NBER WORKING PAPER SERIES

WHAT IS THE GENERAL WELFARE? WELFARE ECONOMIC PERSPECTIVES

Charles F. Manski

Working Paper 33376
http://www.nber.org/papers/w33376

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2025

The author received no external funding to support this research and has no financial relationships requiring disclosure. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Charles F. Manski. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

What is the general Welfare? Welfare Economic Perspectives
Charles F. Manski
NBER Working Paper No. 33376
January 2025
JEL No. D60, H0

ABSTRACT

Researchers cannot definitively interpret what the framers of the United States Constitution had in mind when they wrote of the general Welfare. Nevertheless, welfare economics can contribute to policy choice in democracies. Specifying social welfare functions enables coherent analysis, by formalizing mechanisms for preference aggregation and studying the policies they yield. This paper argues that it is essential for welfare economics to adequately express the richness and variety of actual human preferences over social states. I first discuss devices that economists have used in attempts to circumvent or grossly simplify specification of social welfare functions. I next discuss the common welfare economic practice of assuming that personal preferences are homogeneous, consequentialist, and self-centered. I then call for incorporation of broader forms of personal preferences into social welfare functions. Individuals may hold heterogeneous social preferences, being concerned in various ways with the distribution of outcomes in the population. They may hold heterogeneous deontological preferences, placing value on their own actions and the actions of others. They may have preferences for the mechanism used to aggregate preferences in a social welfare function. These potential aspects of personal preference should be recognized in welfare economics.

Charles F. Manski
Department of Economics
Northwestern University
2211 Campus Drive
Evanston, IL 60208-2600
and NBER
cfmanski@northwestern.edu

1. Introduction

A foundational objective of the Constitution of the United States is to promote the *general Welfare*. The Constitution uses the term twice. The Preamble states:

> "**We the People** of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defence, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America."

Article I, Section 8, Clause 1, often called the "General Welfare Clause," begins as follows: "The Congress shall have Power To lay and collect Taxes, Duties, Imposts and Excises, to pay the Debts and provide for the common Defence and general Welfare of the United States;"

The Constitutional premise that the United States should promote and provide for the *general Welfare* has rhetorical appeal, but it lacks substance. To make it meaningful requires a clear definition of the term. The Constitution contains no definition, nor have Acts of Congress and Supreme Court decisions clarified what constitutes the *general Welfare*. One might attempt to infer the intended meaning from parts of the Constitution such as the Bill of Rights, and from legislation that Congress has enacted and the Supreme Court has deemed constitutional. Inference might be straightforward if Americans were to hold a consensus perspective on what the nation should seek to achieve. It is evident, however, that the preferences of citizens have been heterogeneous to a considerable degree.

The vagueness of the Constitutional term *general Welfare* contrasts with the specificity of economic study of public policy. Economists assume a particular social welfare function or a well-defined class of welfare functions. Research in public economics seeks to characterize the social welfare achieved by alternative feasible policies, aiming to find one that maximizes welfare. Cost-benefit analyses commissioned by government agencies are used to inform policy design and evaluation.

Economists studying policy choice in democracies strive to specify social welfare functions that express the values of society rather than the preferences of dictators; thus, the welfare function should adequately formalize the *general Welfare*. Paul Samuelson placed responsibility for specification of the

welfare function on society rather than on the economist, writing (Samuelson, 1947, p. 220): "It is a legitimate exercise of economic analysis to examine the consequences of various value judgments, whether or not they are shared by the theorist." However, the vagueness of the term *general Welfare* and related terms has placed economists in the awkward position of having to hypothesize welfare functions rather than draw them from society.

In my own research, I have studied policy choice with *pragmatic* welfare functions. I have characterized pragmatic welfare functions as ones motivated by (Manski, 2024, p. 58): "some combination of conjecture regarding societal values, empirical study of population preferences, and concern for analytical tractability." I have found the notion of a pragmatic welfare function useful. I nevertheless worry that economists commonly hypothesize welfare functions that are remote from the actual heterogeneous preferences of society. I have cautioned that economists sometimes use a veneer of pragmatism to conflate science with advocacy (Manski, 2011). That is, they sometimes intentionally specify welfare functions that justify policy conclusions they favor.

In this paper I review and critique the various ways that economists have sought to cope with the vexing problem of specifying social welfare when studying public policy. I call attention to arguably important aspects of the *general Welfare* that economists have generally ignored. My objective is to clarify basic issues. It is not to prescribe use of a particular welfare function.

Economists have mainly studied *personalist* (aka *welfarist*) welfare functions that aggregate the personal welfare (aka *utility*) of individuals, rather than ones that view a society holistically.[1] I maintain this stipulation throughout. However, I find it crucial to observe that economists have commonly placed strong assumptions on personal preferences. Moreover, they have specified mechanisms to aggregate

[1] Sen (1977) used the term *welfarism*, writing (p. 1559): "The general approach of making no use of any information about the social states other than that of personal welfares generated in them may be called 'welfarism.' " He then wrote that "welfarism as an approach to social decisions is very restrictive." As I see it, the word *welfarism* does not express Sen's distinction well, because we regularly speak of social welfare and well as personal welfare. Hence, I instead use the word *personalism*, which seems to me more informative.

personal welfares, typically utilitarian addition, rather than having the aggregation mechanism emerge from personal preferences. Thus, welfare economics has been only partially personalist in practice.

Contrasting Welfare Economics and Moral Philosophy

Before proceeding, I think it important to contrast the usual mindsets of researchers in welfare economics and moral philosophy. Whereas welfare economics has mainly been at least partially personalist, a pervasively non-personalist perspective has been common in moral philosophy. For thousands of years, from the philosophical writings of the ancient Greeks through those of the Old and New Testaments through those of the Enlightenment to the present, philosophers have given themselves the authority and responsibility for interpreting the *general Welfare*. Thus, philosophers have made normative arguments that a social welfare function *should* have certain properties, regardless of whether the members of society agree.

Philosophers often assert that certain deontological principles---equity, fairness, justice---should supersede consequentialist ones. They often argue for lexicographic evaluation of policies, wherein society first restricts attention to policies deemed deontologically acceptable and only then considers the consequences of these policies. Lexicographic planning prohibits the nuanced quantitative weighing of welfare tradeoffs that economists usually recommend.[2]

Lexicographic planning can be coherent in principle, but I have often found it to be distressingly incoherent in philosophical discourse. Philosophers may privilege multiple broad deontological principles, often called *rights*. However, they commonly fail to ask how different rights should be reconciled when

[2] Hildreth (1953) is a relatively rare instance of an economist who explicitly considered placing lexicographic ethical constraints on the space of feasible policies. Using the letter S to denote the policy space, he wrote (p. 91):

> "It is possible that some ethical values which should be recognized in making social choices apply independently of individual preferences. In the present approach at least some of these values might be expressed as restrictions on the set of achievable alternatives and entered along with technical restrictions in the determination of S. Thus, if certain social states were judged to be ethically undesirable independently of individual preferences, these states could be regarded as not achievable and could be excluded before the ordering based on individual preferences was applied."

they conflict with one another. Coherent lexicographic planning requires specifying the sequence in which society should order conflicting rights. But philosophers rarely engage in this essential task.

The difficulty of coherent lexicographic planning is compounded because philosophers commonly write as if values are binary rather than measurable or comparable. Thus, a policy may be classified as "just or unjust" or as "equitable or inequitable," without conceptualization of degrees of justice and equity. See Hsieh and Andersson (2021) for a recent discussion of this longstanding problem in moral philosophy.

It is notable that philosophers do not agree on what properties of welfare functions warrant normative approval. To the contrary, heterogeneity in their perspectives abounds. It is also notable that, in contrast to economics, philosophical discourse rarely engages concrete policy choices. Philosophers appear most comfortable discussing abstract general principles, using idealized examples to illustrate.[3] This paper focuses on research in welfare economics rather than the normative scholarship of moral philosophers.

Organization of the Paper

Section 2 discusses devices that economists have used in attempts to circumvent or grossly simplify specification of explicit social welfare functions. These include the Hicks (1939) introduction of the *new welfare economics* and the proposal by Kaldor (1939) and Hicks (1939) that economists use the artifice of Kaldor-Hicks efficiency to separate the study of Pareto efficiency and the distribution of personal welfare. They include the longstanding macroeconomic practice of using *representative-agent* models to avoid consideration of heterogeneity in personal preferences. And they include the as-if thought experiments proposed by Harsanyi (1955) and Rawls (1970) to argue that individuals should have a consensus view on the mechanism used to aggregate personal preferences.

[3] Research in jurisprudence has aspects in common with moral philosophy and economics. Some jurisprudence discusses broad principles of justice, equity, and rights in abstraction, as in moral philosophy. Other legal writing aims to be pragmatic, recognizing that reasonable construction and interpretation of particular statutes may require attention to subtle welfare tradeoffs. Research of the second type sometimes restricts itself to qualitative recognition of tradeoffs, but it sometimes proposes relatively precise rules for the practice of criminal or civil justice. Zamir and Medina (2008, 2010) discuss the tension between the various perspectives in legal research.

Section 3 discusses the common welfare economic practice of assuming that personal preferences are consequentialist and self-centered. These assumptions are not fundamental to welfare economics. In particular, they are not required in utilitarian welfare analysis. Nevertheless, economists have usually presumed that individuals only contemplate their personal circumstances and only value the consequences of actions, not the actions themselves. To illustrate, I use the utilitarian optimal income-tax theory pioneered by Mirrlees (1971), which assumes that individuals have homogeneous preferences for (consumption, leisure) outcomes and differ only in their abilities to generate income. I also describe the use of the distribution of quality-adjusted life years to measure health-related welfare in a heterogeneous population. I observe that, in practice, research in personalist welfare economics has only partially been based on personal preferences. Researchers have themselves specified basic aspects of social welfare functions.

Section 4 considers incorporation of broader forms of personal preferences into social welfare functions. Individuals may hold consequentialist social preferences, being concerned not only with the outcomes they experience themselves but with the distribution of outcomes in the population. They may hold deontological preferences, placing value on their own actions and on the actions of others. They may have preferences for the mechanism used to aggregate preferences in a social welfare function. These potential aspects of personal preference should be recognized in welfare economics.

2. Personalist Welfare Economic Attempts to Circumvent Specification of Personal Preferences

2.1. The New Welfare Economics and Kaldor-Hicks Efficiency

In the Introduction I quoted Samuelson (1947), who counseled economists that it is legitimate to perform policy analysis using welfare functions that express "various value judgments, whether or not they are shared by the theorist." Even so, economists have often sought to avoid basing analysis on explicit value judgements, Samuelson's perspective notwithstanding. Much economic research has used various devices to circumvent specification of a welfare function.

One route was taken in the 1930s and 1940s by the economists who initiated study of the *new welfare economics* (Hicks, 1939). Not wanting to specify a particular welfare function, which would require taking a stand on interpersonal welfare tradeoffs, they retreated to the study of Pareto efficiency. Indeed, a sequence of analyses relating perfect competition to Pareto efficiency under certain strong assumptions eventually became labelled as the *Fundamental Theorems of Welfare Economics*. See Blaug (2007) for a historical review.

However, restriction of attention to the Pareto frontier severely limited the ability of economists to study actual planning problems. Hildreth (1953) wrote (p. 82): "The new welfare economics has been primarily criticized for the narrowness of the range of questions to which it provides answers." He added (p. 91): "if we wish to go beyond the comparisons that are possible using only the principle of new welfare economics, the issue is not whether we can do so without making interpersonal comparisons of satisfactions. It is rather, what sorts of interpersonal comparisons are we willing to make." Chipman and Moore (1978) wrote (p. 548): "we shall argue that, judged in relation to its basic objective of enabling economists to make welfare prescriptions without having to make value judgments and, in particular, interpersonal comparisons of utility, the New Welfare Economics must be considered a failure."

Recognizing that study of the Pareto frontier alone cannot suffice to inform policy choice, Kaldor (1939) and Hicks (1939) proposed that economists separate the study of Pareto efficiency and the distribution of personal welfare. Rather than consider the actual distribution of welfare achieved by a policy, they argued that it suffices to contemplate fictional redistributions that might in principle be achievable by transfers of wealth. It became common for economists to suggest the use of a hypothetical *lump-sum taxation* mechanism that would levy a fixed tax on each individual, one whose magnitude cannot be altered by changes in individual behavior. See the discussion in Atkinson and Stiglitz (1980).[4]

[4] Public economists have commonly viewed lump-sum taxation as an ideal approach to generation of tax revenue. It is often favorably called "non-distortionary," a term sometimes misinterpreted to mean that the tax does not affect individual behavior. However, a lump-sum tax necessarily affects some aspect of behavior, as it changes personal wealth. Atkinson and Stiglitz (1980) called attention to this misinterpretation when they wrote (p. 469): "Even a lump-sum tax would have allocative effects. Yet (by definition), it is non-distortionary." Thus, the use of the word "non-

Fictional redistribution has become a peripheral topic in economic theory, but it continues to be used widely in applied research on planning. Focusing on efficiency, macroeconomists often seek a policy that maximizes the presented discounted value of Gross Domestic Product (GDP), without analysis of the distribution of GDP within the population. Economic study of climate policy provides an apt illustration.

2.1.1. Integrated-Assessment Analysis of Climate Policy

Integrated assessment (IA) models enable quantitative evaluation of alternative climate policies. IA models provide long-run descriptions of the global economy, including the energy system and its role in economic production. They represent the climate and the links between the climatic effects of greenhouse gas (GHG) emissions and their impacts on the economy. IA models have become primary tools for comparing potential policies to reduce GHG emissions. A leading example is the Dynamic Integrated Climate Economy (DICE) model described in Nordhaus (2019).

In DICE and similar models, the economic losses from climate change are represented by damage functions that give the decreases in world-wide output resulting from increases in mean global temperature, as a proportional reduction or in dollar terms. Policy comparisons have been performed by considering a planner who seeks to make optimal trade-offs between the costs of carbon abatement and the global economic damages from climate change. The planner is assumed to face an optimal-control problem, the objective being to minimize the present discounted costs of abatement and damages over a time horizon. Thus, present discounted gross world product expresses social welfare.

This measure of social welfare ignores the distribution of personal welfare in the population. Discussions of climate policy often refer to distributional impacts, but only verbally. Consider, for example, a report on decarbonization policy published by a committee of the National Academies of Science, Engineering, and Medicine (2023). The Committee wrote that society should aim to achieve (p. 4): “a fair, equitable, and just 30-year transition” to decarbonization. It went on to write (p. 51):

---

distortionary” taxation in economics appears to be tautological, meaning only that an individual cannot alter (distort) the tax he must pay in a specified lump-sum tax assessment.

> "the objective to reduce emissions to net zero by midcentury (or 50 percent by 2030) can be thought of as a constraint with the goal to minimize cost while maximizing desirable societal objectives of equity, employment, health, and public engagement."

The Committee did not formalize the terms "fair, equitable, and just." Nor did it assess what society should do if the "objectives of equity, employment, health, and public engagement" should be in tension with one another.

2.2. Aggregating Willingness-to-Pay in Cost-Benefit Analysis

Fictional Kaldor-Hicks redistribution is prominent in forms of cost-benefit analysis (CBA) that aggregate the monetary amount that each member of a population would be willing to pay for a specified change in policy relative to a given status quo (or, alternatively, the amount each person would be willing to pay to preserve the status quo). Individual willingness to pay may be positive or negative depending on how a change in policy would affect an individual. The methodology aggregates willingness to pay across the population and uses the result to evaluate the worthiness of a policy change.

To illustrate, I draw on my critique (Manski, 2015) of a review article by Domınguez-Rivera and Raphael (2015) of the use of this type of CBA to evaluate criminal-justice policies. I quote from this article at length not to critique it, but rather because the writing is unusually clear and thoughtful. See Bar-Gill (2022) for related critique of CBA using willingness-to-pay concepts.

Using Willingness-to-Pay to Evaluate Criminal-Justice Policy

Domınguez-Rivera and Raphael (2015) call attention to some prima facie unpalatable features of measuring the social welfare of a policy by aggregate willingness-to-pay. They caution in their Introduction (p. 590):

> "Of course, there are limitations to this conceptual framework for public choice . . . To start, cost–benefit analysis and cost-effectiveness analysis provide a specific weighting (or social accounting) of the relative welfare of alternative groups in society that often conflicts with widely held beliefs

> regarding fairness and equity. This is a direct result of the use of money as the common metric used to place various benefits and costs on a common footing."

They observe that willingness to pay is positively associated with ability to pay and state that (p. 596): "This positive relationship between income and benefit and/or cost valuation ultimately results in greater weight being placed on the welfare of the well-to-do in cost-benefit calculations." They subsequently write that (p. 597): "the systematic tendency to place greater weight on the welfare of the wealthy is certainly of concern." They consider deployment of police across poor and wealthy neighborhoods and write "A cost-effectiveness analysis would recommend reallocating police officers from the poor neighborhood to the wealthy neighborhood." They point out that this type of recommendation is common, stating (p. 600): "Examples of policy proposals that pass the benefit-cost test yet have questionable equity implications abound in many policy domains."

A reader contemplating the above and related excerpts from the article might reasonably conclude that CBA is a methodology invented by economists to serve the wealthiest segment of society. A reader might also anticipate that Dominguez-Rivera and Raphael, who repeatedly express concern with equity and fairness, would ultimately reject the methodology that they have described. To the contrary, they write (p. 601): "Nevertheless, there is a strong case to make for cost-benefit analysis as a principal input for policy making, equity concerns notwithstanding." To justify this, they rely on the concept of Kaldor-Hicks efficiency, writing (p. 601): "A policy proposal that would generate net positive benefits is one that has the potential to be a Pareto improvement if a mechanism exists to redistribute some of the net benefits to those who would otherwise be the policy losers."

As have many other economists, Dominguez-Rivera and Raphael suggest that society consider equity separately from CBA. In the Introduction they write (p. 590): "Responsible analysis requires . . . .a careful parallel analysis of the equity implications of policy alternatives." In their Conclusion, they write (p. 628): "Given the unequal distribution of the cost of criminal victimization and anti-crime enforcement efforts, as well as the potential for perceived illegitimacy of the criminal justice system to undermine various public institutions, equity considerations deserve careful attention in all criminal justice policy choices." Yet

nowhere does their article provide guidance on how society might combine the willingness-to-pay type of CBA with equity considerations so as to make desirble policy decisions.

2.3. Representative-Agent Macroeconomics

It has long been the prevailing practice in macroeconomic research to assume a fictitious *representative agent.* This construct, invoked to simplify analysis of economy-wide market interactions, eliminates consideration of heterogeneity in personal preferences. Macroeconomists not only assume a representative agent but they further simplify analysis by supposing that the agent has only simple self-centered preferences. It typically is assumed that the agent has mathematically tractable preferences placing instantaneous utility on the magnitude of consumption of a single fictitious divisible commodity and that the agent maximizes a present-discounted value of this utility.

Efforts by macroeconomists to provide so-called *micro-foundations* for representative agent models, beginning with Gorman (1953), revealed that formal aggregation of heterogeneous personal preferences into the preferences of a fictitious representative agent is possible only when stringent simplifying assumptions holds. No claim has been made that these assumptions are empirically realistic. Cherrier, Duarte, and Saïdi (2023) provide an informative historical review.

The hegemony of representative-agent models in macroeconomics has begun to diminish in the 21st century, with increasing performance of research assuming various types of analytically or computationally tractable heterogeneous-agent models. However, the focus of this recent work has not been to develop social welfare functions that realistically aggregate personal preferences. It has rather been to learn how different forms of heterogeneity may affect macroeconomic phenomena of longstanding concern, including short-term business cycles and long-term growth. Rather than empirically learn the actual distribution of personal preferences, macroeconomists continue to assume that persons (or households) have tractable preferences placing instantaneous utility on the magnitude of consumption of a fictitious divisible commodity and that they maximize a present-discounted value of this utility. Rather than specify social welfare functions that

aim to be realistic, they continue to mainly use GDP as a surrogate for the *general Welfare*, viewing inequality of consumption in the population as a separate and secondary concern. Again, the review article of Cherrier, Duarte, and Saïdi (2023) is informative.

2.4. Harsanyi and Rawls As-If Consensus on the Mechanism Aggregating Personal Preferences

Within the broad subject of personalist welfare economics, the mechanism used to aggregate personal preferences has long been controversial. In principle, members of society may have heterogeneous preferences regarding the mechanism to be used. See Section 4 for discussion.

Rather than having the aggregation mechanism somehow emerge from personal preferences, the prevalent research practice has been to specify it externally. This may be done in a neutral manner as suggested by Samuelson (1947), with a researcher posing the question: "What policy would maximize the *general Welfare* using a conjectured social welfare function?" Or a researcher may make a normative argument for a particular aggregation mechanism. Doing this places the researcher in the position of having to justify why society as a whole should agree with his own aggregation preferences.

Harsanyi (1955) and Rawls (1971) both asserted that society as a whole would have consensus in favor of a particular aggregation mechanism. However, they disagreed on what this consensus mechanism would be. For Harsanyi, it was utilitarian addition of utilities. For Rawls, it was maxmin evaluation of utilities. I summarize here. I begin with the Rawls argument, which is more widely known, even though the Harsanyi proposal preceded it by about fifteen years.

The 'Initial Position' Arguments of Harsanyi and Rawls

Rawls argued that the social welfare function should be determined by a social contract. He maintained that this contract should express a consensus that he argued all rational people would accept in an *initial position*, characterized by a *veil of ignorance*. He wrote (p. 10):

> "the guiding idea is that the principles of justice for the basic structure of society are the object of the original agreement. They are the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality."

He declared that he knew what principles free and rational persons would accept, writing (p. 13):

> "I shall maintain instead that the persons in the initial situation would choose two rather different principles: the first requires equality in the assignment of basic rights and duties, while the second holds that social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society."

Thus, Rawls assumed that personal welfares are ordinally comparable across individuals. and he argued that social welfare should be the minimum personal welfare of all members of society.

Rawls did not originate the idea that all rational people would agree on a unique social welfare function in a hypothetical original position. Harsanyi (1955) posed a thought experiment of this type and reached a different conclusion. He argued that, not knowing their positions in society, individuals in the original position would place equal probability on realizing each possible position and would maximize expected utility. He thus concluded that all rational persons would accept a utilitarian social welfare function. Writing fifteen years later, Rawls barely acknowledged the precedent Harsanyi argument, mentioning Harsanyi by name only briefly in a footnote. Nevertheless, he attacked utilitarianism sharply.

Critics have questioned how one could know that all free and rational persons would accept either the Harsanyi or Rawls principles. In his review of the Rawls book, Arrow (1973) wrote (p. 247): "How do we know other peoples' welfare enough to apply a principle of justice?" . . . ."the criterion of universalizability may be impossible to achieve when people are really different, particularly when different life experiences mean that they can never have the same information." He concluded by writing (p. 263): "To the extent that individuals are really individual, each an autonomous end in himself, to that extent they must be somewhat mysterious and inaccessible to each other. There cannot be any rule that is completely acceptable to all."

3. Personalist Welfare Economics Assuming Consequentialist Self-Centered Personal Preferences

In this section and the next, I discuss research that specifies an explicit personalist social welfare function. In principle, such a function can aggregate the welfares of individuals who hold heterogeneous consequentialist and deontological preferences over abstract social states. Arrow (1978) put it this way (p. 224):

> What remains is the determination of the social ordering. On what data is it based? In particular, how does it relate to individual preferences over social states, what might be termed, "individual utilities." For purpose of this paper, I am accepting the viewpoint of the utilitarians and of welfare economics. It is assumed that each individual has some measure of the satisfaction he draws from each social state and that the social ordering is determined by the specification of these utilities for all possible social states."

This 1978 statement by Arrow differs notably from his argument against utilitarian aggregation of utilities in his famous 1951 book. He wrote there that it (Arrow, 1951, p. 11): "seems to make no sense to add the utility of one individual, a psychic magnitude in his mind, with the utility of another individual. Comparison of Arrow's statements in 1951 and 1978 illustrates well that economists have long vacillated on the meaningfulness of interpersonally comparable cardinal utilities.

In practice, economists rarely consider personal preferences over abstract social states. The convention has been to assume that preferences are consequentialist and self-centered, sometimes called egoism (Driver, 2022). Fleurbaey (2021) stated this in striking fashion, writing (p. 39):

> "It is standard in normative economics, as in political philosophy, to evaluate individual well-being on the basis of self-centered preferences, utility or advantage. Feelings of altruism, jealousy, etc. are ignored in order not to make the allocation of resources depend on the contingent distribution of benevolent and malevolent feelings among the population."

Fleurbaey did not argue that it is realistic to assume that personal preferences are self-centered; he only observed that this assumption simplifies analysis.

In this section, I provide two instructive illustrations of research assuming that personal preferences are consequentialist and self-centered. Section 3.1 discusses the utilitarian optimal income-tax theory pioneered by Mirrlees (1971). This has mainly assumed that individuals have homogeneous (consumption, leisure) preferences but heterogeneous abilities to generate labor income. Section 3.2 describes work in

health economics that uses the distribution of quality-adjusted life years to measure health-related welfare in a heterogeneous population. Section 3.3 fleshes out the aforementioned fact that, in practice, research in personalist welfare economics has not been based fully on personal preferences. Rather, researchers have themselves specified essential aspects of social welfare functions.

3.1. Optimal Income-Tax Theory

The seminal Mirrlees (1971) study of optimal income taxation has spawned a large body of pragmatic utilitarian analysis of tax policy. This body of work has mainly studied a simple static setting where individuals have homogeneous preferences for (consumption, leisure) bundles, utility increases in both goods, and individuals choose how much labor to supply. Heterogeneity stems only from interpersonal differences in abilities to generate labor income, resulting in wage heterogeneity. A social planner is unable to tax wages directly but is able to tax gross income, which is wage multiplied by labor supply. Wages are assumed to be predetermined; thus, tax policy does not affect equilibrium wages.

The planner selects a tax schedule to maximize a utilitarian welfare function, subject to a constraint that net tax revenue must suffice to fund a predetermined level of government expenditure. The welfare function used by the planner applies a specified monotone-concave transformation to the homogeneous personal utility function. This motivates progressive income tax schedules that impose higher tax rates on persons with higher incomes and lower (or negative) rates on those with lower incomes.

Mirrlees showed that, even with the many simplifying assumptions stated above, the structure of the optimal tax schedule is complex. The reason is that the tax schedule affects labor supply, in a manner that depends on consumption-leisure preferences and abilities. The sensitivity of labor supply to taxes affects how much redistribution a society can accomplish. Atkinson and Stiglitz (1980) provides a textbook exposition of Mirrlees (1971) and the contributions made soon after. Mirrlees (1997) gives a retrospective on his work and later developments. Kaplow (2024) is a recent review of the literature.

Absence of a Credible Basis for Specification of (Consumption-Leisure) Preferences

As a theorist, Mirrlees conjectured specific forms (consumption, leisure) preferences, his aim being to simplify analysis of a complex mechanism design problem. He did not assert that his preference assumptions were realistic. Indeed, the fact that he titled his 1971 article "An Exploration in the Theory of Optimal Income Taxation" makes plain that he viewed his work as exploratory rather than as providing a basis for design of actual income tax policy. Further evidence of Mirrlees' recognition that he lacked empirical knowledge of preference is found in the conclusion to his article, where he wrote (p. 207): "The examples discussed confirm, as one would expect, that the shape of the optimum earned-income tax schedule is rather sensitive to the distribution of skills within the population, and to the income-leisure preferences postulated. Neither is easy to estimate for real economies."

Credible specification of personal preferences is essential to evaluation of income tax policy. Utilitarian policy analysis should be based on actual rather than conjectured preferences if it strives to be relevant to the real world rather than only to serve as a challenge for economic theorists. In welfare-economic study of income taxation, the distribution of preferences matters in two ways. One is that it determines the personalist social welfare function. The other is that prediction of labor-supply choices, which is necessary to predict tax revenue, requires knowledge of personal preferences.

It has long been known that economic theory does not predict the direction or magnitude of the response of labor supply to income taxation. To the contrary, it shows that a rational worker may respond in disparate ways. As tax rates increase, a person may decide to work less, work more, or not change labor supply at all. See Robbins (1930). Modern labor economics envisions labor supply as a complex sequence of schooling, occupation, and work effort decisions made under uncertainty over the life course, perhaps with only bounded rationality. However, we need only consider a simple static scenario to see that a person may respond rationally to income taxes in disparate ways.

Suppose that a person with a predetermined wage and no unearned income allocates each day between paid work and the various non-paid activities that economists have traditionally called leisure. Let a

proportional income tax reduce his wage by the prevailing tax rate, yielding his net wage. Assume that the person allocates time to maximize utility, which is an increasing function of net income and leisure.

Different utility functions imply different relationships between the tax rate and labor supply. The labor supply implied by utility functions in the Constant-Elasticity-of-Substitution (CES) family increases or decreases with the tax rate depending on the elasticity of substitution. Other utility functions imply that labor supply is *backward-bending*. That is, hours worked may initially increase as net wage rises from zero but, above some threshold, decrease as net wage rises further. Still other utility functions yield more complex non-monotone relationships between net wage and labor supply. A review article by Stern (1986) describes a broad spectrum of possibilities.

Given that theory does not predict how income taxation affects labor supply, prediction requires empirical analysis. Robbins (1930) emphasized this, writing (p. 129): "we are left with the conclusion . . . . that any attempt to predict the effect of a change in the terms on which income is earned must proceed by inductive investigation of elasticities." However, it was optimistic to think that empirical analysis would resolve the matter.

Fifty years after Robbins, Atkinson and Stiglitz (1980) wrote (p. 47): "Neither economic theory nor empirical evidence can provide a conclusive answer to the effect of income taxation on labour supply." Concluding his detailed comparison of alternative utility and labor supply functions, Stern (1986) wrote (p. 173): "Our general conclusion must be in favour of diversity of functions and great caution in drawing policy conclusions on results based on a particular form." Close to thirty years later, I reached this pessimistic conclusion after studying the problem of using data on the distribution of labor supply to identify the preference distribution (Manski, 2014, p. 146): "As I see it, we lack the knowledge of preferences necessary to credibly evaluate income tax policies." Thus, I think it prudent to continue today to regard optimal income tax theory as exploratory rather than as providing a credible basis for study of actual tax policy.

3.2. QALY Evaluation of Health Welfare Using Stated-Choice Data in Hypothetical Scenarios

Whereas empirical studies of (consumption, leisure) preferences have generally used actual choice data to perform revealed-preference analysis, some other types of policy analysis have used stated-choice data elicited from survey respondents. In empirical research of this type, a researcher poses multiple hypothetical choice settings to a person and asks the person to predict the choice he would make in each setting. If the respondents are a random sample of the population of interest, the data may be used to estimate the distribution of preferences in the population. Inference on personal preferences from data on stated choices has a long history in econometric analysis of discrete choice. Ben Akiva, McFadden, and Train (2019) provide a comprehensive review.

A practical advantage of stated-choice analysis is that the choice settings considered are not limited by what nature offers up. A researcher can elicit predictions of behavior in a wide spectrum of hypothetical choice settings. A potentially serious issue is that interpretation of stated-choice data requires assumptions about the way that persons construe the scenarios posed and the cognitive processes they use when responding to questions. Manski (1999) offers a perspective on aspects of this subject.

Stated-choice data have been widely used in utilitarian health-economic assessment of medical treatments. The prevalent practice is to measure health-related cardinal utility on a [0, 1] scale called a quality-adjusted life year (QALY). In a review article, Weinstein, Torrance, and McGuire (2009) describe the scale as follows (p. S5):

> "Health states must be valued on a scale where the value of being dead must be 0, because the absence of life is considered to be worth 0 QALYs. By convention, the upper end of the scale is defined as perfect health, with a value of 1. To permit aggregation of QALY changes, the value scale should have interval scale properties such that, for example, a gain from 0.2 to 0.4 is equally valuable as a gain from 0.6 to 0.8."

Although the basic QALY scale is [0, 1], health economists often consider the possibility that persons may view some adverse health states as having value less than 0, thus extending the scale.

To measure lifetime health utility under a specified policy, health economists commonly predict future longevity, assign a QALY level to each predicted future year alive, and sum the total. They typically do so without discounting the future. In this and other respects, QALY characterization of health utility makes questionable assumptions about health preferences to simplify analysis. Nevertheless, I find it admirable in its serious effort to empirically measure preference heterogeneity to some degree.

Especially prominent has been the preference elicitation approach known as EQ-5D, developed by the EuroQol Research Foundation; see Devlin and Brooks (2017). In the United Kingdom, the National Institute for Health and Care Excellence (NICE) uses estimates of mean QALY preferences obtained with EQ-5D to make recommendations regarding the cost effectiveness of treatments proposed for coverage in the government's National Health Service (NHS); see NICE (2013). The NHS use of mean QALY valuations to inform government decisions to pay for medical treatments is a remarkable instance of an actual social planner using an empirically-based form of utilitarian welfare analysis to make large-scale policy choices.

3.3. Researcher-Specified Aspects of Personalist Welfare Functions

In practice, economists studying ostensibly personalist welfare economics have placed strong assumptions on personal preferences. Moreover, they have specified mechanisms to aggregate personal welfares rather than having these mechanisms emerge from personal preferences. Thus, welfare economics has been personalist only to a limited degree.

3.3.1. Specification of Personal Preferences

Consider research on optimal income taxation. Empirical labor economists have long strived to use data on labor supply to infer possibly complex heterogeneous preferences that persons hold for consumption and leisure. As discussed in Section 3.1, they have not had much success. Nevertheless, public economists studying optimal taxation commonly conjecture homogeneous preferences with a simple function form.

Thus, preference specification comes mainly from the mind of the researcher, not from study of actual preferences.

Health-economic study of QALYs aims to empirically measure preference heterogeneity. However, as mentioned in Section 3.2, the QALY structure places a questionable function form on preferences. Moreover, significant issues surround the use of stated-choice data to estimate preferences. Thus, the realism of QALY analysis is uncertain.

Considering research in optimal income taxation, Fleurbay and Maniquet (2018) remarked that preference specifications considered by researchers to be approximations to actual preferences may largely express the values of the researchers. They wrote (p. 1036): “Once one acknowledges that the choice of a particular utility measure is always strongly value laden, it is a small step to accept the second view and treat utilities as normative constructs.”

Writing in philosophy, Riley (1990) went further. He questioned whether analysis of social welfare should even seek to be based on actual population preferences. He considered two polar conceptualizations of utility, being (p. 338): “crudely subjective or ethical.” His pejorative wording “crudely subjective” indicated that, from a philosophical perspective, welfare computation should not necessarily be based on actual personal preferences. If some preferences are deemed not normatively acceptable, Riley suggested that a researcher place a (p. 339): “morally restricted domain condition” on the space of ethically permitted utility functions.

3.3.2. Aggregation of Personal Preferences

Harsanyi (1955) and Rawls (1971) argued that there may be societal consensus to use certain procedures for aggregating personal preferences. However, I find it hard to take their arguments as more than ingenious thought experiments. In practice, public economists studying policy specify the aggregation mechanism, commonly some version of utilitarian addition. They do not attempt to empirically elicit how members of the population think preferences should be aggregated.

I write that economists study "some version" of utilitarian addition because utilitarianism encompasses a class of aggregation mechanisms. The concept is generally understood as maximization of some weighted average of monotone transformations of cardinal personal utility functions. The particular weighted average and transformation that a researcher specifies determines the version of utilitarianism.

In research on optimal income taxation, concern to reduce inequality in personal welfare across persons who differ in ability motivates researcher specification of a concave-monotone transformation to the assumed homogeneous utility function. The degree of concavity, which affects the optimal income tax schedule and the redistribution achieved, has not been based on empirical analysis of population preferences for equity. It has been a choice made by a researcher.

Also in study of optimal taxation, researchers sometimes propose aggregation of a weighted average of cardinal utilities, the weights being determined by observable personal attributes. A recent example is Saez and Stantcheva (2016), who wrote (p. 24): "Weights directly capture society's concerns for fairness without being necessarily tied to individual utilities. Suitable weights can help reconcile discrepancies between the welfarist approach and actual tax practice." Observe their reference to "society's concerns for fairness." How to interpret this may be as difficult as interpretation of the Constitutional reference to the *general Welfare*.

3.3.3. Specification of the Discount Rate in Analysis of Climate Policy

To conclude this discussion, I continue the description of integrated-assessment analysis of climate policy begun in Section 2.1.1. Recall that, in an IA model, the economic loss from climate change is represented by a damage function that gives the decrease in world-wide output resulting from increases in global temperature. Temperature increases may be mitigated by policies that reduce GHG emissions. It is presumed that a planner seeks to make optimal trade-offs between the costs of GHG abatement and the global economic damages from climate change. In practice, the researchers who develop an IA model specify the welfare function that expresses this tradeoff.

IA models incorporate discount rates to quantify the present value of future costs and benefits. The appropriate discount rate has been a long-standing and contentious issue in climate economics. Controversy persists in part due to the fact that choice of an appropriate discount rate is not only an empirical question. It is also a normative matter concerning social preferences for equity across future generations, which will vary in their periods of existence and levels of consumption. See Dasgupta (2008).

It is not feasible to elicit the discount-rate preferences of future humans who are as yet unborn. Elicitation might in principle be feasible from the population who are alive currently, but that is not the practice. Rather, the discount rate is a choice made by the researchers developing an IA model.

The chosen discount rate can be a highly consequential determinant of conclusions on optimal climate policy. This became plain in 2007 when Nordhaus (2007) and Stern (2007) reported the findings of studies using different discount rates. Stern used a relatively low rate and concluded that policy should seek to reduce GHG emissions aggressively and rapidly. Nordhaus used a relatively high rate and concluded that policy should act moderately and slowly.

4. Welfare with Heterogeneous Preferences over Social States

I wrote earlier that a social welfare function can aggregate personal preferences over abstract social states, but economists have commonly assumed individuals to be consequentialist and self-centered. When welfare economists have studied holistic preferences across social states, they have typically viewed such preferences as ethical judgements made by a planner. Thus, economists have behaved much like moral philosophers. The work of Amartya Sen has been prominent. Recall that Sen (1977) wrote (p. 1559): "welfarism as an approach to social decisions is very restrictive."

Given that individuals may have preferences over abstract social states, I am puzzled by Sen's perspective that welfarism "is very restrictive." In his book titled *The Idea of Justice*, Sen (2009) took the stance of a moral philosopher who stands outside of society and offers ethical prescriptions for it. In Chapter

13, he belittled utilitarianism. He wrote that it presumes individuals aim to maximize "happiness," which he interpreted in a self-centered fashion.

The basic economic construct of personal welfare places no restrictions on what individuals seek to maximize. In principle, each individual may be a liberal moral philosopher, holding normative beliefs that society should adherence to notions of justice, fairness, and equity. Contrariwise, some individuals may want to enhance inequalities in directions they favor and may advocate discrimination against members of certain groups. Each individual may hold a distinct view on the composition of the population—past, present, and future generations of humans and other forms of life—whose interests should be considered when measuring social welfare. Thus, utilitarianism can be applied to a society whose members have any mix of affinity or antipathy towards other entities. Empirical study by experimental economists has accumulated an array of evidence for various forms of *other-regarding* preferences; see Cooper and Kagel (2017). Yet the implications for policy choice remain unclear.

Personal preferences need not be entirely consequentialist, having deontological components. For example, an individual may prefer an income tax structure that adheres to some concept of *equal treatment of equals* (aka *horizontal equity*) to one that violates this concept. Considering criminal justice policy, an individual may believe that the legal system should respect some concept of *due process* and that "punishment should fit the crime."

There is vast scope for expanding welfare economics to study policy choice using social welfare functions that aggregate the personal welfares of a population whose members hold heterogeneous consequentialist and deontological preferences. Consider preferences for income tax policies. Individuals may be concerned with much more than the consequences of a policy for their personal consumption and leisure. They may be concerned with the population distribution of (consumption, leisure). They may be concerned with whether the tax schedule respects some version of equal-treatment-of-equals. The welfare functions that have been specified to date in research on optimal taxation have rarely expressed these

possibilities.[5]

The scope for expanding welfare economics is much too large for me to attempt a comprehensive discussion here. In what follows I consider one aspect, often called *equity*, that has drawn substantial attention in discussions of policy choice.

4.1. Perspectives on Equity

In much popular discourse and scholarly writing, the word *equity* and related words such as *equality* and *fairness* are used loosely. Similar looseness pervades use of the word *disparity*, intended to mean the absence of some type of equity. One may surmise from the widespread use of these words and the values often attached to them—equity is good and disparities are bad—that they matter to what many individuals perceive as the *general Welfare*. However, the looseness of the usage severely impedes giving these values expression in a social welfare function.[6]

In their textbook on public economics, Atkinson and Stiglitz (1980) distinguished several broad ways in which economists have used the term *horizontal equity* when considering taxation and other policies. They wrote (p. 293): "The principle of horizontal equity states that those who are in all relevant senses identical should be treated identically." They juxtaposed three interpretations of horizontal equity that they had observed in the public economics literature:[7]

(1) (p. 294): "horizontal equity is simply an implication of the more general principle of welfare maximization."

---

[5] An early exception is Boskin and Sheshinski (1978). They assumed that personal welfare is a function not only of individual (consumption, leisure) but of the individual's position relative to the population distribution. A recent exception is Aronssohn and Johansson-Stenman (2023). They considered a society in which individuals are concerned with the mean of a transformation of disposable income in the population.

[6] Indeed, the looseness of the usage impedes achievement of a consensus interpretation of the various terms. In this section I treat equality and fairness as synonyms of equity. Some writers may view the three as distinct concepts.

[7] Considering people who are not identical in all relevant aspects, Atkinson and Stiglitz (1980) also observed that public economists studying taxation have referred to various forms of *vertical equity* and the related term *ability to pay*, advocating principles of *equal sacrifice* or *equal marginal sacrifice*. See Sections 11-4 and 13-1.

(2) (p. 294): “it is an independent principle of justice, which has to be set into the balance alongside maximization of welfare.”

(3) (p. 295): “a restriction on instruments rather than as based on a comparison of distributions.”

In interpretation (1), they meant that analysis of policy choice assuming certain personalist welfare functions and certain policy spaces yields the conclusion that homogeneous treatment of observationally identical persons maximizes social welfare. Interpretation (2) makes horizontal equity a non-personalist ethical consideration, which society may want to weigh along with maximization of a personalist welfare function. Interpretation (3) views satisfaction of horizontal equity lexicographically, as an ethical mandate constraining the space of policies that society considers. Atkinson and Stiglitz emphasized that the interpretations differ, writing (p. 295): “These three interpretations of horizontal equity are rather different. The first and second are concerned with the *results* of policy; the third is concerned with the *means* used to achieve the results.”

4.1.1. Incompatibility of Multiple Types of Equity

When one attempts to pin down concepts of equity, one must confront the fact that it may be logically impossible to jointly achieve multiple types of equity. A simple example occurs in medicine. Suppose that two groups of persons differ in their response to treatment of an illness. Then equalization of the rate at which they receive treatment will yield disparities in their health outcomes. Contrariwise, equalization of their health outcomes will require disparities in the rate at which they receive treatment. See Section 4.2 for further discussion.

Dominitz (2003) called attention to the logical problem in the context of studies of a controversial issue in criminal justice policy: racial profiling in traffic stops searching for illegal drugs. Racial profiling has been widely criticized as a deleterious form of disparity. Dominitz observed that profiling has been measured in multiple ways, among which are comparison of race-specific search rates, drug find rates, thoroughness of search, rates of detention of the innocent, and rates of apprehension of the guilty. He showed that, if crime rates differ across races, it is logically impossible to simultaneously eliminate

disparities in all of these respects. He cautioned (p. 415): “policy makers must decide whether to sacrifice equality of detention rates of the innocent, equality of apprehension rates of the guilty, or both, because they cannot be simultaneously satisfied.”

4.1.2. Ex Ante and Ex Post Treatment of Equals in Treatment Diversification

Considering planning under uncertainty, I have been intrigued by the incompatibility of ex ante and ex post equal treatment of equals. In Manski (2009, 2024) and elsewhere, I have studied treatment diversification as an approach to coping with uncertainty regarding policy outcomes. A simple and instructive case is utilitarian allocation of a population of observationally identical individuals to two feasible treatments. Contemplating assignment of everyone to a single treatment, suppose that the planner does not know which treatment yields larger social welfare, but the planner can bound these welfare values. In this setting, I have shown that a specific fractional (aka diversified) allocation minimizes maximum regret.

Diversification is consistent with the equal-treatment principle in the *ex-ante* sense that all members of the population have the same probability of receiving a particular treatment. It violates the principle in the *ex-post* sense that different persons ultimately receive different treatments. Thus, equal treatment holds ex ante but not ex post.

In a brief discussion, Atkinson and Stiglitz (1980) recognized the distinction between ex-ante and ex-post equal treatment. They wrote (p. 296): “In some writing on welfare economics, it has been assumed that *ex ante* welfare is the natural welfare function; others would argue that the *ex ante* criterion is unacceptable and even, indeed, unconstitutional as a basis for taxation.” They did not write more deeply on the subject because their book focused on policy choice in the absence of uncertainty about population-wide policy impacts. In contrast, uncertainty has been the central concern of my own research on policy choice.

Manski (2009) observed that democratic societies sometimes adhere to the ex post sense of equal treatment. Americans with identical income, deductions, and exemptions are required to pay the same federal income tax. The Equal Protection clause in the 14th Amendment to the Constitution is held to mean

that all persons in a jurisdiction are subject to the same laws, not that all persons have the same chance of being subject to different laws.

Nevertheless, many policies yield equal treatment ex ante but not ex post. American examples include random tax audits, drug testing and airport screening, calls for jury service, and the Green Card and Vietnam draft lotteries. These policies have not been prompted by the desire to cope with uncertainty that motivates treatment diversification. Yet they indicate some willingness of society to accept policies that provide ex ante equal but ex post unequal treatment.

Combining ex ante equal and ex post unequal treatment is the norm in planning under uncertainty, where observationally identical individuals respond heterogeneously to potential treatment. I next use medical treatment as a case study.

4.2. Case Study: Inclusion of Race as a Covariate in Medical Risk Prediction

4.2.1. Utilitarian Treatment Choice

Medical economists have viewed a clinician as a utilitarian planner who treats a population of patients. The clinician observes certain covariates for each patient, who has some risk of illness. The objective is to maximize mean personal welfare, often measured in QALYs. It is common to assume that care is individualistic, meaning that the care received by one patient may affect that person but does not affect others. This assumption is generally realistic when considering non-infectious diseases.

A common problem in clinical decision making is that treatments must be chosen with incomplete knowledge of their health outcomes. Medical economists often assume that the clinician knows the probability distribution of personal outcomes that will occur if a patient with specified observed covariates is given a specified treatment. In this setting, the problem of optimizing utilitarian patient care has a simple solution: patients should be divided into groups having the same observed covariates and all patients in a group should be given the care that yields the highest within-group mean patient welfare. Patients with the same observed covariates should be treated uniformly. When a within-group distribution of treatment

response is non-degenerate, the optimal policy yields ex-ante equal and ex-post unequal treatment of equals.

Achievable utilitarian welfare weakly increases as more patient covariates are observed. Observing more covariates enables a clinician to refine the probabilistic predictions of treatment outcomes on which decisions are based. Refining these predictions is beneficial if doing so affects optimal treatment choices. This important result has been discussed in the literature on medical economics in Phelps and Mushlin (1988), Basu and Meltzer (2007), Manski (2013), and elsewhere. Manski, Mullahy, and Venkataramani (2023) provide proof in the simple instructive setting of choice between two treatments.

Using all available covariate information to maximize mean patient welfare, utilitarian treatment choice formally interprets the idea that clinical decision making should be equitable: clinicians should do as well as possible for their patients, given what is known about them. Utilitarian optimization does not imply that patients with different observed covariates should receive the same treatment or that they will experience the same health. The utilitarian sense of equity may be accompanied by disparities in treatments and/or health outcomes across groups of patients. Recall the discussion in Section 4.1.1, pointing out that it may be logically impossible to jointly achieve multiple types of equity.

4.2.2. Non-Utilitarian Arguments to Exclude Race as a Predictor

Disagreeing with the utilitarian perspective of medical economics, a growing segment of the medical community in the United States have deemed certain treatment and health disparities undesirable from non-utilitarian perspectives on equity, particularly when the disparities are by race. An influential movement to remove race as a covariate in existing algorithms for medical risk prediction has developed. Commentaries by Cerdeña *et al.* (2020), Vyas *et al.* (2020), and Briggs (2022) exemplify calls to cease the use of race as a covariate. Leading institutions have recommended race-free risk prediction. A notable case is Delgado *et al.* (2021), which recommended removal of race as a predictor of kidney disease. This recommendation has since been implemented in some major medical centers.

Manski (2022) questioned four assertions that have been advanced as arguments against the inclusion of race as a covariate in medical risk prediction. These assertions are: (i) race is a social, not biological,

concept. (ii) there is no established causal link between race and the illness. (iii) using race may perpetuate or worsen racial health inequities. (iv) many persons are offended by the use of race in risk assessment.

I observed that assertions (i), (iii), and (iv) are empirical assertions that have been the subject of considerable controversy, with evidence being scant. I observed that assertions (i) and (ii) are not relevant to assessment of race as an informative predictor of illness. With the stated goal of making clinical decisions that would maximize utilitarian welfare, I concluded with this observation (p. 2113):

> "If an alternative perspective is to have a compelling foundation, it should explain why society should find it acceptable to make risk assessments using other patient characteristics that clinicians observe, but not race. It should explain why the social benefit of omitting race from risk assessment is sufficiently large that it exceeds the harm to the quality of patient care."

I made this statement in frustration because advocates of removing race from medical risk assessments have argued loosely. They have not made clear the logic of claiming that use of race as a predictor in medical risk assessment is not appropriate if race is a social construct. Nor have they made clear the logic of concern with causality, given that statistical association suffices for successful prediction within a population.

Moreover, empirical analysis has been lacking. It is not known to what extent patient populations would be willing to give up some accuracy in the medical predictions made for them, in order to mitigate the types of racial disparities that medical commentators have argued are normatively undesirable. Nor has there been much empirical study of how elimination of race as a predictor affects the magnitudes of the various types of disparities that commentators have deemed problematic. Manski, Mullahy, and Venkataramani (2023) elaborate on these themes.

4.3. Heterogeneous Preferences for Preference Aggregation

To reiterate points made earlier, personalist welfare economics specifies a social welfare function that aggregates personal welfare. To the extent that personal preferences are heterogeneous, optimal policy choice is necessarily sensitive to the aggregation mechanism, which determines how society makes interpersonal tradeoffs. Welfare economics has viewed the aggregation mechanism as a meta-decision made

by a social planner, who stands outside of society and acts in its behalf. Thus, personalist welfare economics has in fact been only partially personalist.

To enrich the personalist aspect, economists could seek to learn what preferences individuals have over alternative aggregation mechanisms. They could then conjecture a planner who decides how to aggregate preferences for the aggregation mechanism. It appears that welfare economics cannot be completely personalist. A planner would have to specify a second-order welfare function to adjudicate heterogeneity in personal preferences for the aggregation mechanism.

However the aggregation mechanism is chosen, it must somehow make interpersonal tradeoffs. The additive structure of utilitarian aggregation has long drawn criticism because it provides no inherent guarantee against extreme inequality in realized personal welfares. I next call attention to alternative aggregation mechanisms that provide guarantees of different types, but have their own arguably unattractive features.

4.4. Maximin and Minimax-Regret Policy Choice

The mathematical resemblance of social planning to individual decision making under uncertainty can be used to generate alternatives to utilitarian aggregation of personal preferences. Consider a population of individuals having heterogeneous preferences over social states, these preferences being cardinal and interpersonally comparable. The aggregation problem faced in personalist welfare economics is similar to the one studied in decision theoretic analysis of an individual who must choose an action without knowing the true state of nature. This suggests application of aggregation criteria in that literature.

Decision theory contemplates an individual who specifies a state space S and chooses an action from a choice set C, without knowing the true state. An objective function $w(\cdot, \cdot)$: $C \times S \rightarrow R^1$ maps actions and states into personal welfare. Prominent prescriptions for decision making are maximization of subjective expected utility (SEU), the maximin criterion, and the minimax-regret (MMR) criteria.

A decision maker maximizing SEU places a subjective distribution, say π, on the state space and solves the problem

$$(1) \qquad \max_{c \in C} \int w(c, s) d\pi.$$

The maximin choice selects an action that works uniformly well over S, in the sense of maximizing minimum welfare. The criterion is

$$(2) \qquad \max_{c \in C} \min_{s \in S} w(c, s).$$

The MMR criterion expresses a different formalization of the idea of selecting an action that works uniformly well over S. It solves the problem

$$(3) \qquad \min_{c \in C} \max_{s \in S} [\max_{d \in C} w(c, s) - w(c, s)].$$

Here $\max_{d \in C} w(d, s) - w(c, s)$ is the *regret* of action c in state s; that is, the loss in state s arising from choosing c rather than the optimal choice in that state. The true state being unknown, one evaluates c by its maximum regret over all states and selects an action that minimizes maximum regret. The maximum regret of an action measures its maximum distance from optimality across states.

In the welfare-economic setting, the choice set is the policy space. The members of the population play the role of the state space and their collection of personal welfare functions plays the role of the state-dependent objective function. The similarity of decision theory and welfare economics is incomplete because the latter has no analog to the true state of nature. An individual in the decision-theoretic setting realizes only the welfare attained in the true state, but society in the planning setting realizes the entire collection of personal welfares. These differences notwithstanding, decision criteria (1) through (3) can be applied to social planning.

Classical utilitarian welfare analysis has form (1), π being the uniform distribution. Versions of utilitarianism that weight individuals differentially use non-uniform π. The maximin and MMR criteria have forms (2) and (3) respectively. These criteria provide equity guarantees that are lacking in utilitarian maximization. A maximin planner maximizes the minimum welfare attained in the population, as advocated by Rawls (1971). An MMR planner minimizes the maximum degree of sub-optimal personal welfare experienced in the population.

The maximin and MMR criteria both warrant consideration in social planning. Rawls (1971) stimulated considerable discussion of maximin planning, including the critique of Arrow (1973) that I mentioned earlier. In contrast, MMR planning has received little attention.

I have applied the decision theoretic version of the MMR criterion to problems of social planning under uncertainty in a program of research that began in the early 2000s; See Manski (2024) for a comprehensive exposition. This work has studied the decision problem of a planner with incomplete knowledge, who does not know the true state of nature. The specified social welfare function need not be personalist. In contrast, my concern here is with the conceptually different but mathematically similar problem of a planner whose social welfare function aggregates the known heterogeneous preferences of a population.

As far as I am aware, the only work using the MMR criterion in the latter context is DeCanio, Manski, and Sanstad (2022). We used the MMR criterion to account for normative disagreement about the appropriate discount rate to use when evaluating climate policy. As discussed in Section 3.3.1, there has been substantial debate among economists about the appropriate discount rate to use when evaluating alternative policies. It has been found that the optimal policy is highly sensitive to the discount rate. We found that the interval [0.01, 0.07] approximately encompasses the set of discount rates that have been used by economists across their many studies of climate policy. We computed the MMR policy assuming a collection of welfare functions indexed by this interval of rates.

We argued that the MMR criterion may have appeal as an aggregation mechanism for discount rates when choosing climate policy. The *regret* of a policy measures its degree of sub-optimality. Maximum

regret measures the maximum degree of sub-optimality across individuals who prefer to use different discount rates. Hence, the MMR policy minimizes the degree to which anyone suffers relative to their personal optimum.

While maximin and MMR planning are attractive for the equity guarantees they provide, both have arguably negative features that are innately associated with these guarantees. Maximin planning has long been criticized because each policy is evaluated only by the minimum welfare it achieves across the population, making it insensitive to the distribution of welfare above the minimum. Similarly, MMR planning may be criticized because each policy is evaluated only by its maximum degree of sub-optimality across the population, making it insensitive to the distribution of sub-optimality across the population.

The appeal of both aggregation mechanisms may be particularly questionable when preference heterogeneity in the population is binary, with a fraction $0 < \alpha < 1$ of the population having the same preferences over social states and the remaining fraction $1 - \alpha$ having a different preference. The maximin and MMR policy choices do not vary with the magnitude of α. This differs sharply from the utilitarian choice that Bentham (1776) presumably had in mind when he wrote (p. ii): a "fundamental axiom, it is the greatest happiness of the greatest number that is the measure of right and wrong."

5. Conclusion

No external planner exists in a democracy. Policy choices are made by a political process that need not maximize any social welfare function. In the United States, this process was established in the Constitution, which itself was created by a political process.

Researchers cannot definitively interpret what the framers of the Constitution had in mind when they wrote of the *general Welfare*. I nevertheless feel that personalist welfare economics can contribute usefully to policy choice in democracies. Well-reasoned analysis of public policy is extremely difficult if the *general Welfare* remains an undefined concept. Study of policy choice with specified social welfare functions enables coherent analysis.

I expect that the standard economic practice of studying maximization of utilitarian welfare should yield satisfactory policy guidance for a democracy in settings where personal preferences are relatively homogeneous. However, utilitarian maximization may be inadequate to guide policy in a highly heterogeneous society.[8] Other aggregation mechanisms such as the maximin and MMR criteria may have appeal in some contexts.

Contrary to Harsanyi and Rawls, I do not expect society to agree on an aggregation mechanism. Nevertheless, the act of formalizing aggregation mechanisms and studying the policies they yield can help guide democratic processes of policy choice. To be useful, I think it essential for welfare economics to move beyond conjecture of homogeneous, self-centered, consequentialist personal preferences. It should build on empirical analysis that adequately expresses the richness and variety of actual human preferences over social states.

[8] Observation of the American setting suggests that heterogeneity of preferences is substantial with regard to many policies. The structure of income taxation and governmental involvement in health care have long been controversial. So have policy relating to access to firearms, legality of abortion, and separation of church and state.

References

Anderson, C. (1993), "Recommending a Scheme of Reason: Political Theory, Policy Science, and Democracy," *Policy Sciences*, 26, 215-227.

Aronssohn, T. and O. Johansson-Stenman (2023), "Optimal Taxation and Other-Regarding Preferences," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4606987

Arrow, K. (1951), *Social Choice and Individual Values*, New York: Wiley.

Arrow, K. (1973), "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *The Journal of Philosophy*, 70, 245-263.

Arrow, K. (1978), "Extended Sympathy and the Possibility of Social Choice," *Philosophia*, 7, 223-237.

Atkinson, A. and J. Stiglitz (1980), *Lectures on Public Economics*, Princeton, NJ: Princeton University Press, revised edition, 2015.

Bar Gill, O. (2021), "Willingness to Pay: A Welfarist Reassessment," *Yale Journal on Regulation*," 38, 503-542.

Basu, A. and D. Meltzer (2007), "Value of Information on Preference Heterogeneity and Individualized Care," *Medical Decision Making* 27, 112-27.

Ben Akiva, M., D. McFadden, and K. Train (2019), "Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis," *Foundations and Trends® in Econometrics*, 10, 1-144.

Bentham, J. (1776), *A Fragment on Government*, Payne: London.

Blaug, M. (2007), "The Fundamental Theorems of Modern Welfare Economics, Historically Contemplated," *History of Political Economy*, 39, 185-207.

Boskin, M. and E. Sheshinski (1978), "Optimal Redistributive Taxation When Individual Welfare Depends Upon Relative Income," *Quarterly Journal of Economics*, 92, 589-601.

Briggs, A. (2022), "Healing the Past, Reimagining the Present, Investing in the Future: What Should Be the Role of Race as a Proxy Covariate in Health Economics Informed Health Care Policy?" *Health Economics.* 31, 2115-2119.

Cerdeña, J. M. Plaisime, and J. Tsai (2020), "From Race-Based to Race-Conscious Medicine: How Anti-Racist Uprisings Call Us to Act," *Lancet,* 396, 1125-1128.

Cherrier, B., P. Duarte, and A. Saïdi (2023), "Household Heterogeneity in Macroeconomic Models: A Historical Perspective," *European Economic Review*, 158, 1-16.

Chipman, J. and J. Moore (1978), "The New Welfare Economics 1939-1974," *International Economic Review*, 19, 547-584.

Cooper, D. and J. Kagel (2017), "Other Regarding Preferences," *Handbook of Experimental Economics*, Vol. 2, edited by J. Kagel and A. Roth, 217-289.

Dasgupta, P. (2008), "Discounting Climate Change," *Journal of Risk and Uncertainty,* 37, 141-169.

Dasgupta, P. (2023), "Ramsey and Intergenerational Welfare Economics," *Stanford Encyclopedia of Philosophy*.

DeCanio, S., C. Manski, and A. Sanstad (2022), "Minimax-Regret Climate Policy with Deep Uncertainty in Climate Modeling and Intergenerational Discounting," *Ecological Economics,* 201, https://doi.org/10.1016/j.ecolecon.2022.107552.

Delgado, C. M. Baweja, D. Crews *et al.* (2021), "A Unifying approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease," *Journal of the American Society of Nephrology,* 32, 2994-3015.

Devlin, N. and R. Brooks (2017), "EQ-5D and the EuroQol Group: Past, Present and Future," *Applied Health Economics and Health Policy*, 15, 127-137.

Domınguez-Rivera, P. and S. Raphael (2015), "The Role of the Cost-of-Crime Literature in Bridging the Gap Between Social Science Research and Policy Making: Potentials and Limitations," *Criminology & Public Policy,* 14, 589-632.

Dominitz, J. (2003), "How Do the Laws of Probability Constrain Legislative and Judicial Efforts to Stop Racial Profiling?" *American Law and Economics Review*, 5, 412-432.

Driver, J. (2022), "The History of Utilitarianism," Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/win2022/entries/utilitarianism-history/, accessed October 23, 2024.

Fleurbaey, M. (2021), "Normative Economics and Economic Justice," *Stanford Encyclopedia of Philosophy*.

Fleurbaey, M. and M. Maniquet (2018), "Optimal Taxation Theory and Principles of Fairness," *Journal of Economic Literature*, 56, 1029-1079.

Gorman, W. (1953), "Community Preference Fields," *Econometrica*, 63-80.

Harsanyi, J. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63, 309-321.

Hicks, J. (1939), "The Foundations of Welfare Economics," *The Economic Journal*, 49, 696-712.

Hildreth, C. (1953), "Alternative Conditions for Social Orderings," *Econometrica*, 21, 81-94.

Hillinger, C. (2005), "The Case of Utilitarian Voting," *Homo Oeconomicus*, 22, 295-321.

Hsieh, N. and H. Andersson (2021), "Incommensurable Values," *Stanford Encyclopedia of Philosophy*.

Kaldor, N. (1939), "Welfare Propositions in Economics and Interpersonal Comparisons of Utility," *The Economic Journal*, 49, 549-552.

Kaplow, L. (2024), "Optimal Income Taxation," *Journal of Economic Literature*, 62, 637-738.

List, C. (2022), "Social Choice Theory," *Stanford Encyclopedia of Philosophy*.

Manski, C. (1999), "Analysis of Choice Expectations in Incomplete Scenarios," *Journal of Risk and Uncertainty*, 19, 49-66.

Manski C. (2006), "Search Profiling with Partial Knowledge of Deterrence," *Economic Journal,* 116, F385-F401.

Manski, C. (2009), "Diversified Treatment under Ambiguity," *International Economic Review*, 50, 1013-1041.

Manski, C. (2011), "Policy Analysis with Incredible Certitude," *The Economic Journal*, 121, F261-F289.

Manski, C. (2013), "Diagnostic Testing and Treatment under Ambiguity: Using Decision Analysis to Inform Clinical Practice," *Proceedings of the National Academy of Sciences*, 110, 2064-2069.

Manski, C. (2015), "Narrow or Broad Cost-Benefit Analysis?" *Criminology and Public Policy*, 14, 647-651.

Manski, C. (2022), "Patient-Centered Appraisal of Race-Free Clinical Risk Assessment," *Health Economics,* 31, 2109-2114.

Manski, C. (2024), *Discourse on Social Planning under Uncertainty*, Cambridge: Cambridge University Press.

Manski, C., J. Mullahy, and A. Venkataramani (2023), "Using Measures of Race to Make Clinical Predictions: Decision Making, Patient Health, and Fairness," *Proceedings of the National Academy of Sciences*, 120, e2303370120, https://www.pnas.org/doi/10.1073/pnas.2303370120

Meyer, L. (2021), "Intergenerational Justice," *Stanford Encyclopedia of Philosophy*.

Mirrlees J. (1971), "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175-208.

Mirrlees, J. (1997), "Incentives and Information: The Economics of Carrots and Sticks," *Economic Journal*, 107, 1311-1329.

National Academies of Sciences, Engineering, and Medicine (2023), *Accelerating Decarbonization in the United States: Technology, Policy, and Societal Dimensions*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25931.

NICE (2013), *Guide to the Methods of Technology Appraisal 2013*, https://www.nice.org.uk/process/pmg9/chapter/the-appraisal-of-the-evidence-and-structured-decision-making, accessed December 15, 2024.

Nordhaus, W. (2007), "A Review of the *Stern Review on the Economics of Climate Change*," *Journal of Economic Literature*, 45, 686-702.

Nordhaus, W. (2019), "Climate Change: The Ultimate Challenge for Economics," *American Economic Review,* 109, 1991-2014.

Phelps, C. and A. Mushlin (1988), "Focusing Technology Assessment Using Medical Decision Theory," *Medical Decision Making* 8, 279-289.

Polinsky, M. and S. Shavell (2000), "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature*," 38, 45-76.

Rawls, J. (1971), *A Theory of Justice*, Cambridge: Harvard University Press.

Riley, J. (1990), "Utilitarian Ethics and Democratic Government," *Ethics*, 100, 335-348.

Samuelson, P., (1947), *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.

Saez, E. and S. Stantcheva (2016), "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," *American Economic Review*, 106, 24–45**.**

Sen, A. (1977), "On Weights and Measures: Informational Constraints in Social Welfare Analysis," *Econometrica*, 45, 1539-1572.

Sen, A. (2009), *The Idea of Justice*, Cambridge, MA: Harvard University Press.

Stern, N. (2007), *The Economics of Climate Change: The Stern Review,* Cambridge: Cambridge University Press.

Vyas, D., L. Eisenstein, and D. Jones (2020), "Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms," *New England Journal of Medicine,* 383, 874-882.

Weinstein, M., G. Torrance, and A. McGuire (2009), "QALYs: The Basics," *Value in Health*, 12, S5-S9.

Zamir, E. and B. Medina (2008), "Law, morality, and economics: integrating moral constraints with economic analysis of law," *California Law Review*, 96, 323-392.

Zamir, E. and B. Medina (2010), *Law, Economics, and Morality*, Oxford: Oxford University Press.