population surveys. Various non-informational 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 likelier to behave strategically.[121] On the other hand, patients – by virtue of their direct experience of the health state – will be better informed about it.

> 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 ….[122]

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.[123] 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.[124]

---

[119] *See Implementing Cost-Benefit Analysis*, *supra* note 36, at 283; Matthew D. Adler, *Fear Assessment: Cost-Benefit Analysis and the Pricing of Fear and Anxiety*, 79 CHI.-KENT L. REV. 977, 1009-10 (2004).

[120] On conjoint analysis in the CV context, see, e.g., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 248-95; Thomas P. Holmes & Wiktor L. Adamowicz, *Attribute-Based Methods*, in CHAMP ET AL., *supra* note 14, at 171.

[121] *See, e.g.*, COST-EFFECTIVENESS IN HEALTH AND MEDICINE 99-100 (Marthe R. Gold et al. eds., 1996); Dolan, *supra* note 25, at 1739.

[122] COST-EFFECTIVENESS, *supra* note 121, at 100.

[123] *See* Brazier, *supra* note 26, at 57-81;Dolan, *supra* note 25, at 1731-32, 1744-45.

[124] *See* David H. Feeny et al., *Health Utilities Index*, in QUALITY OF LIFE AND PHARMACOECONOMICS IN CLINICAL TRIALS 239 (Bert Spilker ed., 1996).

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.[125]

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 to information provision is more prosaic.[126] 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.[127] Part of a solution, here, 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.

Information provision is not a concern of the standard happiness surveys. 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.[128]

Kahneman's competitor proposal to the standard happiness-survey methodology[129] 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 bit of 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.

---

[125] *See infra* Part V (describing the WHOQOL survey).
[126] *See* Munro & Hanley, *supra* note 112, at 276-77.
[127] *See, e.g.,* Maxine E. Dakins, *The Value of the Value of Information*, 5 HUM. & ECOL. RISK ASSESSMENT 281 (1999).
[128] Schwartz & Strack, *supra* note 28, at 63.
[129] *See supra* notes 60-61 and accompanying text.

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. But *however* "preference distortions" are defined, it seems clear that respondents to welfare polls are often in their grip.

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,[130] 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 effect).[131] 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 thus endowed). [132] 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.[133] 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.[134] Whether tradeoff resistance is really a preference distortion depends on large issues about the incommensurability of welfare dimensions that I cannot pursue here.[135] But there is a plausible case that certain aspects of tradeoff resistance are irrational. Finally, a variety of other distortions

---

[130] *See* SCOTT PLOUS, THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 84-188 (1993).

[131] *See, e.g.*, James K. Hammitt & John D. Graham, *Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?* 18 J. RISK & UNCERTAINTY 33 (1999); Baron, *supra* note 102, 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 (1998).

[132] 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. ENVT'L ECON. & MGMT. 426 (2006). There is a large theoretical literature about the sources of this disparity. Two good discussions are FREEMAN, *supra* note 14, at 43-94; and Robert Sugden, *Alternatives to the Neo-classical Theory of Choice*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 152.

[133] Amos Tversky & Daniel Kahneman, *Loss Aversion in Riskless Choice: A Reference-Dependent Model*, 106 QUARTERLY 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 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 affects choice under certainty.

[134] *See* Baron, *supra* note 102; Payne et al., *supra* note 112, at 257-58; Adler, *supra* note 22, at 37-38.

[135] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___; Matthew Adler, *Incommensurability and Cost-Benefit Analysis*, 146 U. PA. L. REV. 1371 (1998).

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.[136]

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.[137] The "time tradeoff" format asks respondents for the time span spent in a perfectly healthy state that is equivalent to the time in a diseased state. This format demands a trade-off between life and health, one that respondents sometimes resist.[138] 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 zero-one scale, even where the health states being valued are quite similar.[139]

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."[140] 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.[141]

Schwartz and Strack, in a comprehensive critical review of happiness surveys, identify a number of recurrent distortions. 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 (thus 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, before you came to the university."[142]

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

---

[136] *See* ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 138-39; Ian J. Bateman et al., *Willingness-to-Pay Question Format Effects in Contingent Valuation Studies*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 511, 512-16; Boyle, *supra* note 14, at 137-43; Venkatachalam, *supra* note 14, at 105-10.

[137] *See* Brazier et al., *supra* note 26, at 30-34.

[138] *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).

[139] *See* Brazier et al., *supra* note 26, at 34-35.

[140] Kahneman, *Back to Bentham*, *supra* note 61, at 381.

[141] *See* Kahneman, *Objective Happiness*, *supra* note 61, at 21

[142] *See* Schwarz & Strack, *supra* note 28, at 62-74.

probability distortions by employing devices to help respondents better grasp what probabilities mean, such as verbal analogies, pie charts, risk ladders, or graph paper with blacked-out squares.[143] Probability aids are routinely used in eliciting QALY values when the standard gamble format is employed.[144] Research on the efficacy of these aids in reducing probability distortions is mixed.[145] 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 competitor 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. [146] 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.[147] 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, which as discussed above are in wide use.

A second possible approach to preference distortions is to change 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 payment cards or the open-ended question format, and can be reduced if not eliminated by shifting to other formats.[148] Carthy finds that a novel "chained" method for determining WTP/WTA for the risk of death reduces probability distortions[149]; and Hammitt and Graham find the same for a novel "indifference risk" elicitation method that holds constant the price of a safety device and varies the risk reduction.[150]

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

---

[143] *See* Phaedra S. Corso, *Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation*, 23 J. RISK & UNCERTAINTY 165, 169-70, 177-79 (2001).
[144] *See* Brazier et al., *supra* note 26, at 25.
[145] *See* Corso, *supra* note 143, at 169-70, 177-79.
[146] *See* Horowitz & McConnell, *supra* note 132, at 440-42.
[147] *See* Robin S. Gregory, *Valuing Environmental Policy Options: A Case Study Comparison of Multiattribute and Contingent Valuation Survey Methods*, 76 LAND ECON. 151 (2000); Payne et al., *supra* note 112, at 257-58.
[148] *See* sources cited *supra* note 136. 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 especially strongly associated with certain formats; if those biases produce a particularly large degree of preference distortion, shifting to a different format may be justified.
[149] *See* 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).
[150] *See* Hammitt & Graham, *supra* note 131, at 47, 57-58.

trade off certain health states that vary in their duration. Probability distortions are thereby circumvented.[151]

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.[152] Part of the advantage is that QALY surveys never ask respondents to value health or life in money – a particular 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 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 times 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 $(.15)\times 10{,}000\times(1/12)=125$ QALYs. If the policy abates the toxin for a year rather than a month, then the QALY formula would automatically value that impact as 12 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 12 times larger.

Kahneman's proposal to use hedonic surveys to value moments of experience, and then have policymakers (not citizens) value temporally extended experiential episodes by aggregating momentary values, represents a closely analogous proposal to circumvent preference distortions by shifting welfare polling formats: in this case, shifting away from traditional happiness surveys.[153]

Mental Effort

---

[151] *See* Adler, *supra* note 22, at 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 SG as gold standard has faded during years of practice. TTO 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 trade-off values, then shifting to the time trade-off technique will not wholly eliminate the underlying biases. But probability weighting, at least, should go away. *See* Bleichrodt, *supra* note 138, at 453.

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).

[152] *See* Adler, *supra* note 22, at 24-42, 69-83.

[153] 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, either those whose values are transparently distorted, or those whom additional questions suggest have especially distorted preferences; or (2) using calibration factors. On the first approach, see, e.g., 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 131, at 50-52; Murphy & Stevens, *supra* note 105, at 186. On the second, see, e.g., van Osch et al., *supra* note 151.

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.[154] It is therefore quite possible that respondents to welfare surveys will economize on mental effort. Problems of rational apathy, exhaustively discussed by economists in contexts such as voting,[155] readily carry 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 mental effort -- strategic laziness, as it were -- is little discussed by the welfare-polling literatures.[156] 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.[157]

Nor is mental effort a key focus of the general literature on survey design. There is a large subliterature on survey nonresponse.[158] But survey or individual item nonresponse – an outright refusal to answer the question -- is only one manifestation of 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

---

[154] The *locus classicus* for discussions of bounded rationality is, of course, Herbert Simon's scholarship. *See* HERBERT SIMON, MODELS OF BOUNDED RATIONALITY: ECONOMIC ANALYSIS AND PUBLIC POLICY (1982); HERBERT SIMON, MODELS OF BOUNDED RATIONALITY (1997). For more recent work in this area, see, e.g., ARIEL RUBENSTEIN, MODELING BOUNDED RATIONALITY (1998); GERD GIGERENZER ET AL., SIMPLE HEURISTICS THAT MAKE US SMART (1999).

[155] *See* DENNIS C. MUELLER, PUBLIC CHOICE III 303-332 (2003).

[156] *See* Adamowicz, *supra* note 18, at 435. By contrast, the issue has been more fully discussed in the literature on policy deliberation. Deliberation, after all, involves making a substantial effort to think about some issue. *See, e.g.*, Luskin et al., *supra* note 7, at 456-61.

[157] *See, e.g.*, Kahneman, *Objective Happiness*, *supra* note 61, at 21; Schwarz & Strack, *supra* note 28, at 63.

[158] *See, e.g.*, DON A. DILLMAN, MAIL AND INTERNET SURVEYS: THE TAILORED DESIGN METHOD (2d ed. 2000); ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 113-16, 177-79; ROBERT M. GROVES & MICK P. COUPER, NONRESPONSE IN HOUSEHOLD INTERVIEW SURVEYS (1998); SURVEY NONRESPONSE (Robert M. Groves et al. eds., 2002).

differentiate among a set of diverse objects in ratings," "saying `don't know' instead of reporting an opinion," and "randomly choosing among the response alternatives offered."[159]

If (1) lazily constructed valuations were normatively on a par with effortful valuations, or if (2) lazy valuations were statistically unbiased estimates of effortful valuations, then welfare pollsters wouldn't 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 for P than O_2 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[160]; therefore actual, distorted, poorly informed preferences will not (in some cases) be unbiased estimates of idealized preferences; therefore preferences constructed without efforts to absorb information or to participate in debiasing won't (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), writes:

> [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.[161]

This purely statistical technique may or may not be adequate to the problem of outright nonresponse, but it offers no solution to the problem that all or many responses may be marred by apathy. A solution to that problem, presumably, will mean designing surveys to trigger the motivators of mental effort. Krosnick suggests that these 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 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 behavior"; the respondent's sense of "accountability" to the interviewer; and the length of the interview.[162] The general, and substantial survey literature on steps to reduce outright nonresponse rates will be helpful here, since the norms, emotions, interests and so on which motivate respondents to take

[159] Jon A. Krosnick, *Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys*, 5 APPLIED COG. PSYCH. 213, 215 (1991).
[160] *See* sources cited *supra* notes 114, 145, 146.
[161] ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 178-79.
[162] *See* Krosnick, *supra* note 159, at 223-25.

the initial step of participating in a survey can also presumably be deployed to motivate more effortful participation.[163]

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, rational, effortful respondents might well lie to their interviewers. The problem of eliciting truthful valuations has been discussed at some length in the CV literature,[164] 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 the good plus 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. A higher stated WTP increases the chance that the good will be provided, without changing its cost to her. 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.[165]

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, pretty clearly, she best advances her preferences by voting for the option she actually prefers. Similarly, if an individual is asked, as per the dichotomous choice format, whether she prefers (a) having a good provided, at cost to her of $X, or (b) the status quo; and she believes that the governmental agency will pick one of these two options; and she believes that her statement to the interviewer raises the probability of the agency picking whichever option she claims to prefer; then she should truthfully articulate her preference.[166]

The research on "incentive compatibility" shows, further, that dichotomous-choice questions may not prompt truthful answers.[167] (Imagine a question where the respondent believes that, if the good is provided, she will not be taxed $X, but instead can choose voluntarily whether to pay anything. In that case, she has an incentive to answer the question affirmatively, even if $X

---

[163] For example, Dillman, an influential expert on nonresponse, argues that increasing surveying 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 158, at 14-21.
[164] Two particularly important discussions are Richard T. Carson et al., *Incentive and Informational Properties of Preference Questions* (University of California, San Diego, Dept. of Economics, Feb. 2000); and Robert Sugden, *Public Goods and Contingent Valuation*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 131. The literature is summarized in Carson et al., *supra* note 14 , at 189-93.
[165] Carson et al., *Incentive and Informational Properties*, *supra* note 164, at 28.
[166] *See id.* at 10-20.
[167] *See id.* at 11.

exceeds her true WTP.) Finally, and most strikingly, it emerges that *only* dichotomous-choice questions will be incentive compatible. The Gibbard-Satterthwaite theorem, a deep result in social choice theory, shows that no other format will be.[168]

The apparent conclusion from this line of research is that contingent-valuation survey designers can address strategic bias by using the incentive-compatible variants of dichotomous choice question. Indeed, the prestigious NOAA panel on CVs makes 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 strategic reason for the respondent to do other than answer truthfully ....[169]

But there is a problem. Consider the fully rational, self-interested individual whose valuation of cleaner air (say) is elicited by means of the 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. Then she will 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. 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."[170]

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."[171] 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 the 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 truth-telling 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 actually help the polling enterprise, here. Actual respondents may not realize that lying is in their

---

[168] More precisely, no other format can be incentive compatible unless the shape of preferences is restricted. *See id.* at 11, 31-32.
[169] Arrow et al., *supra* note 70, at 21.
[170] *See* Sugden, *supra* note 164, at 136-37.
[171] *Id.* at 137.

interest, or may not be up to the cognitive strain of keeping track of their lies.[172] As for sanctions: formal sanctions for lying are of course unavailable in the survey context, but social norms may come into play. "[T]he social setting of interviewer and interviewee [may] evoke[] norms of honesty."[173] 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).[174]

Further, even if some (perhaps large) fraction of survey respondents do 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 might be seen by policymakers as upper bounds to true WTP amounts. More generally, if stated valuation amounts are correlated, to some extent, with true valuations, rather than being random, then the statements will be useful to policymakers, to some extent, 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.[175] In one particularly extensive study of this sort, Carson and co-authors performed a meta-analysis of studies that provided both CV and revealed preference estimates of the same good. They found that the average CV/RP ratio was .89, .77, or .92, 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.[176] 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 Duchenne smiles; heart rate, blood pressure, and skin resistance measures of stress; psychosomatic illnesses; and EEG measures of brain activity.[177] There is less work in this vein on QALYs.[178] But

---

[172] *See id.* at 137; ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 381; Chris William Sanchirico, *Evidence, Procedure, and the Upside of Cognitive Error,* 57 STAN. L. REV. 291 (2004).

[173] Sugden, *supra* note 164, at 137.

[174] On the use of money or gifts as incentives in surveys, see, e.g., DILLMAN, *supra* note 158, at 14-15. Dillman suggests providing a "token of appreciation" – not full monetary compensation – as a step to establish the respondent's trust.

[175] *See, e.g.*, Venkatchalam, *supra* note 14, at 110-12; Carson, *supra* note 14, at 194-95; James J. Murphy et al., *A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation*, 30 ENVTL. & RESOURCE ECON. 313 (2005); 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. ENVT'L ECON & MGMT. 631 (2003).

[176] *See* Richard T. Carson et al., *Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods*, 72 LAND ECON. 80 (1996). A different body of scholarship focuses on comparing CV valuations for particular goods with the amounts that are actually paid in experiments. This work tends to find that the CV value is higher than the value actually paid. *See* Murphy, *supra* note 175, at 313-15. 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.

[177] *See* Blanchflower & Oswald, *supra* note 29, at 1360-61. For other discussions of the correlation between survey and non-survey evidence of individual happiness, see, e.g., Diener et al., *supra* note 28, at 278; Rafael Di Tella et al.,

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-trade-off formats correlate reasonably well.[179] All of this would be puzzling if QALY responses were just random.

A natural thought, given this evidence of the correlation between stated and true valuations, is 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 50% discount to CV values. Indeed, a substantial body of scholarship seeks to estimate such calibration factors.[180] A note of caution 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 first-best solution to problems of truthtelling would be to use norms and incentives to minimize deception; but if that seems unavailing, calibration factors offer a second-best approach to deriving information about underlying valuations from welfare polls.[181]

A more general point. Welfare polls are just a small bit of a vast survey literature, encompassing political opinion polls, censuses, psychological surveys, and consumer-product research. One estimate is that "[a]bout 20 million interviews are conducted each year in the United States."[182] 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.[183] Still, there remains great demand

---

The Macroeconomics of Happiness, 85 REV. ECON. & STAT. 809, 812 (2003); Ed Sandvik et al., *Subjective Well-Being: The Convergence and Stability of Self-Report and Non-Self-Report Measures*, 61 J. PERSONALITY 317 (1993).

[178] *See* Brazier et al., *supra* note 26, at 18.

[179] *See id.* at 30-46; Adler, *supra* note 22, at 41-42. There is analogous "internal" evidence for CV and happiness surveys which, like the behavioral evidence summarized in the text, suggests some correlation with respondents' true valuations. *See, e.g.,* Carson et al., *supra* note 14, at 193-95; Richard C. Bishop et al., *Contingent Valuation*, in HANDBOOK OF ENVIRONMENTAL ECONOMICS 626-46 (Daniel W. Bromley ed., 1995); Sandvik et al., *supra* note 177, at 319-20.

[180] *See, e.g.,* Murphy et al., *supra* note 175. NOAA once considered requiring that a 50% calibration factor be applied to CV values in calculating natural resource damages. *See* Navrud, *supra* note 40, at 12-13.

[181] This Section has focused on the main source of concern about truth-telling 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. *See, e.g.,* Colin Green & Sylvia Turnstall, *A Psychological Perspective*, in VALUING ENVIRONMENTAL PREFERENCES, *supra* note 14, at 206, 237-38. Here, as with strategic deception, the first-best response is to use survey design to mitigate the bias. *See id.* at 237-38. Calibration is second-best.

[182] FISHKIN, *supra* note 7, at 80.

[183] 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 164, at 3 (emphasis omitted). They argue that economic theory makes no prediction how respondents will answer inconsequential questions – in particular, theory does not predict truthful responses – and then go on to analyze when consequential surveys will be incentive compatible. *See id.* at 3-5.

for these surveys: by politicians, psychologists, and so on.[184] Consider, in particular, consumer product research — the closest parallel to welfare polls. Incentive-compatibility problems afflict these surveys: for example, the respondent 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 change.[185] But firms continue to conduct these surveys, at substantial cost; if the surveys had little informational value about consumers' preferences, that would be surprising.[186]

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.[187] Correlational data of this sort is now regularly used to derive calibration factors for consumer surveys[188] – a close analogue to the use of calibration factors for welfare polls.

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 . . . .[189]

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-trade-off variants of the QALY technique ask respondents to use an esoteric method which they may 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

[184] *See* Sugden, *supra* note 164, at 137-38; Di Tella et al., *supra* note 177, at 811-12.
[185] *See* Carson et al., *Incentive and Informational Properties*, *supra* note 164, at 10-20.
[186] *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 the main mechanisms for collecting primary data, along with the observation of behaviors).
[187] *See, e.g., id.* at 210; William J. Infosino, *Forecasting New Product Sales from Likelihood of Purchase Ratings*, 5 MARKETING SCI. 372 (1986).
[188] *See* CHURCHILL & IACOBUCCI, *supra* note 186, at 209-11; Diamond & Hausman, *Is Some Number Better than No Number?*, *supra* note 98 , at 54.
[189] ROGER TOURANGEAU ET AL., THE PSYCHOLOGY OF SURVEY RESPONSE 25 (2000).

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 in virtue of communicative norms may interpret it non-literally. 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 domain. Synonymous questions about happiness, if separated in a survey, will tend to receive the same answers; but if asked in succession are likely receive different answers, since respondents -- avoiding an interpretation that creates redundancy – will try to read them differently.[190]

The problem of misunderstood questions is, of course, a very general one for survey research, and a wide range of responses to the problem have 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.[191] 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.[192]

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 literature on second-order techniques,[193] and some of the more sophisticated approaches have percolated into welfare polls.[194]

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; what respondents' informational state is; whether they are confused about probabilities, loss averse, or otherwise irrational; and whether they are making a mental effort.

Representativeness

---

[190] *See* Schwarz & Strack, *supra* note 28, at 64.
[191] *See* STANLEY PRESSER ET AL., *Introduction*, in METHODS FOR TESTING AND EVALUATING SURVEY QUESTIONS 2 (2004).
[192] *Id.* at 4.
[193] *See, e.g.*, PAUL P. BIEMER & LARS E. LYBERG, INTRODUCTION TO SURVEY QUALITY 258-304 (2003); COGNITION AND SURVEY RESEARCH (Monroe Sirken et al. eds., 1999); ROBERT GROVES ET AL., SURVEY METHODOLOGY 241-53 (2004); METHODS FOR TESTING AND EVALUATING SURVEY QUESTIONNAIRES, *supra* note 191; GORDON B. WILLIS, COGNITIVE INTERVIEWING: A TOOL FOR IMPROVING QUESTIONNAIRE DESIGN (2005).
[194] *See, e.g.*, ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 151-56l; Champ, *supra* note 16, at 85-87.

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.[195] Indeed, the origins of polling in the United States are bound up with random sampling: George Gallup become famous because his random sample of a few thousand 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.[196]

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.[197] 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 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 those whose uses are especially sensitive to aesthetics, will tend to overstate the average CV 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.[198] 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, random sampling techniques are available – and indeed regularly employed by CV, QALY, and happiness researchers alike.

Deliberative Welfare Polls as a Solution?

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

[195] *See, e.g.*, BIEMER ET AL., *supra* note 193, at 305-50;GROVES ET AL., *supra* note 193, at 93-135.
[196] *See* FISHKIN, *supra* note 7, at 76-80.
[197] *See* BIEMER ET AL., *supra* note 193, at 30 (distinguishing between random sampling and different kinds of nonrandom sampling, such as convenience, purposive, and quota sampling).
[198] *Cf.* Paul Dolan, *Aggregating Health State Valuations*, 2 J. HEALTH SERV. RES. POL'Y 160, 160 (1997) (noting that QALY surveys might be aggregated using median rather than mean values).

about policy questions can improve on solitary policy polling.[199] Mightn't the same be true for welfare polls?

Consider that the scholars who favor deliberative-polling 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, wouldn't it also in the *welfare* context? A single presentation by an expert to the respondents assembled en masse is a cheap way to provide them 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 incorporate collective discussions about QALY, CV, happiness, or other welfare valuations, with ancillary informational, debiasing, and question-clarifications techniques administered to the group, culminating perhaps in a collective verdict, perhaps 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.[200]

A few of these group-deliberative CV studies have actually been conducted, although they remain very unusual.[201]

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."[202] But this is really a generic objection

---

[199] *See* sources cited *supra* note 7.

[200] Simon Niemeyer & Clive L. Spash, *Environmental Valuation Analysis, Public Deliberation, and Their Pragmatic Syntheses: A Critical Appraisal*, 19 ENVT. & PLANNING C: GOVT. & POL'Y 567, 576-77 (2001). For other scholarship on deliberative money valuation, see sources cited *id.* at 576; M. Sagoff, *Aggregation and Deliberation in Valuing Environmental Public Goods: A Look Beyond Contingent Pricing*, 24 ECOL. ECON. 213, 223-27 (1998).

[201] *See* 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. ENVT'L PLANNING & MGMT. 257, 259 (2005). This study is an example of deliberative money valuation; for another example, see Gregory, *supra* note 147. Focus groups do seem to be fairly common in CV research. *See* Michael D. Kaplowitz & John P. Hoehn, *Do Focus Groups and Individual Valuations Reveal the Same Information for Natural Resource Valuation?*, 36 ECOL. ECON. 237, 237 (2001). But these are group-based techniques for designing the questionnaire, not for collecting the data itself. *See also* Alan Shiell, *Reliability of Health Utility Measures and a Test of Values Clarification*, 67 SOCIAL SCI. MED. 56 (2003) (employing a deliberative, although not group-based, approach to eliciting QALYs).

[202] Niemeyer & Spash, *supra* note 200, at 579

to all welfare polls, both traditional solo polls and group-based formats such as DMV; it hardly shows why welfare polls, if they are justifiably conducted, are best conducted without inter-respondent deliberation.

A different objection, more to the point, 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."[203] These objections may be apt – but what is harder to see is why the objections 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. Part II described the various roles that welfare polls currently play, or might plausibly play, in administrative governance. 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. Synthesizing the material from the preceding two parts, I present a moderate defense of welfare polling and entertain general objections, from two quarters -- first the "revealed preference" tradition in economics which is generally skeptical of surveys, and second the deliberative-democratic tradition in political theory which is skeptical of questions about preference, interest, or welfare rather than the public good.

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."[204] Weak welfarism says that overall well-being is one of the moral considerations that bear 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 $M \geq 0$. 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

[203] *See id.* at 578-79; Cass R. Sunstein, *Group Judgments: Statistical Means, Deliberation, and Information Markets*, 80 N.Y.U. L. REV. 962 (2005).
[204] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___; *Rethinking Cost-Benefit Analysis*, *supra* note 36, at 204-16, 243-45; Adler, *supra* note 96, at 288-319.

economists,[205] and defended by Louis Kaplow and Steven Shavell in a recent, high-profile book,[206] allows for distributive considerations but insists that only information about well-being is relevant to moral evaluation. According to Kaplow and Shavell, morality has the structure $\{W_1 .. W_n\}$, 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 $F_1 ... F_M$, 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 non-welfarist moral factors.[207] 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 doesn't speak directly to the questions of *legal* obligation and authority that primarily concern governmental officials and legal scholars. Law and morality can come apart. Only a pure natural-law view – a 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 governmental officials are morally required to be sensitive to well-being. It doesn't 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.[208] 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.[209] 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 onto features of the world more readily ascertainable than welfare impacts[210] – *all* statutes have some degree of open-texture, some area where agencies have legal discretion.[211] 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 25 years, have imposed a legal obligation on executive agencies – flowing from the President's legal powers to oversee executive agency decisionmaking -- to consider overall welfare where that is statutorily permissible.[212] 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

---

[205] *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).

[206] *See* LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE (2002).

[207] Moral rights prohibit certain kinds of infringements (for example, intentional physical harms), and are nonwelfarist in that the degree of prohibition isn't 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.

[208] *See* Cass R. Sunstein, *Cost-Benefit Default Principles*, 99 MICH. L. REV. 1651, 1666-67 (2001).

[209] *See supra* text accompanying notes 49-51.

[210] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ___.

[211] *See* H.L.A. HART, THE CONCEPT OF LAW 124-36 (2d ed. 1994).

[212] *See supra* text accompanying note 36.

form that requires agencies to focus on some aspect of welfare. Or it can pass (and indeed has considered passing) a welfare super-mandate that would give statutory teeth to the general legal obligation to consider overall welfare now embodied in the Presidential cost-benefit orders.[213]

In short, questions about human well-being have substantial relevance, 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.[214] 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 that lie beyond the scope of this Article -- about the item whose fair distribution morality requires, and about the optimal institutional structure for redistribution.[215] Still, it is as 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.[216]

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 governmental 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 substantive rule, can and should describe welfare impacts. And the periodic statistics about the polity which governmental offices announce might include data about welfare, for example in the form of national well-being accounts.[217]

Nothing said to this point hinges on the specific content of administrative regulations.

---

[213] *See* Fred Anderson et al., *Regulatory Improvement Legislation: Risk Assessment, Cost-Benefit Analysis, and Judicial Review*, 11 DUKE ENVT'L L. & POL'Y FORUM 89, 89-107 (2000).
[214] *See supra* text accompanying notes 36-61.
[215] *See, e.g.*, Richard T. 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).
[216] *See supra* text accompanying notes 55-57.
[217] *See supra* text accompanying notes 83-95.

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. The clearest example 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 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 cost of the activity. In both of these contexts, CV, happiness, or QALY surveys can be useful in ascertaining what the welfare costs of private activities are.[218]

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– a point I have reiterated in preceding parts –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 a general schedule of valuations for different types of impacts, or a general "benefits function."

The Revealed Preference Objection to Welfare Polling

The argument thus far, synthesizing the material in Part II, shows that governmental officials are (morally and legally) justified in securing information about well-being. It doesn't yet establish that 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 behaviors, more precisely their non-verbal 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 marketed goods.[219] 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 of

[218] *See supra* text accompanying notes 62-74.

[219] On these techniques, see RICHARD E. JUST ET AL., THE WELFARE ECONOMICS OF PUBLIC POLICY: A PRACTICAL APPROACH TO PROJECT AND POLICY EVALUATION (2004).

course 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).[220]

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 questions marks an important divide between economists and other social scientists.[221]

This generic suspicion of subjective surveys has fueled much of the opposition to the CV method.[222] 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.[223]

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

---

[220] *See* FREEMAN, *supra* note 14, at 95-136, 353-452; George R. Parsons, *The Travel Cost Method*, in CHAMP ET AL., *supra* note 14, at 269; Laura O. Taylor, *The Hedonic Method*, in CHAMP ET AL., *supra* note 14, at 331; Mark Dickie, *Defensive Behavior and Damage-Cost Methods*, in CHAMP ET AL., *supra* note 14, at 395; Nancy E. Bockstael, *Travel Cost Methods*, in HANDBOOK OF ENVIRONMENTAL ECONOMICS, *supra* note 179, at 655; A. Myrick Freeman III, *Hedonic Pricing Methods*, in HANDBOOK OF ENVIRONMENTAL ECONOMICS, *supra* note 179, at 672.

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

[222] *See* Carson et al., *supra* note 14, at 176.

[223] OMB Circular A-4 (Sept. 17, 2003), at 24.

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.

The noncomparative objection to welfare polls is a bit of a straw man – although some of the critical literature comes pretty to making this strong claim.[224] 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 no 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 wouldn't be altered by welfare surveys, but the head of EPA's office of policy analysis isn't 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 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.[225]

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.[226]

---

[224] *See, e.g.*, Diamond & Hausman, *Is Some Number Better than No Number?*, *supra* note 98, at 46.

[225] 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 22, at 54-56.

[226] Carson et al., *supra* note 14, at 176. For similar defenses of CV surveys, see, e.g., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES, *supra* note 14, at 21-22; FREEMAN, *supra* note 14, at 154; Arrow et al., *supra* note 70, at 2-3, 41-42; Nichols Flores, *Conceptual Framework for Nonmarket Valuation*, in CHAMP ET AL., *supra* note 14, at

The problem with this traditional defense of welfare polls is twofold. First, it doesn't 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 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. But I believe such an argument can be furnished, as follows.[227]

To begin, revealed preference studies raise special problems of measurement. The ultimate objective of any valuation study is to estimate a numerical measure of the strength of some subject's subjective preference, on some numerical scale (QALY, dollar, happiness, or other), for some good. Welfare polls inquire directly about the strength of preference, and intervene to shape and correct 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.

One large measurement problem involves the gap between preferences (an 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 the good or that P is the average WTP. Rather, P is the WTP of the "marginal" consumer, the one who values the good the least.

A slightly more complicated example, again from the heartland of revealed preference work -- market behavior. Imagine that the price of the 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

---

47-50. *See also* Boyle & Bergstrom., *supra* note 105, at 191 (noting "conventional wisdom that CV is the only game in town when it comes to measuring non-use values").

[227] *See also* CHURCHILL & IACOBUCCI, *supra* note 186, at 212-30 (discussing pros and cons of surveys versus behavioral data in marketing research, with a particular focus on versatility as the main informational advantage of surveys).

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.[228]

The problem of estimating the subject's beliefs about the good being valued is a different, and equally pervasive, 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* ... C_n*, or a change in some good from characteristics C_1* ... C_n* to C_1+ ... C_n+, is valued by some group of subjects. That information will feed into policy analysis, government communications, and the design of fines and fees, as discussed above and documented in detail in Part II. 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, doesn't 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), 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 people's 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?"[229] Ditto for 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.[230]

Or consider the use of hedonic wage models or defensive behavior models to estimate WTP to avoid risk, a very standard use. A critical problem here is that the objective risk associated with a job or a defensive behavior, and the risk perceived by the worker or actor, may be different. "Valuation methods [for risk] based on revealed preference have the virtue of relying on actual behavior but can be applied only when the analysts knows (or can reasonably infer) what decision alternatives and consequences (including their pecuniary, health, and other attributes) were perceived by the decision maker."[231]

A third, general, 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* ... C_n* to C_1++, C_2* ... Cn* If existing goods

[228] A good review of this literature, and the basic theoretical flaws of ordinary consumer surplus, is FREEMAN, *supra* note 14, at 49-72.
[229] *Id.* at 363.
[230] *Id.* at 428.
[231] Hammitt & Graham, *supra* note 131, at 34.

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) .... [And they] may suffer from collinearity among attributes. Collinearity 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.[232]

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."[233]

A final measurement problem is this. "Revealed preference" methods standardly use the observed 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. 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 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). Or the good may be marketed, but the consumer may be insured – so that its price will exceed his cost. In particular, inferring WTP for health from the prices of health care services or products is very tricky, given the wide existence of health insurance.

Let us turn, now, to the second set of reasons why welfare polls have informational value even with revealed preference studies in the picture. The first set of reasons, just discussed, have to do with measurement. The second have to do with 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 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.

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 premia for riskier jobs are problematic, in part,

---

[232] 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 collinearity is a major one for defensive-behavior methods, often referred to there as the problem of "joint production." *See* 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); Dickie et al., *supra* note 220, at 412-13; Maureen L. Cropper, *Has Economic Research Answered the Needs of Environmental Policy?*, 39 J. ENVTL. ECON. & MGMT. 328, 335 (2000).

[233] CHURCHILL & IACOBUCCI, *supra* note 186, at 209.

because workers may be incurring risk out a sense of responsibility to family.[234] Ditto 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).[235]

Admittedly, as discussed in Part III, welfare pollsters do not seem to have experimented much with discursive devices for screening out disinterested preferences.[236] 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 here 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[237] — a problem already touched upon, in my discussion of the gap between the true and perceived characteristics of goods. 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 problems of sample size or statistical bias. Here, current welfare polling techniques, as described in Part III, do offer an important advantage. 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.[238]

A parallel analysis applies to the problem of preference distortion. To begin, it is clear that preferences measured in behavioral 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[239] 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 which seems descriptively much more accurate than expected utility theory.[240] "[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).[241] John Payne and co-authors have

---

[234] *See* ELIZABETH ANDERSON, VALUE IN ETHICS AND ECONOMICS 197-99 (1993).

[235] *Cf.* James R. Bettman et al., *Constructive Consumer Choice Processes*, 25 J. CONSUMER RES. 187, 197-199 (1998) (suggesting that consumers often must justify their choices to others, and use choice procedures that make justification easier).

[236] *See supra* text accompanying notes 105-106.

[237] *See* OMB Circular A-4 (Sept. 17, 2003), at 20; Bettman et al., *supra* note 235, at 201-02.

[238] *See supra* text accompanying notes 118-129, 199-203.

[239] *See, e.g.*, PLOUS, *supra* note 130, at 84-188; HASTIE & DAWES, *supra* note 133, at 289-312; Baron, *supra* note 102, at 82-84.

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

[241] 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).

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.[242] 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 trade-off rates for dimensions (such as health or life) that are particularly important.[243]

The other distortions that have raised concerns about welfare polls — framing, anchoring, availability, range effects — have also been observed in consumer behavior. [244] More generally, 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 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; more successful may be changing elicitation format, or changing the kind of scale entirely, for example switching from CV to QALY surveys to overcome tradeoff biases in valuing health states, or switching from standard happiness studies to Kahneman's moment-based surveys to overcome availability biases. And switching from solo to group surveys may help to facilitate debiasing, just as it may help to disseminate information cheaply.[245]

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.[246] 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; survey respondents will not. On the other hand, as discussed in Part III, the polling format can presumably intensify the internal and social pressures which induce additional effort – by remunerating respondents for their answers, by using face-to-face interviews rather than mail or internet surveys, and (once

---

[242] *See* Bettman et al., *supra* note 235, at 190. For a discussion of the separability 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).

[243] *See* Bettman et al., *supra* note 235, 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 241.

[244] *See* Bettman et al., *supra* note 235, at 201-02 (availability); *id.* at 208 (framing); Joseph C. Nunes & Peter Boatwright, *Incidental Prices and their Effect on Willingness to Pay*, 42 J. MKT'G RES. 457 (2004) (anchoring); Ronald Niedrich et al., *Reference Price and Price Perceptions: A Comparison of Alternative Models*, 28 J. CONSUMER RES. 339 (2001) (range effects).

[245] *See supra* text accompanying notes 143-153, 199-203.

[246] *See* Bettman et al., *supra* note 235, at 189-99; JOHN W. PAYNE ET AL., THE ADAPTIVE DECISIONMAKER 9-15 (1993).

more) by shifting from solo to group formats.[247]

To sum up: 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, mental effort).[248]

My response to the revealed-preference objection assumes, to be sure, that the communicative difficulties which are unique to 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 cancel their informational value. *If* the respondents to these polls assigned semantic content to the questions asked which didn't correlate with their objective semantic content, or if the respondents understood the questions but provided untruthful answers that didn't 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 (politicians, firms), wouldn't use them if respondent lying, shirking, and misunderstanding seriously undercut their informational value.

Why doesn't this happen? Shirking has been 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.[249]

As for the problem of 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 truth-telling 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, again, be designed to intensify this pressure. The fact that stated valuations correlate with behaviorally inferred valuations underscores the informational value of surveys. If respondents were not constrained (by norms, internal pressure,

---

[247] *See supra* text accompanying notes 162-163, 199-203.

[248] Two other generic problems that affect revealed preference techniques should also be noted. The first are problems of bias and self-selection, for example the fact that workers with a greater appetite for risk would tend to take risky jobs, which means that the observed wage premium for a given risk will tend to be lower than population WTP for that risk. *See* U.S. EPA, OFFICE OF RESEARCH AND DEVELOPMENT, *Human Health Metrics for Environmental Decision Support Tools: Lessons from Health Economics and Decision Analysis* 18-19 (Sept. 2001). *See also* OMB Circular A-4 (Sept. 17, 2003), at 20 (noting that, in revealed preference studies, "the specific market participants studied should be representative of the target populations to be affected by the rulemaking under consideration"). The second involve assumptions about market structure, for example the assumption in property value models that switching costs are low. *See supra* text accompanying notes 48; OMB Circular A-4 (Sept. 17, 2003), at 20 (stating that, in revealed preference study, market should be competitive).

[249] *See supra* text accompanying notes 189-194.

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.[250]

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 have been traditional economists who agree that policymakers need information about preferences, but are suspicious about surveys. A different set of critics have 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. The argument is 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. 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 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.[251]

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

[250] *See supra* text accompanying notes 164-188.

[251] MARK SAGOFF, THE ECONOMY OF THE EARTH: PHILOSOPHY, LAW, AND THE ENVIRONMENT 100 (1988). This book draws on a number of Sagoff's prior articles. *See id.* at x. For more recent statements, drawing a similar distinction between consumer preferences and citizen preferences or judgments, see Sagoff, *supra* note 200, at 213-14; MARK SAGOFF, PRICE, PRINCIPLE, AND THE ENVIRONMENT 2-3 (2004).

> 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. [252]

These quotations exemplify a standard line of critical scholarship about CV studies,[253] but it should be noted that the criticisms would also seemingly apply to welfare polls that seek to elicit non-monetary measures of respondents' self-interested preferences, such as QALY or happiness surveys. Indeed, Heinzerling and Ackerman explicitly criticize QALY measures.[254]

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 which this survey approach instantiates. By deliberation, Fishkin means public-regarding deliberation, or at least deliberation which is open to claims about the public good rather than individual interest. "Deliberation" approximates Habermas' 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."[255] 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."[256] Ned Crosby, who has spearheaded the development of "citizen juries" in the United States, argues that these embody a "social contract" approach to policy choice, as opposed to a "utility calculation" or a "political power" approach.[257] This is (roughly) 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 CJs [citizen juries] as a method of preference revelation allows ... deliberation on the environmental issue in terms of what is best for society.[258]

---

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

[253] *See, e.g.*, ANDERSON, *supra* note 234, 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 (Robert W. Hahn ed., 1996); Laurence H. Tribe, *Policy Science: Analysis or Ideology?* 2 PHIL. & PUB. AFF. 66, 95-97 (1972).

[254] *See* ACKERMAN & HEINZERLING, *supra* note 252, at 98-102.

[255] FISHKIN, *supra* note 7, at 40.

[256] 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; Kenyon et al., *supra* note 7, at 558.

[257] *See* Crosby, *Using the Citizens Jury Process*, *supra* note 7, at 401-02.

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

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, not welfare polls, have the appropriate public-regarding structure for citizen involvement in policymaking.

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 appropriate 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).

There are a range of possible arguments, intrinsicalist and instrumentalist, for why governmental decisions 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.[259] Some of these arguments are quite plausible, and I will assume here that they succeed in demonstrating that policy-deliberation formats should be convened to address some subset of governmental decisions. I say "some subset" because, clearly, policy deliberation formats can't be used to address all government 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.[260]

The subset of governmental decisions appropriately informed by "deliberative polls," "citizen juries," or other policy-deliberation formats might be second-order or first-order decisions.[261] Second-order decisions concern the decision-procedures that agencies should use in issuing regulations, building projects, and making other first-order choices. 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, rather, some competitor procedure in choosing regulatory measures. Congress also faced a second-order choice when it debated the enactment of a generic cost-benefit "supermandate." 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.[262] For example, a citizens' jury might be convened on the question whether an agency 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. More generally, second-order decisions can eventuate in administrative decision procedures that require the agency to attend to human welfare, or aspects of welfare. The agency, in line with the citizens' recommendation, will implement the welfare-regarding decision procedure, and in doing so may

---

[259] *See* Thomas Christiano, *The Significance of Public Deliberation*, in DELIBERATIVE DEMOCRACY, *supra* note 6, at 243, 244-46 (describing the different kinds of instrumental and intrinsic value that public deliberation might have).

[260] The proponents of "citizen juries" and other policy-deliberation formats do not suggest that they would be structured to comply with the panoply of constitutional and statutory rights that are afforded defendants tried before traditional criminal juries – for example, to limit the sorts of evidence that the deliberating citizens could consider.

[261] This should not be confused with my earlier distinction between second- and first-order techniques for designing surveys.

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

well employ welfare polls.

First-order policy deliberation formats and welfare polls, too, are compatible – although this is a little less obvious. Parenthetically, the case for first-order formats is debatable. As Crosby notes, "the [citizens' jury] process works better on value questions" – about the evaluative criteria that should guide some subsequent choice – "than technical issues."[263] But let me bracket the point and assume that a policy-deliberation format has been convened to determine (say) the content of an agency regulation. For example, an environmental agency drafting a pollution-control regulation for pollutant X might ask a citizens' 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 virtue 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.[264] 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) increases overall welfare and (if the effects of pollution are disproportionately borne by the poor) diminishes 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.

Here's one way to put the lesson that, I hope, emerges from the last few paragraphs. 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. Administrative decisionmaking is a kind of legal reasoning that should begin with applicable legal sources (statutes, the Constitution, executive orders); should integrate moral considerations where legally required or permissible; and should invite citizen participation where legally required or where legally permissible and morally required. At some junctures in this complicated reasoning process, the concept of "welfare" or of aspects of welfare ("happiness," "health") may become relevant. If weak welfarism, the moral view I defended earlier in this Part, is correct, those junctures should be fairly frequent. The entry of "welfare" or its aspects into the legal-cum-moral reasoning process may then lead the reasoners to wonder how much some outcome or policy affects welfare – and thence to welfare polls.

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; but from that starting point 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

[263] Crosby, *Citizens Juries*, *supra* note 7, at 157.

[264] *See, e.g.,* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at __ (suggesting that moral rights and distributive considerations – perhaps cashed in terms of the distribution of well-being – as well as overall welfare may have moral relevance); *Rethinking Cost-Benefit Analysis*, *supra* note 36, at 244-45 (same).

policies on our narrow interests. This is no more paradoxical than the observation that preference-utilitarianism or preference-egalitarianism are genuinely moral views even though they construe the content of morality in terms of the distribution or the maximization of preference-satisfaction. 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 care about.

A final claim that deliberative democrats might make to criticize welfare polls is the following: "There are intrinsicalist moral reasons to encourage citizen participation in government – reasons linked to the process of participation itself – and welfare polls do not constitute the sort of participation that is intrinsically valuable." Is this claim true? To decide, we'd need a clearer conception of the way in which participation is intrinsically valuable. In particular: Does the realization of the intrinsic value truly require participants to adopt a moral perspective, or might other perspectives (for example, a self-interested perspective that participants provisionally adopt so as to provide information about the nature of well-being) also be acceptable? But we need not pursue these questions. Even if welfare polling fails the test of intrinsic value, by contrast with other citizen-participation mechanisms, the welfare-polling format still has *instrumental* value. It still provides information useful in producing substantively better governmental decisions – for all the reasons provided in Parts II and III above. The intrinsicalist may have good reason to encourage deliberative polling or other policy-deliberation formats; but because welfare polls are complementary with such formats, not preclusive of them, the intrinsicalist claim is not in fact an objection to welfare polls.

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 current play in agency decisionmaking and other aspects of administrative governance[265]; canvassing other possible roles;[266] describing a variety of possible improvements to the survey enterprise within this trio of formats;[267] and defending welfare polling from objections from two quarters, the "revealed preference" camp within economics and the "deliberative democracy" camp within political theory.[268]

But should the trio of CV, QALY, and happiness surveys be changed – perhaps supplemented by different welfare polling formats, 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.[269] QALYs became popular, in substantial part, because public health researchers were disinclined to monetize

[265] *See supra* Part II.
[266] *See id.*
[267] *See supra* Part III.
[268] *See supra* Part IV.
[269] For critical discussion of Kaldor-Hicks efficiency, with citations to other critical literature, see, e.g., Adler, *supra* note 96, 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* sources cited *supra* note 36. On the continuing dominance of the Kaldor-Hicks criterion in applied economics, see, e.g., RICHARD E. JUST ET AL., THE WELFARE ECONOMICS OF PUBLIC POLICY 646 (2004).

health and longevity and wanted a scale that was different from the WTP/WTA scale.[270] 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.[271] Only recently have economists and others with an interest in policymaking latched on to happiness surveys.

Despite their "accidental" origins, CVs, QALYs, and happiness surveys turn out to be a vital governmental tool, as discussed. QALYs 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 multi-dimensional 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.

There is a long philosophical tradition, going back to Aristotle, of drawing up lists of the different aspects of well-being.[272] 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, the use of the "senses, imagination, and thought," the emotions, practical reason, affiliation, interaction with other species, play, and control over one's environment.[273] Finnis' is: life, knowledge, play, aesthetic experience, sociability, practical reasonableness, religion. [274] Griffin's is: accomplishment, autonomy, understanding, enjoyment, and deep personal relations.[275] 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 traditionally conceived.[276] Research groups in 15 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."[277] Based on these focus groups, a preliminary questionnaire was

---

[270] *See* Adler, *supra* note 22, at 14.
[271] *See, e.g.*, Ryan & Deci, *supra* note 28, at 142.
[272] *See, e.g.*, L.W. SUMNER, WELFARE, HAPPINESS AND ETHICS 45-80 (1996); SABINA ALKIRE, VALUING FREEDOMS 77-84 (2002). 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* note 97.
[273] MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH 78-80 (2000)
[274] JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 85-90 (1980).
[275] JAMES GRIFFIN, VALUE JUDGMENT: IMPROVING OUR ETHICAL BELIEFS 29-30 (1996)
[276] *See* Amy 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); The WHOQOL Group, *The World Health Organization Quality of Life Assessment (WHOQOL): Development and General Psychometric Properties*, 46 SOC. SCI. MED. 1569 (1998).
[277] The WHOQOL Group, *supra* note 276, at 1570.

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 15 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 questionnaire. The final WHOQOL structure consists of 24 facets or dimensions of well-being, grouped into 6 domains.

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 such as Nussbaum's, Griffin's, and Finnis's, 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 in survey work designed to characterize the fine structure of particular well-being dimensions. Consider the WHOQOL domains. 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.[278] Happiness surveys, if focused on the respondents' 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.[279] 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.[280] In theory, both of these benefits could also be realized through an inclusive nonmonetary measure – an idea I'll return to in a moment – but dimension-specific measures such as QALYs have certain advantages. A QALY survey asks the respondent to imagine changes in her health state, holding non-health characteristics fixed. To put this in terms of the WHOQOL, imagine a QALY survey that focuses on dimensions 1, 2, 3, 5, 8, 9, 10, 11, and 12, which would be reasonably typical. These are the "health" facets; the other WHOQOL facets are background facets. Then the QALY survey will instruct the respondent that the number 1 means the best possible state *with respect to all the health dimensions*, together with some set of background characteristics (most straightforwardly, the respondent's actual characteristics), and 0 means the worst possible state *with respect to all the health dimensions*, together with that same set of background characteristics.[281] The respondent will then be asked 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 1 is an ideal state *with respect to all the WHOQOL dimensions*, and that 0 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

[278] *See* Adler, *supra* note 22, at 47-52.

[279] *See supra* note 54 (describing standard conception of happiness as a 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 31. The respondent to a standard, global question about life-satisfaction (for example, that used in the Eurobarometer series, *see supra* text accompanying note 30) could reasonably answer by thinking about his achievements with respect to the totality of WHOQOL dimensions, but the question does not ask him to do that, or otherwise draw his attention to those dimensions. Further, an answer to a question such as "How satisfied are you with your life?" isn't 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. Contrast the QALY method, which specifies a state for the respondent to evaluate.

[280] *See supra* text accompanying notes 42-43,152.

[281] 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 heath characteristics, rather than both health *and* background characteristics, and is therefore presumably more sensitive to health changes, remains true. *See* Adler, *supra* note 22, at 47-52.

collection of QALY scores gives angina a score of .75 and pancreatitis a score of .80).[282] 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 QALYs and happiness surveys) as opposed to the last three. Social quality-of-life metrics and surveys, corresponding to the fourth WHOQOL domain, should be experimented with. These 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. Ditto 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.[283] 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.

The point to see is that 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 its well-being value on some quantitative scale. Questions of this sort would seem to have both advantages and disadvantages vis a 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.[284]

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.[285] Whether, on

[282] *See* Harvard Center for Risk Analysis, Cost-Effectiveness Analysis Registry, http://www.hsph.harvard.edu/cearegistry.
[283] *See supra* text accompanying notes 60-61 (discussing Kahneman's approach to measuring experience).
[284] *See supra* text accompanying notes 199-203 (discussing possibility of group-based welfare polls).
[285] *See supra* text accompanying note 109.

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 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).[286] Imagine that the interpersonal utility scale ranges from 0 to 1. Then, if person A is at level .3 and person B is at level .4, this means that B is better off than A. If a policy changes A's level to .31 and B's to .37, this means that the policy decreases overall well-being (because the .01-unit positive difference it makes to A's well-being is less than the .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.[287] 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 format analogous to the QALY standard-gamble. 1 is the best possible life; 0 the worst possible life; the respondent is then asked to contemplate a possible life and to place it on the 0-1 scale by expressing the probability p that makes her indifferent as between that life and a p probability of the best possible one.[288] Different respondents might express different indifference probabilities for a given life – which raises a problem of aggregation. But that problem is no different from the aggregation problem currently faced by QALY surveys.[289]

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 fixed non-health dimensions; a parallel but more inclusive format that defined 1 as the optimal state with respect to all 24 of the WHOQOL dimensions, and 0 as the worst state with respect to all 24, 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;[290] a third, to change format. What the change of format would be is itself a matter for experimentation. But just as the time-tradeoff method has emerged, within QALY

---

[286] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ch. 2 (discussing how different moral theories require measurement of well-being levels and differences).
[287] *See id.*; Adler, *supra* note 22, at 17-24; Adler, *Beyond Efficiency and Procedure*, *supra* note 96, at 289-302; Adler & Posner, *Rethinking Cost-Benefit Analysis*, *supra* note 36, at 204-09.
[288] *Cf.* Andrews & Robinson, *supra* note 31, at 66, 74 (describing Cantril's Ladder Scale, which asked respondents to locate themselves on a ladder with the top representing "the best possible life for you" and the bottom "the worst possible life for you").
[289] *See* Adler, *supra* note 22, at 41; Paul Dolan, *Aggregating Health State Valuations*, 2 J. HEALTH SERV. RES. POL'Y 160 (1997).
[290] *See supra* text accompanying notes 143-144, 153 (discussing the use of these devices within QALY, happiness, or CV formats).

research, as a better technique for eliciting QALY values than the QALY standard gamble – once thought to be the gold standard [291]– so the standard-gamble format 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 doesn't preclude other utility-elicitation formats that are easier to use and produce values which approximate standard-gamble values.[292]

Why an interpersonal utility scale? We already possess, in dollars, an inclusive scale – one that covers all 24 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.[293] If P is willing to pay $100 for a policy, and Q is willing to accept $50 in exchange for the policy, it doesn't 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 imperfect inaccuracy of WTP/WTA amounts in tracking well-being, cost-benefit analysis is probably the best currently available welfarist policy-analytic tool: 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.[294] 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 distributive weights, or by guiding the choice of QALY-to-dollar or happiness-to-dollar conversion factors.[295] 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 sum up: the trio of QALYs, CVs, and happiness surveys can be usefully supplemented (1) by surveys such as the WHOQOL that attempt to characterize the multi-dimensional structure of well-being; (2) by dimension-specific analogues to QALYs 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) by surveys that ask respondents to evaluate others' lives, not their own; and (4) by 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

---

[291] *See supra* note 151 and accompanying text.

[292] 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 112, at 257-57 (discussing multi-attribute utility theory techniques, such as eliciting "swing weights" for different dimensions).

[293] Similarly, money measures are only a rough index of welfare levels. Take wealth as the most obvious monetary proxy for welfare levels. P may have more wealth than Q but be worse off, given his physical condition or lack of access to public goods.

[294] *See* ADLER & POSNER, NEW FOUNDATIONS, *supra* note 36, at ch. 3; Adler, *supra* note 22.

[295] *See* Adler, *supra* note 22, at 57-74 (discussing use of interpersonal utility numbers in setting QALY-to-dollar conversion factor).

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, 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.

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 pretty much 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 the condition of survey respondents: on whether they 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 motivational, 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.