1

Credible Planning under Uncertainty

A foundational objective of the Constitution of the United States is to "promote the general Welfare." 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.

The Constitution does not define "general Welfare."

A century later, Marshall (1890) began his *Principles of Economics* with this sentence (p. 1):

> Political economy or economics is a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of wellbeing.

The word "wellbeing" may be synonymous with welfare.

In this century, a report on clinical practice guidelines by the US Institute of Medicine (IOM) stated (Institute of Medicine, 2011, p. 4):

> Clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.

The report did not specify what it means to optimize patient care.

The Constitutional premise that the United States should promote the general welfare, Marshall's concern with social action to promote wellbeing, and the IOM premise that clinicians should optimize patient

care exemplify broad assertions that entities making societal decisions should aim to maximize social welfare. Such assertions may have rhetorical appeal but they lack substance. They become meaningful only when several questions are answered: What constitutes social welfare? What are the feasible actions? What is known about the welfare consequences of alternative choices?

Maximization of welfare is a well-defined objective if enough is known about the welfare consequences of alternative choices to determine an unambiguous best action. Maximization is ill defined if the consequences are sufficiently uncertain that no action is clearly best. My concern is reasonable societal decision making in such settings.

What are the uncertainties with which planning must cope? They are too many and varied to summarize easily. For now, I will simply list those that I have studied, each of which will be discussed in this book. These include numerous uncertainties in medical risk assessment and prediction of treatment response; see Manski (2019a) for a broad exposition. There is much uncertainty in the epidemiological models used to predict the spread of infectious diseases, which inform choice of vaccination policy (Manski, 2010, 2017). There is also much uncertainty in the physical science climate models used to predict future climate change, which inform choice of climate policy (Manski, Sanstad, and DeCanio, 2021), and in the discount rate used to form a social welfare function (DeCanio, Manski, and Sanstad, 2022).

Challenging uncertainties arise when studying the preferences and behavior of human populations. Knowledge of preferences is essential to policy evaluation when welfare is utilitarian. An ability to predict behavior is required to evaluate policy consequences whatever the welfare function may be. Manski (2007c) provides an abstract analysis. Manski (2014a, 2014b) examined how uncertainties about preferences and behavior complicate evaluation of income tax policies, where a central consideration is the relative preferences of potential workers for consumption goods and for availability of time to enable nonpaid activities. I have shown how uncertainty about the effect of policing on criminal behavior complicates evaluation of proactive policing programs (Manski, 2006).

ORGANIZATION OF THE BOOK

I lay out basic themes in abstraction in this opening chapter and flesh them out in what follows. Part I, constituting Chapters 2 through 4, is concerned with characterization of uncertainty. Part II, being Chapters 5

through 9, describes my research analyzing particular classes of planning problems. Chapter 10 looks ahead to performance of future research on social planning under uncertainty.

In this initial chapter, Section 1.1 calls attention to the prevalent research practice that studies planning with *incredible certitude*. Section 1.2 contrasts the conceptions of uncertainty in consequentialist and axiomatic decision theory. Section 1.3 presents the formal structure of consequentialist theory, which will be used throughout the book. Section 1.4 explains the prevalent econometric characterization of uncertainty, which distinguishes identification problems and statistical imprecision. Section 1.5 discusses the distinct perspectives on social welfare expressed in various strands of research on planning.

In Part I, Chapter 2 demonstrates how incredible certitude harms analysis of planning and assesses explanations that have been suggested for the prevalence of incredible certitude. Chapter 3 considers the central econometric problem of identification of treatment response. Chapter 4 discusses the comparably central problem of identification of choice behavior and the distribution of personal welfare in a society.

In Part II, Chapter 5 presents a core part of my work on treatment of individuals under ambiguity, developing the theme that diversification may be socially beneficial. Chapter 6 shows that use of statistical decision theory can improve treatment choice with data from statistically imprecise randomized trials, replacing the common use of hypothesis testing. Chapter 7 discusses my research on personalized treatment under uncertainty, where the planner wants to condition treatment on observed covariates but does not know how treatment response varies across persons.

Chapter 8 considers an important setting where treatment response has social interactions, this being vaccination to prevent transmission of infectious disease. Moving from treatment of individuals to global planning, Chapter 9 exposits my collaborative research on choice of a greenhouse gas abatement policy to reduce planetary warming when the physics of climate determination and the discount rate used in the social welfare function are uncertain. Chapter 10 looks ahead, calling for work that strengthens the foundations for planning under uncertainty, and touching on certain planning problems that need immediate and long-term attention.

As far as I am aware, only a small body of other research engages any of the themes that I will discuss. In the late 1970s, Johansen (1978) called for research on macroeconomic planning under uncertainty,

stating (pp. 263–264): "Uncertainty is not something which should be considered as a theoretically interesting refinement or extension of standard theory and methodology, but a central factor of eminently practical importance. Sometimes uncertainty is itself the heart of the matter when decisions are to be taken." In the early 2000s, Hansen and Sargent initiated a program of work on robust macroeconomic policy, considering certain possible deviations of reality from the assumptions maintained in conventional macroeconomic models; see Hansen and Sargent (2008). Their work uses concepts of robust decision analysis, which I will explain in Section 1.3. Barlevy (2011) reviews work on macroeconomic policy under ambiguity.

1.1 THE PREVALENT STUDY OF PLANNING WITH INCREDIBLE CERTITUDE

Economists have long studied policy choice by an actual or hypothetical social planner who aims to maximize welfare in democracies or other political systems where, in some sense, welfare is intended to express the values of society rather than the preferences of a dictator. The public at large may not be familiar with the formal structure of welfare economics, but basic ideas are familiar through the widespread use of the term *benefit–cost analysis*. Economists often study planning with utilitarian welfare functions. They sometimes specify ones that express a form of paternalism or principles of fairness.

The motivation for studying planning is most transparent when actual planners face specific decision problems. A national government must design an income tax structure and develop a system for national defense. Local governments choose how to maintain roads, perform policing, and organize public education. Planners need not be governmental. Clinicians make medical choices on behalf of patients. Parents act as planners for their families. In these settings and many more, the objective of the planner may be to maximize some idea of social welfare.

Welfare economics has also sought to shed light on noncooperative societal decision processes, where no actual planner exists. In the late 1700s, Adam Smith metaphorically suggested that an *invisible hand* makes decentralized decision making in market economies promote social welfare. Between then and the mid 1900s, economists gradually formalized this notion to develop what have become known as the *fundamental theorems of welfare economics*. These give idealized conditions under which equilibrium outcomes in markets have the desirable welfare

property of Pareto efficiency, which would be sought by a planner using a utilitarian or other welfare function that aggregates personal welfare (aka utility).

A central concern of research in public economics has been to study societal outcomes when the idealized conditions of the fundamental theorems of welfare economics do not hold. The social welfare achieved by a hypothetical planner has served as a benchmark in social choice theory, which studies the outcomes produced by voting and other decentralized mechanisms that attempt to aggregate personal preferences. Even when actual societal decisions are made by processes distant from planning, study of hypothetical planning problems has been valuable to clarify the respects in which the members of society agree and to make explicit the nature of disagreements.

I wrote previously that welfare economics has studied maximization of welfare. Whether performing abstract theoretical studies or applied benefit–cost analyses, researchers have generally assumed that the planner knows enough about the choice environment to be able to determine an optimal action. However, the consequences of decisions are often highly uncertain. Aiming to circumvent this difficulty, researchers commonly invoke strong unsubstantiated assumptions and use them to study solvable optimization problems. I have referred to this practice as policy analysis with *incredible certitude* (Manski, 2011b, 2013c).

Planning with incredible certitude can harm society in multiple ways. Most obviously, it seeks to maximize the social welfare that would prevail if untenable assumptions were to hold rather than actual social welfare. If planners incorrectly believe that existing analysis provides an errorless description of the current state of society and accurate predictions of policy outcomes, they may make substantively poor decisions. Moreover, they will not recognize the value of new research aiming to improve knowledge. Nor will they appreciate the potential usefulness of decision strategies that may help society cope with uncertainty and learn. In Chapter 2, I will present a typology of research practices that generate incredible certitude and discuss many specific cases.

The dearth of study of planning under uncertainty is apparent in the comprehensive textbook on public economics of Atkinson and Stiglitz (1980), which mentions uncertainty only a few times and then only in passing. Mongin and Pivato (2016) began their review article with this sentence (p. 711): "PERHAPS surprisingly, uncertainty plays no role whatsoever in the classical works on social welfare."

Addressing the failure of research in welfare economics to come to grips with uncertainty has motivated my research program on credible social planning under uncertainty, which has developed over the past twenty-five years. The word "credible" is inevitably subjective and difficult to pin down, but I use it nonetheless.

1.2 UNCERTAINTY IN DECISION THEORY

A fundamental difficulty with welfare maximization under uncertainty is apparent even in a simple setting with two feasible actions, say A and B, and two possible choice environments, say s_1 and s_2. Suppose that action A yields higher welfare in environment s_1 and action B yields higher welfare in s_2. If it is not known whether s_1 or s_2 is the actual choice environment, it is not known which action is better. Thus, maximization of welfare is logically impossible. At most one can seek a reasonable way to make a choice. A basic issue is how to interpret and justify the word "reasonable."

Research in decision theory has posed and characterized various principles for reasonable decision making under uncertainty. Decision theory is not specifically concerned with societal decisions. It presumes the existence of an abstract decision maker who must choose among a specified set of actions. The decision maker could be an individual, a firm, or another institution. When the decision maker is an entity making societal decisions, it is a social planner. Thus, decision theory provides the formal basis for the study of social planning under uncertainty.

The description of uncertainty in decision theory is abstract. One supposes that outcomes are determined by the chosen action and by some feature of the environment, called the *state of nature*. The decision maker is assumed able to list all states of nature that could possibly occur. This list, called the *state space*, is a primitive concept which provides the most basic expression of uncertainty. The larger the state space, the less the decision maker knows about the consequences of each action. Decision theorists usually describe the state space mathematically, without reference to an actual choice problem. For example, they might describe it as a finite or a convex set.

Much of decision theory adds a secondary expression of uncertainty in the form of a probability distribution over the state space. Some studies view the probability distribution as a cognitive concept, expressing how decision makers might actually perceive uncertainty. Others view it as a mathematical construct, whose existence might be inferred from analysis

of choice behavior. Arguing for the psychological realism of subjective probabilities, Tversky and Kahneman (1974) made plain the difference between the two perspectives, writing (p. 1130):

> It should perhaps be noted that, while subjective probabilities can sometimes be inferred from preferences among bets, they are normally not formed in this fashion. A person bets on team A rather than on team B because he believes that team A is more likely to win; he does not infer this belief from his betting preferences. Thus, in reality, subjective probabilities determine preferences among bets and are not derived from them.

Two conceptually distinct but mathematically related approaches have been used to develop criteria for reasonable decision making. Consequentialist theory focuses on the substantive consequences of choices. Axiomatic theory poses choice axioms that characterize consistency of behavior across choice settings and proves *representation theorems* relating choice axioms to consequentialist decision criteria. My research has applied consequentialist rather than axiomatic theory. I explain why in Sections 1.2.1 and 1.2.2.

1.2.1 Consequentialist Decision Theory

Consequentialist decision theory specifies a welfare function and an expression of uncertainty as primitives. It then seeks reasonable criteria to make decisions. The most prevalent recommendation has been maximization of expected utility. One places a probability distribution on the state space and chooses an action that maximizes the expected value of welfare with respect to this distribution.

To assist decision makers who do not find it credible to express uncertainty through a probability distribution, decision theorists have mainly studied criteria that, in some sense, works uniformly well over all of the state space. Two prominent interpretations of this broad idea are the maximin and minimax regret criteria. I will formalize these criteria in Section 1.3 and apply them throughout the book, particularly minimax regret.

The decision theory used in my research on planning is consequentialist. I suppose that the objective is to make substantively good societal decisions in particular settings. To accomplish this, I suppose that a planner specifies a suitable welfare function, expresses uncertainty in a credible manner, and uses these primitives to make a decision. The suitability of a welfare function and the credibility of an expression of uncertainty are context specific. These matters will be discussed in general terms in Sections 1.4 and 1.5 and in specific contexts in Part II.

1.2.2 Axiomatic Decision Theory

Axiomatic decision theory poses principles, called axioms, for consistency of hypothetical behavior across a class of potential choice problems. Researchers introspect and assert it to be reasonable, or rational, that a decision maker should adhere to these choice axioms. The central research activity of axiomatic decision theorists has been to pose and prove representation theorems establishing that adherence to a specified set of axioms is equivalent to acting as if one wants to use some consequentialist decision criterion, coping with uncertainty in some manner.

Perhaps the most famous representation theorems are those of Von Neumann and Morgenstern (1944) and Savage (1954). Both theorems establish that adherence to certain axioms is equivalent to maximization of expected utility. They differ mainly in that the probability distribution on the state space used to form expected utility is pre-specified in the former work and determined within the theory in the latter. Von Neumann and Morgenstern (VN-M) viewed the probability distribution as a primitive concept. Savage viewed the distribution as a construct that may in principle be inferred from analysis of choice behavior. I explain this distinction further on. I emphasize that in neither theorem does the probability distribution have any necessary connection to an objective reality.

Axiomatic theorists have long debated which axioms have normative appeal. Appraisal of normative appeal rests on introspection, so there should be no expectation that consensus will emerge. Indeed, decision theorists exhibit considerable difference in opinion. Binmore (2009) catalogues and assesses a wide spectrum of consistency axioms.

Why should one consider the VN-M, Savage, or other axioms to be compelling? No theorem answers this question. Instead, decision theorists call for introspection. In lecture notes for a Ph.D. course in decision theory, Kreps (1988) counseled a decision maker contemplating application of the VN-M theorem that he must first (p. 5): "Decide that you want to obey the axioms because they seem reasonable guides to behavior."

Considering the matter at length, Savage (1954) put it this way (p. 7):

> I am about to build up a highly idealized theory of the behavior of a "rational" person with respect to decisions. In doing so I will, of course, have to ask you to agree with me that such and such maxims of behavior are "rational." In so far as "rational" means logical, there is no live question; and, if I ask your leave there at all, it is only as a matter of form. But our person is going to have to make up his mind in situations in which criteria beyond the ordinary ones of logic will be necessary. So, when certain maxims are presented for your consideration, you must ask yourself whether you try to behave in accordance with them, or, to put it differently, how you would react if you noticed yourself violating them.

After discussing the positive role of logic in guiding actual human behavior, Savage wrote (p. 20):

> The principal value of logic, however, is in connection with its normative interpretation, that is, as a set of criteria by which to detect, with sufficient trouble, any inconsistencies there may be among our beliefs, and to derive from the beliefs we already hold such new ones as consistency demands. It does not seem appropriate here to attempt an analysis of why and in what contexts we wish to be consistent; it is sufficient to allude to the fact that we often do wish to be so.

Then, addressing his basic axiom P1, which assumes that the decision maker places a complete binary preference ordering on all potential actions, he wrote:

> Pursuing the analogy with logic, the main use I would make of P1 and its successors is normative, to police my own decisions for consistency and, where possible, to make complicated decisions depend on simpler ones. Here it is more pertinent than it was in connection with logic that something be said or why and when consistency is a desideratum, though I cannot say much.

Thus, Savage opined that humans may want their behavior to be consistent beyond the degree required by logic, but he was unable to explain why.

In a famous critique of the Savage axioms, Ellsberg (1961) sharply questioned the Savage conclusion that a rational decision maker must behave as if he places a subjective probability distribution on the state space. He observed that thoughtful persons sometimes exhibit behavioral patterns that violate the Savage axioms in ways implying that they do not hold subjective distributions. Considering this behavior, he wrote (p. 669):

> Are they foolish? It is not the object of this paper to judge that. I have been concerned rather to advance the testable propositions: (1) certain information states can be meaningfully identified as highly ambiguous; (2) in these states, many reasonable people tend to violate the Savage axioms with respect to certain choices; (3) their behavior is deliberate and not readily reversed upon reflection; (4) certain patterns of "violating" behavior can be distinguished and described in terms of a specified decision rule.
>
> If these propositions should prove valid, the question of the optimality of this behavior would gain more interest. The mere fact that it conflicts with certain axioms of choice that at first glance appear reasonable does not seem to me to foreclose this question; empirical research, and even preliminary speculation, about the nature of actual or "successful" decision making under uncertainty is still too young to give us confidence that these axioms are not abstracting away from vital considerations. It would seem incautious to rule peremptorily that the people in question should not allow their perception of ambiguity, their unease with their best estimates of probability, to influence their decision: or to assert that the manner in which they respond to it is against their long-run interest

and that they would be in some sense better off if they should go against their deep-felt preferences. If their rationale for their decision behavior is not uniquely compelling …, neither, it seems to me, are the counterarguments. Indeed, it seems out of the question summarily to judge their behavior as irrational: I am included among them.

In any case, it follows from the propositions above that for their behavior in the situations in question, the Bayesian or Savage approach gives wrong predictions and, by their lights, bad advice. They act in conflict with the axioms deliberately, without apology, because it seems to them the sensible way to behave. Are they clearly mistaken?

When studying consistency axioms of the types posed by VN-M and Savage, decision theorists ordinarily do not differentiate between private entities and social planners. The presumption is that all decision makers should behave consistently in the same manner. However, some theorists have proposed that social planners should adhere to additional ethical axioms that require them, in some sense, to respect the preferences of their populations and/or behave fairly. Review articles include Fleurbaey (2018) and Mongin and Pivato (2016).

Representation Theorems

I now remark further on representation theorems. The staple formalism of axiomatic decision theory considers a collection of hypothetical choice settings and proposes axioms that mandate specific forms of consistency of behavior across settings. A representation theorem proves that adherence to the axioms is necessary and sufficient for behavior across settings to be representable as solution of some consequentialist optimization problem.

Consider the VN-M and Savage representation theorems. Both begin with a basic axiom stipulating that a decision maker has a complete binary preference ordering over a universe A of actions. They then propose further axioms mandating certain consistency properties for the preference ordering. The theorems prove that adherence to the axioms is necessary and sufficient for representation of behavior when facing any hypothetical choice set $D \subset A$ as maximization of expected utility.

Consequentialist decision theory takes the utility function to be a primitive specified by the decision maker to express what he wants to achieve. In contrast, the representation theorems of axiomatic theory view the utility function as a mathematical construct implied by hypothetical choice behavior. In neither the VN-M nor the Savage theorem does the distribution on the state space have any necessary connection to an objective reality. Considering this distribution, Berger (1985) cautioned that (p. 121)

"a Bayesian analysis may be 'rational' in the weak axiomatic sense, yet be terrible in a practical sense if an inappropriate prior distribution is used." Berger's comment expresses the consequentialist perspective that a decision maker should express uncertainty in a realistic manner.

Although the VN-M and Savage theorems both represent behavior as maximization of expected utility, they differ in how they view uncertainty. A central primitive of VN-M is an externally specified probability distribution on the state space. This could be a subjective distribution formed by a cognitive process but research in the VN-M tradition often presumes it to be a credible objective distribution.

The Savage theorem does not pre-specify a distribution on the state space. Instead, it proves that a decision maker who adheres to the axioms behaves as if he maximizes expected utility using a (utility function, state-space distribution) pair implied by hypothetical choice behavior. Thus, the utility function and the probability distribution of the Savage theorem are both constructs determined within his representation theorem. The credibility of the implied distribution plays no role in the Savage paradigm.

Axiomatic decision theorists often use language that obscures the distinction between hypothetical and actual choice behavior. They often describe axiomatic theory as revealed preference analysis. Consider, for example, this passage in Savage (1954) concerning two actions labeled f and g (p. 17): "I think it of great importance that preference, and indifference, between f and g be determined, at least in principle, by decisions between acts and not by response to introspective questions." The critical phrase in this sentence is "at least in principle." The enormously rich choice data contemplated in the Savage axioms are essentially never available in practice. This has been pointed out repeatedly over the years, at least as early as Sen (1973). Nevertheless, some theorists continue to describe their subject as revealed preference analysis; see Gul and Pesendorfer (2008).

Is Axiomatic Theory Relevant to Planning?

In Manski (2011a), I argued from a consequentialist perspective that a decision maker facing an actual choice problem is not concerned with the consistency of his behavior across hypothetical choice scenarios. The decision maker wants to make a substantively reasonable choice in the setting that he actually faces. I called this idea *actualist rationality*, stating (p. 196): "Prescriptions for decision making should promote welfare maximization in the choice problem the agent actually faces." The

word *actualist* is seldom used in modern English but an old definition captures the idea well. *Webster's Revised Unabridged Dictionary, 1913 Edition* defines an actualist as, "One who deals with or considers actually existing facts and conditions, rather than fancies or theories."

From the perspective of actualist rationality, one need not introspect regarding the normative appeal of choice axioms. Axiomatic theory might become relevant if researchers were to show that adherence to certain axioms promotes substantively good decision making. However, this has not been the objective of axiomatic theory. The representation theorems of the theory are interpretative rather than prescriptive. The decision maker contemplated in axiomatic theory is assumed to know how he would behave when facing any choice set. Hence, he has no need for prescriptions.

Some decision theorists have suggested that axiomatic theory may describe a psychological process in which persons use axioms as a cognitive tool to learn their own preferences. Sugden (1990) alluded to such a process when he wrote (p. 762): "One of the main ways in which we come to know our own preferences is by noting how we in fact choose, or by constructing hypothetical choice problems for ourselves and monitoring our responses." Binmore (2009, section 7.5) interpreted Savage as having in mind a "massaging process," in which a decision maker modifies his hypothetical decisions until he feels comfortable that the implied subjective distribution adequately expresses his beliefs. The idea appears to be that a person holds coherent probabilistic beliefs internally but is psychologically unable to express them directly. Contemplating hypothetical choice problems helps the person discover his internal beliefs.

I find it difficult to reconcile the use of consistency axioms as cognitive tools with the formal structure of axiomatic decision theory. The theory formally contemplates a being who arrives with a complete preference ordering, not a cognitively challenged creature who uses thought experiments with hypothetical choice problems to learn about itself. Thus, efforts to motivate adherence to consistency axioms as tools for cognition lie entirely outside of formal axiomatic theory.

As I see it, a fundamental problem with axiomatic decision theory is that it provides no connection to substantively good decisions. Choice axioms only aim to characterize procedural reasonableness, or rationality, in the sense of consistency of hypothetical behavior across potential choice problems. It is particularly troubling that axiomatic theory is unconcerned with the credibility of a decision maker's expression of

uncertainty. Theory in the tradition of VN-M may assume that a decision maker holds objectively accurate probabilistic expectations (aka rational expectations), but it does not explain how this may be accomplished in practice. The accuracy of probabilistic expectations is not germane to theory in the Savage tradition. The realism of expectations should matter to any decision maker. It should matter particularly to a planner who represents a population.

1.2.3 The Institutional Separation of Research on Planning and Actual Planning

Decision theory presumes a unitary setting in which a planner performs his own research to inform policy choice. I observed in Manski (2013c) that modern democratic societies have created an institutional separation between policy analysis and decision making, with professional analysts reporting findings to representative governments. Separation of the tasks of analysis and decision making, the former aiming to inform the latter, appears advantageous from the perspective of division of labor. No one can be an expert at everything. In principle, having researchers study planning problems and provide their findings to law makers and civil servants enables these planners to focus on the challenging task of policy choice, without having to perform their own research.

I also observed that the current practice of policy analysis with incredible certitude does not serve planners well. The problem is that the consumers of policy analysis cannot trust the producers. I argued that, to improve analysis and to increase trust, research on planning should transparently face up to uncertainty rather than hide it.

Some think this idea to be naive or impractical. I have repeatedly heard policy analysts assert that policy makers are either psychologically unwilling or cognitively unable to cope with uncertainty. Some economists with experience in the federal government of the United States have suggested to me that concealment of uncertainty is an immutable characteristic of the American policy environment. Hence, they assert that the prevailing practice of policy analysis with incredible certitude will have to continue as is.

A more optimistic possibility is that concealment of uncertainty is a modifiable social norm. My hope is that salutary change will occur if awareness grows that incredible certitude is harmful. Then I anticipate that the scientific community will reward policy research based

on credible analysis more than optimization exercises performed with ill-conceived assumptions. Planners and the public will want researchers to provide reasonable policy recommendations that recognize the subtlety of planning under uncertainty, not unequivocal ones that lack foundation.

1.3 THE STRUCTURE OF CONSEQUENTIALIST DECISION THEORY

1.3.1 The Choice Set, State Space, and Welfare Function

I now deepen the discussion of uncertainty in consequentialist decision theory. The starting point is to suppose that the planner or other decision maker faces a predetermined choice set C and believes that the true state of nature s^* lies in a state space S. The welfare function $w(\cdot, \cdot)\colon C \times S \to R^1$ maps actions and states into welfare. The planner wants to maximize $w(\cdot, s^*)$ over C but does not know s^*. Hence, maximization is infeasible except in special cases.

The state space S provides the basic decision theoretic expression of uncertainty. In lay language, S is a list of "known unknowns." States of nature that are not elements of S are presumed impossible to occur. Decision theory supposes that the decision maker does not contemplate the possible existence of unlisted "unknown unknowns."

Discussions of the state space often consider it to express uncertainty purely about the physical and social environment within which choice takes place. However, a state space can also express uncertainty about the welfare function that a planner should maximize. This often occurs when the planner is utilitarian. Then the planner must know the preferences of the population to maximize welfare, but this knowledge may not be available. See Chapter 4 for further discussion.

Being a primitive of the decision problem, the state space is necessarily subjective. This does not imply, however, that it is an arbitrary construction. Credibility is a fundamental matter in consequential decision theory in general and in the study of social planning specifically. If planning decisions are to enhance welfare in the real world, the planner should specify a state space that embodies some reasonable sense of credibility. Research seeks to help by providing at least a partially objective basis for specification of the state space. This basis is obtained by combining plausible theory with empirical analysis. I discuss this further in Section 1.4.

1.3.2 Decision Criteria

It is generally accepted that decisions should respect dominance. Action $c \in C$ is weakly dominated if there exists a $d \in C$ such that $w(d,s) \geq w(c,s)$ for all $s \in S$ and $w(d,s) > w(c,s)$ for some $s \in S$. To choose among undominated actions, decision theorists have proposed various ways of using $w(\cdot, \cdot)$ to form functions of actions alone, which can be optimized. In principle, one should only consider undominated actions, but it is often difficult to determine which actions are undominated. Hence, in practice it is common to optimize over the full set of feasible actions. I define decision criteria accordingly.

I initially consider settings without sample data, describing three prominent criteria. I extend these criteria to settings with sample data in Section 1.3.3. Consequentialist decision theory views the welfare function, state space, and decision criterion as meta-choices made by a decision maker. It views these meta-choices as predetermined rather than matters to be studied within the theory. In this sense consequentialist theory requires introspection, as does axiomatic theory.

A familiar idea is to place a subjective probability distribution π on the state space, average state-dependent welfare with respect to π, and maximize subjective expected welfare (SEW) over C. The criterion solves:

$$\max_{c \in C} \int w(c,s) d\pi. \tag{1.1}$$

Observe that, given a subjective distribution π on S, one need not average over π. Any criterion respecting stochastic dominance has a consequentialist claim to be reasonable. For example, Manski (1988) studied maximization of quantile welfare. However, the prevalent practice has been to average over π.

In the absence of a subjective distribution on S, a prominent idea is to choose an action that, in some sense, works uniformly well over all of S. This yields the maximin and minimax regret (MMR) criteria. The maximin criterion maximizes the minimum welfare attainable across S, solving the problem:

$$\max_{c \in C} \min_{s \in S} w(c,s). \tag{1.2}$$

The MMR criterion solves:

$$\min_{c \in C} \max_{s \in S} \left[\max_{d \in C} w(d,s) - w(c,s) \right]. \tag{1.3}$$

Here $\max_{d \in C} w(d,s) - w(c,s)$ is the *regret* of action c in state s. 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. Hence, maximum regret is uniform nearness to optimality.

The maximin and minimax regret criteria are sometimes confused with one another but they yield the same choice only in certain special cases. Whereas the maximin criterion considers only the worst outcome that an action may yield, MMR considers the worst outcome relative to the best achievable in a given state of nature. Hence, the two criteria generically differ. The leading case where the two criteria coincide occurs when the best achievable welfare has the same magnitude in every state of nature; that is, $\max_{d \in C} w(d,s)$ is constant across $s \in S$. Then (1.3) reduces to (1.2) up to this additive constant.

It is also noteworthy that a decision maker who places a subjective probability distribution π on the state space might choose to minimize subjective expected regret rather than maximum regret, subjective expected regret being $\int [\max_{d \in C} w(d,s) - w(c,s)] d\pi$. The expression $\int \max_{d \in C} w(d,s) d\pi$ is constant across the feasible actions $c \in C$. Hence, minimization of subjective expected regret is equivalent to maximization of subjective expected welfare.

I will discuss criteria (1.1) to (1.3) throughout this book. Readers should be aware that these three, which arguably have been the most prominent in decision theory, are not the only criteria that may warrant attention. For example, Hurwicz (1951) suggested modification of the maximin criterion to maximize instead a weighted average of the minimum and maximum welfare attainable across the state space.

Other decision theorists have studied settings intermediate between the polar cases in which a planner asserts either a complete subjective distribution on the state space, or none. A planner might instead assert a partial subjective distribution, placing lower and upper probabilities on states, as in Dempster (1968) or Walley (1991), and then maximize minimum subjective expected welfare or minimize maximum expected regret. These criteria combine elements of averaging across states and concern with uniform performance across states. Statistical decision theorists refer to them as Γ-maximin and Γ-minimax regret (Berger, 1985). The former has drawn attention from axiomatic decision theorists, who call it "maxmin expected utility" (Gilboa and Schmeidler, 1989).

Complete Class Theorems

Complete class theorems show that, in various contexts, an action is weakly undominated only if it would be chosen by a decision maker who maximizes SEW with respect to some subjective probability distribution on S (Wald, 1950). Bayesian decision theorists sometimes cite such theorems as a reason to focus attention on maximization of SEW rather than other decision criteria (Berger, 1985). They say that the action chosen using any alternative criterion would also be chosen by an expected-welfare maximizer with some subjective distribution. Hence, they claim, one might as well think of decision makers as maximizing SEW.

Complete class theorems have proved to be useful analytical devices when studying the properties of alternative consequentialist decision criteria. However, this does not imply that decision makers should be counseled to maximize SEW. To apply the SEW criterion, one must first decide what subjective probability distribution to use and then solve the maximization problem. Complete class theorems provide no guidance on what constitutes a credible subjective distribution. They only state that choice of any undominated action can be represented as the outcome of SEW maximization with some distribution.

1.3.3 Statistical Decision Theory

Abraham Wald, in a series of contributions culminating in Wald (1950), extended consequentialist decision theory to encompass settings where the decision maker observes sample data. Wald's formulation of statistical decision theory supposes that a decision maker observes data generated by a sampling distribution which is a known function of the state of nature. To express this, let the feasible sampling distributions be denoted $(Q_s, s \in S)$. Let Ψ_s denote the sample space in state s; Ψ_s is the set of samples that may be drawn under distribution Q_s. The literature typically assumes that the sample space does not vary with s and is known. I do likewise and denote the sample space as Ψ. A statistical decision function (SDF), $c(\cdot)$: $\Psi \rightarrow C$, maps the sample data into a chosen action.

An SDF is a deterministic function after realization of the sample data but it is a random function ex ante. Hence, an SDF generically makes a randomized choice of an action. The only exceptions are SDFs that make almost-surely data-invariant choices. An SDF $c(\cdot)$ is almost-surely data-invariant in state s if there exists a $d \in C$ such that $Q_s[c(\psi) = d] = 1$.

Given that SDFs are random functions, welfare using a specified SDF is a random variable ex ante. Wald's theory evaluates the performance

of SDF $c(\cdot)$ in state s by $Q_s\{w[c(\psi),s]\}$, the ex ante distribution of welfare that it yields across realizations ψ of the sampling process. Thus, statistical decision theory is frequentist. In particular, Wald measured the performance of $c(\cdot)$ in state s by its expected welfare across samples; that is, $E_s\{w[c(\psi),s]\} \equiv \int w[c(\psi),s]dQ_s$. Not knowing the true state, a planner evaluates $c(\cdot)$ by the state-dependent expected welfare vector $(E_s\{w[c(\psi),s]\},\ s \in S)$, which is computable.

One need not measure the sampling performance of an SDF by its expected welfare across samples. Manski and Tetenov (2023) observe that any criterion respecting stochastic dominance has a claim to be reasonable. In particular, they study measurement of sampling performance by quantile welfare. However, the prevalent practice has been to measure performance by expected welfare across samples.

Statistical decision theory has mainly studied the same decision criteria as has decision theory without sample data. Let Γ denote the set of feasible SDFs, which map $\Psi \to C$. The statistical versions of criteria (1.1), (1.2), and (1.3) are:

$$\max_{c(\cdot)\in\Gamma} \int E_s\left\{w\left[c(\psi),s\right]\right\} d\pi, \tag{1.4}$$

$$\max_{c(\cdot)\in\Gamma} \min_{s\in S} E_s\left\{w\left[c(\psi),s\right]\right\}, \tag{1.5}$$

$$\min_{c(\cdot)\in\Gamma} \max_{s\in S} \left(\max_{d\in C} w(d,s) - E_s\left\{w\left[c(\psi),s\right]\right\}\right). \tag{1.6}$$

In settings of choice between two actions, SDFs can be viewed as hypothesis tests. However, evaluation of tests in the Wald theory differs fundamentally from Neyman–Pearson hypothesis testing, which I will discuss in Chapter 6. The Wald theory does not restrict attention to tests that yield a predetermined upper bound on the probability of a Type I error. Nor does it aim to minimize the maximum value of the probability of a Type II error when more than a specified minimum distance from the null hypothesis. Wald proposed for binary choice, as elsewhere, evaluation of the performance of SDF $c(\cdot)$ in state s by the expected welfare that it yields across realizations of the sampling process. See Chapter 6 and Manski (2021a) for further discussion.

Robust Decisions

Research on *robust decisions* includes a statistical literature on robust estimation and prediction (e.g., Huber, 1981; Hampel et. al., 1986) and an engineering literature on robust control (e.g., Zhou, Doyle, and

Glover, 1996). The latter has provided the basis for recent econometric analysis of robust macroeconomic policy (e.g., Hansen and Sargent, 2008).

The study of robust decisions proceeds in a different manner than statistical decision theory. Rather than specify a state space that lists all possible states of nature, the researcher poses a model space. The model specifies a single state or a relatively small set of states, typically a finite-dimensional family. Having posed the model space, a researcher may be concerned that it does not contain the true state; that is, the model may not be correct. To recognize this possibility, the researcher enlarges the model space locally, using some metric to generate a neighborhood thereof. He then acts as if the locally enlarged model space is correct. Watson and Holmes write (2016, p. 465): "We then consider formal methods for decision making under model misspecification by quantifying stability of optimal actions to perturbations to the model within a neighbourhood of [the] model space."

Although research on robust decisions differs procedurally from statistical decision theory, one can subsume the former within the latter if one considers the locally enlarged model space to be the state space. It is unclear how often this perspective characterizes what researchers have in mind. Articles often do not state explicitly whether the constructed neighborhood of the model space encompasses all states that authors deem sufficiently feasible to warrant consideration. The models specified in robust decision analyses often make strong assumptions and generated neighborhoods often relax these assumptions only modestly.

1.3.4 Minimax Regret Planning

Among the decision criteria posed above, maximization of SEW places a subjective distribution on the state space, whereas maximin and MMR do not. Concern with the basis for specification of a subjective distribution motivated Wald (1950) to study the minimax criterion (maximin in my description), writing (p. 18): "a minimax solution seems, in general, to be a reasonable solution of the decision problem when an a priori distribution … does not exist or is unknown."

I am similarly concerned with decision making with no subjective distribution on states. However, I have mainly measured performance of decisions by maximum regret rather than by minimum welfare. The maximin and MMR criteria both provide ex ante evaluations of the worst result that a decision maker may experience ex post. However, the criteria are equivalent only in special cases, particularly when optimal welfare

is invariant across states. They differ more generally. Whereas maximin considers the worst absolute outcome that an action may yield across states, MMR considers the worst outcome relative to what is achievable in a given state.

As I see it, a conceptual appeal of using maximum regret to measure performance is that it quantifies how lack of knowledge of the true state of nature diminishes the quality of decisions. The term "maximum regret" is shorthand for the maximum suboptimality of a decision criterion across the feasible states of nature. A decision with small maximum regret is uniformly near optimal across all states. Introspecting, I think this a desirable property. Each study in Part II of this book applies the MMR criterion. Some consider the maximin criterion as well.

MMR has drawn diverse reactions from axiomatic decision theorists. In a famous early critique, Chernoff (1954) observed that MMR decisions are sometimes inconsistent with the choice axiom known as independence of irrelevant alternatives (IIA). Chernoff considered this a serious deficiency, writing (p. 426):

> A third objection which the author considers very serious is the following. In some examples, the min max regret criterion may select a strategy d_3 among the available strategies d_1, d_2, d_3, and d_4. On the other hand, if for some reason d_4 is made unavailable, the min max regret criterion will select d_2 among d_1, d_2, and d_3. The author feels that for a reasonable criterion the presence of an undesirable strategy d_4 should not have an influence on the choice among the remaining strategies.

This passage is the totality of Chernoff's argument. He introspected and concluded that a reasonable decision criterion should always adhere to IIA, without explaining why he felt this way. He did not argue that minimax regret choices have adverse consequentialist consequences.

Chernoff's view has been endorsed by some modern decision theorists, including Binmore (2009). Indeed, Ken Binmore used picturesque language to express this view in my presence during a conference, declaring that violation of IIA is "Death to minimax regret" (statement made in a presentation at the Kellogg School of Management conference on "Decision Theory and its Discontents," May 1, 2009). On the other hand, Sen (1993) argued that adherence to the IIA axiom does not per se provide a sound basis for evaluation of decision criteria. He asserted that consideration of the context of decision making is essential.

Manski (2011a) argued that adherence to the IIA axiom is not a virtue per se. What matters is how violation of the axiom affects welfare.

The MMR decision is necessarily undominated when it is unique. There generically exists an undominated MMR decision when the criterion has multiple solutions. Hence, I concluded that violation of the IIA axiom is not a sound rationale to dismiss minimax regret.

1.4 UNCERTAINTY IN EMPIRICAL RESEARCH

To characterize uncertainty with enough credibility and concreteness to be useful to the study of planning, I draw heavily on my own study of partial identification in empirical research. I explain in general terms in Section 1.4.1, adding specificity in Chapters 3 and 4 and throughout Part II. Section 1.4.2 addresses how statistical imprecision in research affects planning, with specificity added in Chapter 6.

1.4.1 Identification Analysis

It has become standard in econometrics to specify the state space as a set of objective probability distributions that may possibly describe the system under study. Haavelmo (1944) did so for economic systems when he introduced *The Probability Approach in Econometrics*. Studies of treatment choice do so when they consider the population to be treated to have a distribution of treatment response.

The Koopmans (1949) formalization of identification analysis contemplated unlimited data collection that enables one to shrink the state space, eliminating states that are inconsistent with accepted theory and with the information revealed by observation of data. For most of the twentieth century, econometricians commonly thought of identification as a binary event – a feature of an objective probability distribution (a parameter) is either identified or it is not. Empirical researchers applying econometric methods combined available data with assumptions that yield point identification, and they reported point estimates of parameters. Economists recognized that point identification often requires strong assumptions that are difficult to motivate. However, they saw no other way to perform empirical research.

Yet there is enormous scope for fruitful research using weaker and more credible assumptions that partially identify population parameters. A parameter is partially identified if the sampling process and maintained assumptions reveal that the parameter lies in a set, its *identification region* or *identified set*, that is smaller than the logical range of the parameter but larger than a single point. I explain below.

Research on Partial Identification

Isolated contributions to analysis of partial identification were made as early as the 1930s, but the subject remained at the fringes of econometric consciousness and did not spawn systematic study. A coherent body of research took shape in the 1990s and has since grown rapidly. Reviews of this work include Manski (1995, 2003, 2007a), Tamer (2010), and Molinari (2020).

I first connected identification analysis with decision making under uncertainty in Manski (2000), writing (p. 416):

> This paper connects decisions under ambiguity with identification problems in econometrics. Considered abstractly, it is natural to make this connection. Ambiguity occurs when lack of knowledge of an objective probability distribution prevents a decision maker from solving an optimization problem. Empirical research seeks to draw conclusions about objective probability distributions by combining assumptions with observations. An identification problem occurs when a specified set of assumptions combined with unlimited observations drawn by a specified sampling process does not reveal a distribution of interest. Thus, identification problems generate ambiguity in decision making.

Here I followed Ellsberg (1961) in using the word *ambiguity* to signify uncertainty when one specifies a set of feasible states of nature but does not place a probability distribution on the state space. Synonyms for ambiguity include *deep uncertainty* and *Knightian uncertainty*.

The modern literature on partial identification emerged out of concern with traditional approaches to inference with missing outcome data. Empirical researchers have commonly assumed that missingness is random, in the sense that the observability of an outcome is statistically independent of its value. Yet this and other point-identifying assumptions have regularly been criticized as implausible. It was natural to ask what random sampling with partial observability of outcomes reveals about outcome distributions if nothing is known about the missingness process or if only credible assumptions are imposed.

Studying identification with missing outcome data quickly led to analysis of treatment response. A common objective of empirical research is to predict treatment response conditional on specified covariates, using data from a random sample of the population. Analysis must contend with the fundamental problem that counterfactual outcomes are not observable. At most one can observe the outcomes that have occurred under realized policies. The outcomes of unrealized policies are logically unobservable. Yet determination of an optimal policy requires comparison of all

feasible policies. For this and many other reasons, planners usually have only partial knowledge of the welfare achieved by alternative policies.

Findings on partial identification with missing outcome data are directly applicable to analysis of treatment response. Yet analysis of treatment response poses more than a generic missing-data problem. One reason is that observations of realized outcomes, when combined with suitable assumptions, can provide information about counterfactual ones. Another is that practical problems of treatment choice motivate research on treatment response and thereby determine what population parameters are of interest. For these reasons, it has been productive to study partial identification of treatment response as a subject in its own right. See Chapter 3 for further discussion, as well as Part II.

Whatever the specific subject under study, a common theme runs through the literature on partial identification. One first asks what the sampling process alone reveals about the population of interest and then studies the identifying power of assumptions that aim to be credible. This conservative approach to inference makes clear the conclusions one can draw in empirical research without imposing untenable assumptions. It establishes a domain of consensus among analysts who may hold disparate beliefs about what assumptions are appropriate. It also makes plain the limitations of the available data. When identification regions turn out to be large, we should face up to the fact that the available data and credible assumption do not support conclusions as tight as we might like to achieve.

From the perspective of planning, findings on partial identification imply that empirical research may shrink the state space for decision making but not reduce it to a single state of nature. Let S be the state space without observation of the unlimited data assumed in an identification study. Let $S_0 \subset S$ be the shrunken state space using these data. Then decision criteria (1) to (3) posed in Section 1.3.2 have the same forms, but with S_0 replacing S. In (1), the conditional subjective distribution $\pi(s \mid s \in S_0)$ replaces $\pi(s)$.

1.4.2 Statistical Imprecision

Whereas identification analysis contemplates unlimited data collection that enables one to shrink the state space, the data observed in a finite sample generated by a sampling distribution generally are not informative enough to shrink the state space. Nevertheless, Wald's development of statistical decision theory shows how sample data can be informative.

In Wald's paradigm, statistical imprecision is expressed through the state-dependent ex ante distributions $[Q_s\{w[c(\psi), s]\}, s \in S]$ of welfare that an SDF yields across realizations ψ of the sampling process. Wald's concept of an SDF embraces all mappings [data → action]. An SDF need not perform conventional statistical inference; that is, it need not use data to directly draw conclusions about the true state of nature. The prominent decision criteria that have been studied – maximin, minimax regret, and maximization of subjective expected welfare – do not explicitly perform inference on the true state.

Although SDFs need not perform conventional inference, some do. These have the form [data → inference → action], first performing inference and then using the inference to make a decision. There seems to be no accepted term for such SDFs, so I have called them *inference based* (Manski, 2021a).

The general absence of conventional inference in statistical decision theory is striking. Familiar measures of statistical imprecision, such as confidence sets and standard errors, play no role in the Wald theory. On the other hand, statistical imprecision is measured when one computes the maximum regret of an SDF; that is, its maximum distance from optimality. Maximum regret is determined jointly by the identification problem faced and by the statistical imprecision of sample data. When the true state of nature is point identified, maximum regret purely measures statistical imprecision. I will expand on this in Chapter 6 in the context of analysis of randomized trials.

1.5 PERSPECTIVES ON SOCIAL WELFARE

Given a specified choice set and welfare function, analysis of optimal planning is straightforward in abstraction, even if solution of the optimization problem may be difficult in practice. The fundamental subtleties in research on planning are conceptual rather than mathematical. If analysis is to be useful in practice, the welfare function should express normative properties acceptable to some meaningful part of the relevant society. The specified choice set should be realistic, expressing options that may actually be available. Throughout this book, I stress that analysis should appropriately recognize uncertainty in policy outcomes. To conclude this opening chapter, I comment on specification of the welfare function.

Specification of the welfare function has vexed economists and philosophers in broad terms, as well as policy analysts in particular contexts. Most research by economists, and some by philosophers and others, has

supposed that the social welfare function should somehow aggregate the personal welfares of the individuals who compose society. Yet it has long been understood that, in general, a heterogeneous society cannot develop a consensus social welfare function. The Arrow (1950) Possibility Theorem nullified the residual hope that a heterogeneous society might be able to devise a coherent non-dictatorial welfare function. How then should research on planning proceed? The literature is vast and varied.

1.5.1 The New Welfare Economics

One route was taken in the 1930s and 1940s by the economists who initiated the study of the *new welfare economics*, a term apparently first used by Hicks (1939). Wary of any criterion to choose among policies that benefit some people but harm others, they retreated to the study of Pareto efficiency with extension to the fictional redistributions proposed by Kaldor (1939) and Hicks (1939). However, this restriction on their domain of concern drastically limited their ability to study actual planning problems. This led Chipman and Moore (1978) to write (p. 548): "In this paper 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." Efficiency in the fictional Kaldor–Hicks sense has become at most a peripheral topic in economic theory, but it continues to be used in applied benefit–cost analysis, as I explain below.

1.5.2 Utilitarian Welfare

Within the body of economic research that studies planning when policies benefit some people but harm others, it has been common to specify a utilitarian welfare function. The standard theory of rational individual behavior under certainty requires only an ordinal personal concept of welfare. A utilitarian social welfare function specifies interpersonally comparable cardinal personal welfares and sums them. This gives formal expression to what Bentham (1776) may have 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."

In utilitarian welfare analysis, concern with equity is expressed through specification of the personal welfare functions, measured on a commensurate scale, that are summed to compute social welfare. A

prominent example is the Mirrlees (1971) analysis of optimal income taxation. There, personal welfare was taken to be a concave function of income, thus expressing "diminishing marginal utility of money." It follows that, all else equal, transferring a dollar from a wealthy person to a poor one increases social welfare. In the Mirrlees analysis, 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 the structure of an optimal tax schedule is complex because the schedule generally affects the amount of labor that persons choose to supply. This consideration affects how much redistribution a society is able to accomplish in practice.

Willingness to Pay and Kaldor–Hicks Efficiency

Rather than sum interpersonally comparable personal welfare values, economists performing benefit–cost analysis often sum monetary "willingness to pay" values across persons. Willingness-to-pay analysis seeks to measure 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 that each member would be willing to pay to preserve the status quo. Thus, 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 a policy change.

To justify this approach to planning, economists cite the Kaldor–Hicks argument that concerns with equity could in principle be addressed by redistribution of money, even if the redistribution is not accomplished in practice. However, if redistribution is not actually performed, the practical outcome is to choose policies that weight the welfare of the wealthy more than that of the poor.

I commented critically on willingness-to-pay analysis in Manski (2015a), responding to an article on benefit–cost analysis of criminal justice policy by Domınguez-Rivera and Raphael (2015). I find it instructive to summarize what these authors wrote, not to single them out for scrutiny, but because their discussion illustrates how economists have sought to justify the willingness-to-pay approach to planning.

Domınguez-Rivera and Raphael called attention to some unpalatable features of the methodology. They cautioned that it (p. 589): "provides 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." They observed that willingness to pay is

positively associated with ability to pay and stated 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 wrote that (p. 597): "the systematic tendency to place greater weight on the welfare of the wealthy is certainly of concern."

Nevertheless, they wrote that (p. 601): "there is a strong case to make for cost-benefit analysis as a principal input for policy making, equity concerns notwithstanding." Referring to the Kaldor–Hicks idea of fictional redistribution, they wrote that (p. 601): "any policy choice with net positive monetary benefits provides what economists call a potential Pareto improvement." They suggested that society consider equity and fairness separately from benefit–cost analysis, writing (p. 590): "Responsible analysis requires … a careful parallel analysis of the equity implications of policy alternatives." Yet they did not provide guidance on how society might combine willingness-to-pay analysis with equity considerations so as to make sensible policy decisions.

Economists often use willingness-to-pay analysis to evaluate policy quantitatively, paying only qualitative lip service to equity. This is concerning. As I see it, direct specification of interpersonally comparable personal welfares, in the manner of Mirrlees, better expresses utilitarian policy choice.

1.5.3 Maximin Welfare

One need not sum cardinal personal welfares to develop social welfare functions that respect Pareto efficiency. Among alternatives, the work of Rawls (1971) has received considerable attention outside of economics. He recommended evaluation of policy by the minimum value of interpersonally comparable ordinal personal welfares. He argued that society should evaluate social welfare in this manner, rather than the utilitarian one.

The "Initial Position" Arguments of Harsanyi and Rawls

To reach his conclusion, Rawls argued that the welfare function should be determined by a social contract. Attempting to circumvent the deep problem of aggregating heterogeneous personal welfares, he maintained that the social 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 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 under a veil of ignorance. Earlier, Harsanyi (1955) posed a thought experiment of this type and reached a different conclusion than Rawls. Harsanyi 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.

Rawls barely acknowledged the precedent Harsanyi argument, mentioning Harsanyi by name only briefly in a footnote. Nevertheless, he sharply attacked utilitarianism, writing (p. 13):

> It may be observed, however, that once the principles of justice are thought of as arising from an original agreement in a situation of equality, it is an open question whether the principle of utility would be acknowledged. Offhand it hardly seems likely that persons who view themselves as equals, entitled to press their claims upon one another, would agree to a principle which may require lesser life prospects for some simply for the sake of a greater sum of advantages enjoyed by others. Since each desires to protect his interests, his capacity to advance his conception of the good, no one has a reason to acquiesce in an enduring loss for himself in order to bring about a greater net balance of satisfaction. In the absence of strong and lasting benevolent impulses, a rational man would not accept a basic structure merely because it maximized the algebraic sum of advantages irrespective of its permanent effects on his own basic rights and interests. Thus it seems that the principle of utility is incompatible with the conception of social cooperation among equals for mutual advantage. It appears to be inconsistent with the idea of reciprocity implicit in the notion of a well-ordered society. Or, at any rate, so I shall argue.

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 (1973a) 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 his review 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. There must, or so it now seems to me, be the possibility of unadjudicable conflict, which may show itself logically as paradoxes in the process of social decision-making.

This conclusion reminds one of Arrow's Possibility Theorem.

1.5.4 Optimal Paternalism in Populations with Bounded Rationality

The norm in the study of utilitarian planning has been to assume that members of the population maximize their personal welfare. However, the realism of this assumption has long been questioned. Simon (1955) put it this way in the article that spawned the modern literature in behavioral economics (p. 101): "Because of the psychological limits of the organism (particularly with respect to computational and predictive ability), actual human rationality-striving can at best be an extremely crude and simplified approximation to the kind of global rationality that is implied, for example, by game-theoretical models." This idea has come to be called *bounded rationality*. Simon put forward this mission for research on behavior (p. 99):

> Broadly stated, the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist.

A recent development in the field of public economics has been the initiation of research on utilitarian planning in populations with bounded rationality. Behavioral economists have suggested that planners should limit the choice options available to individuals to ones deemed beneficial from a utilitarian perspective or, less drastically, should frame the options in a manner thought to influence choice in a positive way. Thaler

and Sunstein (2003) evocatively wrote that such policies express (p. 175): "libertarian paternalism." However, here and elsewhere, their discussion has been verbal rather than formal.

An early expression of formal analysis was given by O'Donoghue and Rabin (2003), who began their article as follows (p. 186):

> The classical economic approach to policy analysis assumes that people always respond optimally to the costs and benefits of their available choices. A great deal of evidence suggests, however, that in some contexts people make errors that lead them not to behave in their own best interests. Economic policy prescriptions might change once we recognize that humans are humanly rational rather than superhumanly rational, and in particular it may be fruitful for economists to study the possible advantages of *paternalistic policies* that help people make better choices.
>
> We propose an approach for studying optimal paternalism that follows naturally from standard assumptions and methods of economic theory: Write down assumptions about the distribution of rational and irrational types of agents, about the available policy instruments, and about the government's information about agents, and then investigate which policies achieve the most efficient outcomes. In other words, economists ought to treat the analysis of optimal paternalism as a mechanism-design problem when some agents might be boundedly rational.

Economists have subsequently performed a growing set of analyses of the type sought by O'Donoghue and Rabin, addressing different classes of policy choices and assuming various distributions of preferences and deviations from utility maximization.

Research in the developing field of behavioral public economics has thus far assumed that the planner understands bounded rationality in the population well enough to be able to optimize social welfare. Thus, authors have assumed that, while members of the population are boundedly rational, the planner is globally rational. However, detailed knowledge of population preferences and deviations from global rationality is rare. Manski and Sheshinski (2023) argue that a utilitarian planner with limited knowledge should not seek to optimize policy invoking assumptions that lack credibility. Instead, the planner should use a reasonable criterion to plan under uncertainty.

1.5.5 Nonpersonalist Welfare Functions

I have so far discussed research that assumes the social welfare function somehow aggregates personal welfare across society. Sen (1977) called such research *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 (p. 1559): "welfarism as an approach to social decisions is very restrictive." Sen's perspective warrants serious attention, but it seems to me that the word "welfarism" does not express his concern well. I shall instead use the word "personalism."

Nonpersonalist welfare functions place direct societal value on certain ethical concepts, beyond their possible manifestations as determinants of personal welfare. These concepts have been given many appealing names, including justice, fairness, and equity. However, they are devilishly difficult to interpret. Moreover, interpretations vary across the persons who use the concepts. See Backhouse et al. (2021) for multiple discussions.

Economists have long sought to pose and analyse concepts of fairness. For example, Tobin (1970) posed a concept of "specific egalitarianism." This idea moves away from utilitarianism, which concerns itself with the overall utility that a person experiences, instead supposing that society desires that (p. 264): "certain specific scarce commodities should be distributed less unequally than the ability to pay for them." Tobin discussed medical care as a leading case of such a specific commodity.

Foley (1967) and Varian (1974) formalized concepts of *envy-free* and *fair* allocations of resources within a population. Manski, Mullahy, and Venkataramani (2023) formalized concepts of *disparity aversion*. In this book, Chapter 5 discusses welfare functions that formalize the idea of *equal treatment of equals*. Considering policing in Chapter 7, I specify welfare functions that value the personal welfare of law-abiding citizens but do not similarly value the preferences of criminals.

1.5.6 Pragmatic Welfare

Research in welfare economics and moral philosophy has mainly been abstract. In contrast, studies of realistic classes of planning problems have specified pragmatic welfare functions. I use the word "pragmatic" to mean that researchers motivate their welfare functions by some combination of conjecture regarding societal values, empirical study of population preferences, and concern for analytical tractability.

For example, the literature on optimal taxation stemming from Mirrlees (1971) has assumed a utilitarian welfare function and has placed various restrictions on the population distribution of labor–leisure preferences; see Chapter 4 for further discussion. Research on government spending to optimize macroeconomic growth has assumed utilitarian

welfare and a representative infinite-lived household, as in Barro (1990). Integrated assessment studies of optimal climate policy has assumed that the objective is to maximize present-discounted gross world product, as in Nordhaus (2008); see Chapter 9 for further discussion. Analysis of optimal medical care may assume that the objective is to maximize a population survival rate or mean quality-adjusted life years (QALYS) net of treatment cost; see Chapters 5 through 8. Benefit–cost analyses of transportation projects quantify and weigh an array of societal project benefits and costs; see US Department of Transportation (2023).

When academic researchers specify pragmatic welfare functions, they may believe that these functions have sufficient social acceptability to make them worthy of study. However, they usually do not argue that actual planners should necessarily use these welfare functions to make decisions. The less ambitious goal is to learn what decisions would be optimal if specified welfare functions were to be used. This perspective is maintained throughout my own work. Research should be distinct from advocacy.