

Optimal Paternalism in a Population with Bounded Rationality: with Focus on Discrete Choice

Charles F. Manski^a and Eytan Sheshinski^b

February 2024

Abstract

We consider a utilitarian planner with the power to design a discrete choice set for a population with bounded rationality. We find that optimal paternalism is subtle. The policy that most effectively constrains or influences choices depends on both the preference distribution and the choice probabilities measuring the extent to which persons behave suboptimally. We caution against implementation of paternalistic policies that optimize welfare using behavioral assumptions that lack credible foundation. In the absence of firm empirical understanding of behavior, such policies may do more harm than good. To develop these themes, we first consider the planning problem in abstraction. We next examine policy choice when individuals are boundedly rational in a specific way, this being that they measure utility with additive random error and maximize mismeasured rather than actual utility functions. A numerical example shows the subtlety of the planning problem. We then analyze binary treatment choice under uncertainty, supposing that a planner can mandate a particular treatment or can decentralize decision making, enabling variation in treatment.

^a Department of Economics and Institute for Policy Research, Northwestern University

^b Department of Economics, Hebrew University of Jerusalem.

Acknowledgements: This paper developed from research initiated in the early 2000s by Eytan Sheshinski, with the title “Socially Desirable Limits on Individual Choice.” Presentations by Sheshinski of this original work have been made in various settings, including in a seminar at the University of Bonn in 2004 and as the Richard Musgrave Lecture in Munich in April 2012. At various stages, we have benefitted from the comments of Peter Diamond, Jacob Goldin, Joram Mayshar, and Daniel Reck. We have benefitted from the opportunity to present this work in a conference at Duke University in September 2023.



"We're thinking maybe it's time you started getting some religious instruction. There's Catholic, Protestant, and Jewish—any of those sound good to you?"

1. Introduction

A central mission of research in public economics has been to determine policies that optimize utilitarian welfare, recognizing that policy choice affects individual behavior. To ease analysis, economists have maintained simplifying assumptions about behavior. Findings on optimal policy are sensitive to these assumptions.

The classic Mirrlees (1971) study of optimal income taxation assumed that individuals maximize static deterministic utility when choosing labor supply. It assumed that individuals have homogeneous consumption-leisure preferences and are heterogeneous only in ability, hence wage. Among the assumptions that Mirrlees posed in his introductory section, he stated (p. 176): "The State is supposed to have perfect information about the individuals in the economy, their utilities and, consequently, their actions." His analysis showed that the optimal tax structure is sensitive to the assumed utility function and ability distribution. In his conclusion 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."

The ensuing literature on optimal taxation has studied settings where individuals may have jointly heterogeneous preferences and abilities, it being assumed that the planner knows the joint distribution of preferences and ability. It has long been recognized that optimal policy is sensitive to the form of this distribution (e.g., Sheshinski, 1972; Atkinson and Stiglitz, 1980). However, empirical understanding of the actual population distribution of preferences and abilities has remained weak, impeding application of the theory. Manski (2014a) has written (p. 146): “As I see it, we lack the knowledge of preferences necessary to credibly evaluate income tax policies.”

Theoretical study of utilitarian policy choice began in the 1700s, was formalized in the first half of the 1900s, and continued to develop steadily through the latter part of the century, but the subject has received relatively little attention recently. A welcome exception to the recent dearth of research is a new body of analysis of optimization of welfare in populations with bounded rationality. Behavioral economists have suggested that social 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 “libertarian paternalism.” However, here and in their influential work on “nudges” (Thaler and Sunstein, 2008), their discussion has been verbal and casual, rather than formal and careful. A consensus report by a National Academies committee on *Policy Impact and Future Directions for Behavioral Economics* is similarly verbal and casual (National Academies of Sciences, Engineering, and Medicine, 2023).

An early expression of the type of formal analysis that we think desirable 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. Studies that contemplate a population of heterogeneous agents include O'Donoghue and Rabin (2006), Goldin and Lawson (2106), Rees-Jones and Taubinsky (2018), Handel, Kolstad, and Spinnewijn (2019), Moser and Olea de Souza e Silva (2019), Farhi and Gabaix (2020), and Goldin and Reck (2022).

Authors often observe that their findings on optimal policy are sensitive to the assumed population distribution of preferences and deviations from utility maximization. This sensitivity should not be surprising. After all, Mirrlees (1971) and other early studies assuming complete rationality found that conclusions are sensitive to population preferences. Broadening analysis to consider bounded rationality adds a further dimension to behavior that yields even more sensitivity. Recent authors sometimes caution that the sensitivity of optimal policy makes it important to have a firm empirical understanding of behavior in populations with bounded rationality. However, econometric research on identification of structural choice models shows that a firm empirical understanding of behavior is difficult to achieve. See Manski (2007, Chapters 13 through 15, 2014a) and Molinari (2020).

In this paper, we use a simple yet broadly applicable framework to enhance understanding of the sensitivity of optimal policy to population preferences and behavior. We consider a social planner who has the power to design a discrete choice set from which individuals will choose. We suppose that there is no social cost to offering larger choice sets. Hence, classical utilitarian welfare economics recommends that the planner should offer the largest choice set possible. We depart from the classical setting by supposing that individuals may be boundedly rational and, hence, may not choose options that maximize utility. In

such settings, it may be optimal for the planner to constrain the choice set to prevent persons from choosing inferior actions or, less drastically, to frame the choice set in a manner that influences behavior.

Example: Some public social security systems and private pensions have an early eligibility age at which a person can start receiving a pension, with less than full benefits. This age differs widely across countries. In the US, partial benefits are obtainable at age 62 and full benefits later (historically at age 65, in process of advancing to 67). The UK has had a single State Pension Age determining eligibility for full benefits (historically at age 65, in process of advancing to 67), with no option of earlier retirement with lower benefits. Imposing a constraint on the earliest age for eligibility hurts workers who would sensibly stop working before this age due to health and other personal circumstances. On the other hand, it prevents people from retiring too early, reflecting shortsightedness. Setting an early eligibility age should strike a balance between these considerations. ■

The planner's problem is straightforward if all members of the population have the same known preferences. Then the optimal paternalistic policy calls on the planner to determine the population-wide best option and mandate it. This obvious result holds regardless of the nature of bounded rationality in the population.

Our concern is more realistic settings in which persons have heterogeneous preferences and, additionally, may vary in how their choices deviate from utility maximization. We find that optimal paternalism is subtle in these settings. The policy that most effectively constrains or influences individual choice depends both on the distribution of preferences in the population and on the choice probabilities measuring the extent to which persons behave suboptimally, conditional on their preferences.

We show how the interaction of population preferences and choice probabilities determines the utilitarian welfare of policies that seek to ameliorate bounded rationality. The prevailing practice in empirical research in behavioral economics has been to perform experiments that present various options to subjects, framed in particular ways, and observe the choices that subjects make. Research of this type

usually does not seek to measure the preferences of subjects and, hence, cannot study the interaction of choices and preferences, which is essential to evaluation of utilitarian welfare. Section 2 formalizes this interaction in a concise, abstract manner. Our analysis covers both mandates that constrain choice and nudges that aim to influence choice.

Section 3 considers policy choice when individuals are boundedly rational in a specific way, this being that they measure utility with additive random error and maximize mismeasured rather than actual utility functions. Studying this type of bounded rationality yields more detailed analysis and enables presentation of an instructive numerical example. When the errors in utility mismeasurement are independent and identically distributed, we obtain a lower bound on the welfare achieved by a policy that constrains the choice set. When the errors have the Type I extreme value distribution, choice probabilities have the multinomial logit form. When the scale parameter of the error distribution is invariant across policies, this parameter succinctly characterizes the degree of rationality in the population. Our numerical example shows that the optimal paternalistic policy varies in a subtle way with the degree of rationality. Much of the analysis in this section originated in work of one of us initiated in the early 2000s that was presented in several seminars and distributed in slides, but not circulated as a working paper (Sheshinski, 2012).

Section 4 analyzes an important class of settings in which a planner can either mandate that members of a population receive a particular treatment or can decentralize decision making, enabling variation in treatment. The applied contexts in which this policy choice arises range from medical treatment to school tracking to timing of pension eligibility. A central feature of our analysis is to suppose that members of the population have publicly and privately observable covariates that may be used to condition mandates and decentralized decisions. The planner sees only the publicly observable covariates, whereas decentralized decision makers also see the privately observable ones. In these settings, utilitarian theory assuming complete rationality recommends decentralization to enable decision making to condition on more covariate information. However, some decentralized decision makers with bounded rationality may make sub-optimal decisions. We show that mandates improve utilitarian welfare if sub-optimal decentralized decision are sufficiently common.

We apply the analysis to medical treatment. Here, the planning entities are panels that formulate clinical practice guidelines. Guideline panels make treatment recommendations that act as quasi-mandates. Clinicians commonly observe patient covariates beyond those considered in guideline recommendations, enabling further personalization of treatment. Utilitarian theory assuming that both guideline panels and clinicians act with complete rationality implies that decentralized treatment is preferable to mandates. We show that this conclusion may not hold if guideline panels make optimal recommendations conditioning on the covariates they observe, but some clinicians do not.

The sensitivity of optimal policy to the fine structure of preferences and bounded rationality, about which little is known empirically, motivates us in the concluding Section 5 to caution against premature implementation of policies that attempt to ameliorate bounded rationality by constraining or influencing choice behavior. We recommend study of planning under ambiguity, enabling reasonable policy choice with credible partial knowledge of population preferences and behavior.

2. Policy Choice in Abstraction

2.1. Setup and Findings

Let J denote the population of concern to a utilitarian planner. Let C denote a pre-specified largest feasible finite choice set that the planner may make available to each member of J . Let each individual $j \in J$ have a utility function $u_j(\cdot): C \rightarrow \mathbb{R}$ expressing the person's preferences. Let $u_j^* \equiv \max_{c \in C} u_j(c)$.

To formalize utilitarian welfare, consider J to be a probability space with distribution $P(j)$ and let $P[u(\cdot)]$ denote the population distribution of utility functions. Let utility functions be interpersonally comparable. Then the idealized optimum utilitarian welfare, if all persons maximize utility, is $E(u^*)$.

The problem is that, having bounded rationality, individuals may not maximize utility. Let the planner choose among a set S of policies that may constrain or influence choice behavior. Suppose that, with policy

s, person j chooses $c_j(s) \in C$. For each $i \in C$, let $P[c(s) = i|u(\cdot)]$ denote the fraction of persons with utility function $u(\cdot)$ who would choose option i under policy s . The utilitarian welfare achieved by this policy is

$$(1) \quad E\{u[c(s)]\} = \int \sum_{i \in C} u(i) \cdot P[c(s) = i|u(\cdot)] dP[u(\cdot)].$$

The optimal feasible welfare is achieved by a policy that solves the problem $\max_{s \in S} E\{u[c(s)]\}$.

Observe that the value of $E\{u[c(s)]\}$ depends on both the preference distribution $P[u(\cdot)]$ and the conditional choice probabilities $P[c(s)|u(\cdot)]$. It is revealing to consider the regret of a policy, its degree of sub-optimality, relative to the idealized optimum utilitarian welfare $E(u^*)$. The regret of policy s is

$$(2) \quad E(u^*) - E\{u[c(s)]\} = \int \sum_{i \in C} [u^* - u(i)] P[c(s) = i|u(\cdot)] dP[u(\cdot)].$$

For each utility function $u(\cdot)$ and alternative i , $[u^* - u(i)] P[c(s) = i|u(\cdot)]$ is the degree of sub-optimality of i multiplied by its choice probability. Thus, the regret of policy s is a weighted average of the multiplicative interactions of choice probabilities for alternatives and their degrees of sub-optimality. The specific cognitive processes that lead individuals to deviate from utility maximization are immaterial. All that matters for social welfare is the set of resulting choice probabilities.

2.2. Mandates and Nudges

In the Introduction, we distinguished between policies that constrain and influence individual choices. Both types are encompassed in the above general setup. A choice-constraining policy limits the effective choice set to some $C(s) \subset C$, implying that $P[c(s) = i|u(\cdot)] = 0$ for all $i \notin C(s)$ and all $u(\cdot)$. Such a policy is a mandate if $C(s)$ is singleton. Mandating alternative i yields welfare $E[u(i)]$. Hence, the optimal mandate is to an alternative i^m that solves the problem $\max_{i \in C} E[u(i)]$.

The optimal mandate yields the idealized optimal utilitarian welfare if all members of the population have the same preferences. It yields lower welfare when preferences are heterogeneous. This holds by Jensen's Inequality, which implies that $\max_{i \in C} E[u(i)] \leq E[\max_{i \in C} u(i)]$, the inequality being strict when preferences are heterogeneous. The optimal mandate is best for persons who most prefer alternative i^m , but not for those with other preferences.

A choice-influencing policy enhances the prominence of a specified alternative but does not prevent persons from choosing other alternatives. Behavioral economists have sought to enhance prominence by specifying an alternative to be the "default option," by placing it first in the ordering of alternatives, by associating it with favorable images, and in other ways. Whatever mechanism is used, the objective is to increase the probability with which persons choose this alternative.

When considering choice-influencing policies, a behavioral economist may recommend enhancing the prominence of i^m , the optimal mandate. Such policies have been called nudges. In general, the impact of a nudge on choice behavior may depend not only on the manner in which the policy frames the choice set C , but also on the joint distribution of preferences and forms of bounded rationality in the population. It is impossible to evaluate nudges in abstraction. One must consider the context.

3. Policy Choice with Additive Error in Utility Measurement

In this section we consider policy choice when individuals are boundedly rational in a specific way. We assume that they measure utility with additive error and maximize mismeasured rather than actual utility functions. We do not assert that this type of bounded rationality is prevalent in actual populations. We study it because the idea is easy to understand and because it enables us to apply findings on choice probabilities developed in the literature analyzing random utility models.

3.1. Choice Probabilities Generated by Random Utility Models

Let policy s constrain choice to a choice set $C(s)$, which may be any non-empty subset of C . We assume that, under policy s , person j mismeasures the utility of each $c \in C(s)$ as $u_j(c) + \varepsilon_j(c, s)$, chooses an alternative $c_j^\#(s) \equiv \operatorname{argmax}_{c \in C(s)} u_j(c) + \varepsilon_j(c, s)$, and thus achieves utility $u[c_j^\#(s)]$. For simplicity, we assume that the conditional error distribution $P[\varepsilon(c, s), c \in C(s)|u(\cdot)]$ is continuous. This implies that $c_j^\#(s)$ is unique for almost every person j . Hence, choice probabilities are well-defined, with

$$(3) \quad P[c^\#(s) = i|u(\cdot)] = P[u(i) + \varepsilon(i, s) \geq u(c) + \varepsilon(c, s), \text{ all } c \in C(s)|u(\cdot)], \quad i \in C(s).$$

Inserting these choice probabilities into (1) yields the welfare achieved by policy s , which is

$$(4) \quad E\{u[c(s)]\} = \int \sum_{i \in C(s)} u(i) \cdot P[u(i) + \varepsilon(i, s) \geq u(c) + \varepsilon(c, s), \text{ all } c \in C(s)|u(\cdot)] dP[u(\cdot)].$$

Equation (3) provides a standard random-utility model interpretation of bounded rationality (McFadden, 1974; Manski, 1977). The values of the choice probabilities (3) are determined by the conditional error distribution $P[\varepsilon(c, s), c \in C(s)|u(\cdot)]$. In the absence of restrictions on this distribution, any choice probabilities are possible. Hence, assuming that a random utility model expresses bounded rationality does not, per se, yield restrictions on $E\{u[c(s)]\}$. Some knowledge of the error distributions is necessary.

Goldin and Reck (2022) use an additive random utility model with a particular type of error distribution to study policies that distinguish some alternative as the default option. For $i \in C$, let s_i denote a policy specifying i as the default option. Let there exist a person-varying but not alternative-varying quantity $\gamma_i \geq$

0 such that $\varepsilon_j(i, s_i) = 0$ and $\varepsilon_j(c, s_i) = -\gamma_j$ when $c \neq i$. Thus, an individual subtracts an *as-if cost* γ_j from the utility of each alternative that is not the default. They characterize utilitarian welfare in this setting.

In what follows, we consider random utility models with a different type of error distribution, which appears not to have been studied to date.

3.2. Simple Scalability

A lower bound on $E\{u[c(s)]\}$ emerges if, conditional on each utility function $u(\cdot)$, the error components $\varepsilon(c, s)$, $c \in C(s)$ are known to be independent and identically distributed (i.i.d.). We do not assume knowledge of the specific distribution. Indeed, it may vary with $u(\cdot)$. The i.i.d. assumption expresses the idea that individuals make “white-noise” errors in utility measurement. The error distribution may vary across persons j and policies s . However, errors do not vary systematically across alternatives. The i.i.d. assumption is generally not appropriate when studying policies that generate nudges, which asymmetrically influence evaluation of different choice options.

Manski (1975) showed that, when errors are conditionally i.i.d., choice probabilities are related to utility functions by a set of inequalities called *simple scalability*. For each utility function $u(\cdot)$ and alternative pair $(a, b) \in C(s) \times C(s)$,

$$(5a) \quad u(a) > u(b) \Leftrightarrow P[u(a) + \varepsilon(a, s) \geq u(a) + \varepsilon(c, s), \text{ all } c \in C(s)|u(\cdot)] \\ \geq P[u(b) + \varepsilon(b, s) \geq u(c) + \varepsilon(c, s), \text{ all } c \in C(s)|u(\cdot)],$$

$$(5b) \quad u(a) = u(b) \Leftrightarrow P[u(a) + \varepsilon(a, s) \geq u(a) + \varepsilon(c, s), \text{ all } c \in C(s)|u(\cdot)] \\ = P[u(b) + \varepsilon(b, s) \geq u(c) + \varepsilon(c, s), \text{ all } c \in C(s)|u(\cdot)].$$

Let $u^{\text{mean}}(s)$ denote unweighted mean utility in set $C(s)$; that is, $u^{\text{mean}}(s) \equiv [1/|C(s)|] \sum_{c \in C(s)} u(c)$. It follows from (5a)-(5b) that

$$(6) \quad u^{\text{mean}}(s) \leq \sum_{i \in C(s)} u(i) \cdot P[u(i) + \varepsilon(i, s) \geq u(c) + \varepsilon(c, s), \text{ all } c \in C(s) | u(\cdot)].$$

Hence,

$$(7) \quad \int u^{\text{mean}}(s) dP[u(\cdot)] \leq E\{u[c(s)]\}.$$

This bound varies across policies that constrain choice to different subsets of C . Among policies that constrain choice to the same subset of C , the bound does not vary across policies that seek to influence utility measurement in different ways.

3.3. Multinomial Logit Choice Probabilities

A substantial strengthening of the knowledge assumed above supposes that errors are i.i.d. with a type 1 extreme-value distribution, also called the Gumbel distribution. Assume that $\varepsilon(c, s)$, $c \in C(s)$ are independent and have the common distribution function $P[\varepsilon(c, s) \leq t] = \exp[-e^{-q(s)t}]$. Here $q(s)$ is a positive scaling factor whose value may vary with s . Then the conditional choice probabilities have the multinomial logit form (McFadden, 1974):

$$(8) \quad P[c^{\#}(s) = i | u(\cdot)] = P[u(i) + \varepsilon(i, s) \geq u(c) + \varepsilon(c, s), \text{ all } c \in C(s) | u(\cdot)] = e^{q(s)u(i)} / \sum_{c \in C(s)} e^{q(s)u(c)}, \quad i \in C(s).$$

Hence, the welfare achieved by policy s is

$$(9) \quad E\{u[c(s)]\} = \int \sum_{i \in C(s)} u(i) \cdot \left[e^{q(s)u(i)} / \sum_{c \in C(s)} e^{q(s)u(c)} \right] dP[u(\cdot)].$$

In this setting, $q(s)$ concisely measures the success of policy s in inducing individuals to measure utility correctly within the constrained choice set $C(s)$. Thus, $q(s)$ quantifies the *degree of rationality* in the population, under policy s . As $q(s) \rightarrow \infty$, the spread of the error distribution decreases and the choice probability for the alternative that maximizes utility increases to one. As $q(s) \rightarrow 0$, the spread of the error distribution increases and the choice probabilities for all alternatives converge to $1/|C(s)|$. For each alternative $i \in C(s)$, the partial derivative of the choice probability with respect to $q(s)$ is

$$(10) \quad \partial P[c^\#(s) = i | u(\cdot)] / \partial q(s) = P[c^\#(s) = i | u(\cdot)] \{u(i) - v[s, u(\cdot)]\},$$

where $v[s, u(\cdot)] \equiv \sum_{i \in C(s)} u(i) \cdot P[c^\#(s) = i | u(\cdot)]$ is the choice-probability-weighted expected utility for persons with utility function $u(\cdot)$. Thus, increasing $q(s)$ raises the choice probabilities of alternatives whose utility is higher than expected utility and vice-versa.

3.4. A Numerical Example Showing the Subtlety of Optimal Choice-Constraining Policy

In Section 3.3, policy s was described by two factors, the set $C(s)$ constraining individual choice and the degree of rationality $q(s)$ achieved by the policy. We now specialize further, considering a set S of policies that yield the same value of $q(s)$, now labeled q , and that differ only in their choice-constraining sets $C(s)$. A policy cannot exclude every option in C , so there exist $2^{|C|} - 1$ such policies.

The welfare yielded by policy s is

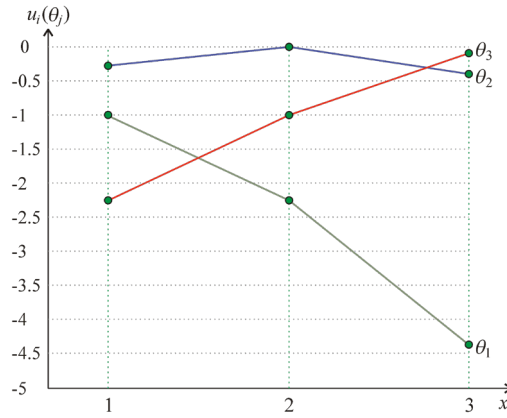
$$(10) \quad E\{u[c(s)]\} = \int \sum_{i \in C(s)} u(i) \cdot \left[e^{q u(i)} / \sum_{c \in C(s)} e^{q u(c)} \right] dP[u(\cdot)].$$

Thus, the optimal choice-constraining policy is a function of q and $P[u(\cdot)]$. Analytical determination of a policy that maximizes welfare does not seem feasible, but numerical calculation of welfare is possible given a specification of q and $P[u(\cdot)]$.

A numerical example demonstrates that optimal policy choice is subtle. The example is based on the famous Hotelling (1929) model of choice when individuals and stores are located on a line. Let C contain three alternatives (potential stores), each identified by a location x_i , $i = 1, 2, 3$ on a line. Let J contain three individuals, each residing at a location θ_j , $j = 1, 2, 3$ on this line. Let the utility of alternative i to person j be $u_i(\theta_j) = -(x_i - \theta_j)^2$. Thus, due to transportation costs, utility decreases with the distance of individual j 's location, θ_j , from store x_i . By construction, preferences are *single-peaked*.

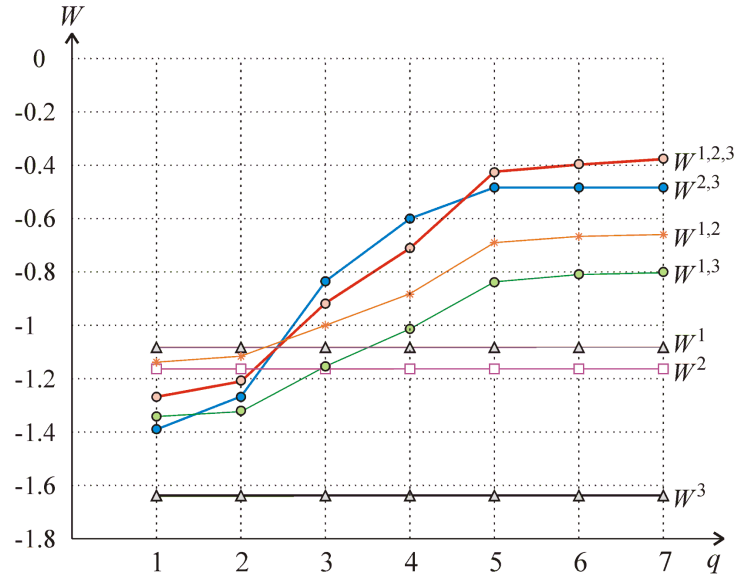
In our example, we specify $x_1 = 0.5$, $x_2 = 1$, $x_3 = 1.6$ and $\theta_1 = -0.5$, $\theta_2 = 1$, $\theta_3 = 2$. This yields the utility values shown in Figure 1:

Figure 1



The seven possible constrained choice sets are $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{2, 3\}$, $\{1, 3\}$, and $\{1, 2, 3\}$. The corresponding social welfare functions are denoted W^1 , W^2 , W^3 , $W^{1,2}$, $W^{2,3}$, $W^{1,3}$, and $W^{1,2,3}$, respectively. Figure 2 plots each of these welfare functions against different levels of q . For each q , the optimum choice-set corresponds to the *outer envelope* of these plots.

Figure 2



Observe that the optimal constrained choice set has a single alternative (W^1) at low values of q and includes all alternatives ($W^{1,2,3}$) at high values of q . Of particular interest is the fact that the ordering of welfare across choice sets is not nested, with reswitching occurring as q rises. For example, choice set $\{1, 2\}$ outperforms set $\{2, 3\}$ when q is smaller than about 2.5, but the welfare ordering reverses when q is larger. Choice set $\{1, 2, 3\}$ outperforms set $\{2, 3\}$ when q is smaller than about 2.2, the welfare ordering reverses for q between 2.2 and about 4.8, and then reverses again for q above 4.8.

This is only an example, but it suffices to demonstrate the subtlety of optimal policy choice. We find that, even in a highly simplified environment assuming multinomial logit choice probabilities, the welfare ordering of different choice-constraining policies is rather sensitive to the degree of rationality in the population. Empirical measurement of the degree of rationality would be a challenging task.

4. Mandated or Decentralized Treatment of a Population with Publicly and Privately Observed Covariates

We now analyze settings in which a planner can either mandate that members of a population receive particular treatments or can decentralize decision making. We suppose that each member of the population has publicly observable covariates $x \in X$ and privately observable covariates $z \in Z$, where X and Z are finite sets. The planner sees the publicly observable covariates, whereas decentralized decision makers also see the privately observable ones. Thus, the planner can condition a treatment mandate on x but not on z . Decentralized decisions can vary with (x, z) .

There are many contexts in which a planner chooses between a mandate and decentralized treatment. For example, a government may mandate that eligibility for a public pension begins at a particular age or may enable workers to receive a smaller benefit at a younger age. A school principal may mandate that high school students with covariates x enroll in a mathematics class taught at a specified intellectual level or can permit these students to self-select course levels. A clinical guideline panel can recommend a particular treatment for patients with covariates x , or the panel can state that physicians should use clinical judgement to choose treatments for these patients.

Utilitarian theory assuming complete rationality recommends decentralization. The basic reason is that observation of (x, z) expands the set of feasible treatment choices relative to observation of x alone. See, for example, Good (1967), Phelps and Mushlin (1988), Basu and Meltzer (2007), Manski (2007), and Kadane, Shervish, and Seidenfeld (2008). The increase in welfare achieved by observation of (x, z) relative to x is sometimes called the *value-of-information*.

We study a simple yet nuanced setting of binary treatment under uncertainty. Utilitarian theory assuming that both the planner and decentralized decision makers forecast an uncertain utility-relevant outcome with rational expectations implies that decentralized treatment is preferable to a mandate. We quantify the value of information, paraphrasing analysis in Manski, Mullahy, and Venkataramani (2023).

A mandate may yield higher utilitarian welfare if some decentralized decision makers have bounded rationality and make sub-optimal decisions. Section 4.1 presents the analysis. Section 4.2 applies the analysis to medical treatment, drawing on discussion in Manski (2018, 2019).

4.1. Analysis

Let there be two feasible treatments, labeled A and B. Treatment choice is made without knowing a utility-relevant binary outcome, $y = 1$ or 0 . For example, if treatments are mathematics courses taught at different levels, we may have $y = 1$ if a student would pass the more difficult course and $y = 0$ if the student would pass the course. If A and B are options for patient care, $y = 1$ may mean that a patient is healthy and $y = 0$ if the patient has an illness of concern.

Each person has observable covariates (x, z) , with x observable by the planner and (x, z) by decentralized decision makers. Let $p_x = p(y = 1|x)$ and $p_{xz} = p(y = 1|x, z)$ be objective probabilities that $y = 1$ conditional on x and on (x, z) . Let each value of z occur for a positive fraction of persons; thus, $P(z|x) > 0$ for all $z \in Z$. Assume that, conditional on x , p_{xz} varies with z .

Let $U_x(y, t)$ denote the expected utility that a person with covariates x would experience with treatment t , should the outcome be y . This specification assumes the absence of social interactions; that is, expected utility varies with a person's own treatment but not with the treatments received by others. It also assumes that, conditional on x , expected utility does not vary across persons with different values of z . However, z still matters to decision making because the outcome probabilities p_{xz} do vary with z . We assume that the planner knows $U_x(y, t)$ for each possible value of (x, t, y) .

Maximum utilitarian welfare using p_{xz} to predict y is always at least as large as using p_x , and it is strictly larger if optimal treatment choice varies with z . Manski, Mullahy, and Venkataramani (2023) derive a simple expression that quantifies the value of information. Section 4.1.1 summarizes the derivation. This result provides the foundation for consideration of planning with bounded rationality in Section 4.1.2.

4.1.1. Optimal Treatment Choice with Complete Rationality

With x observable but not z , the optimal mandate by a utilitarian planner is

$$(11a) \text{ choose treatment A if } p_x \cdot U_x(1, A) + (1 - p_x) \cdot U_x(0, A) \geq p_x \cdot U_x(1, B) + (1 - p_x) \cdot U_x(0, B),$$

$$(11b) \text{ choose treatment B if } p_x \cdot U_x(1, A) + (1 - p_x) \cdot U_x(0, A) \leq p_x \cdot U_x(1, B) + (1 - p_x) \cdot U_x(0, B).$$

With (x, z) observable, optimal decentralized treatment is

$$(12a) \text{ choose treatment A if } p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A) \geq p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B),$$

$$(12b) \text{ choose treatment B if } p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A) \leq p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B).$$

With criterion (11), the maximized welfare for persons with covariates x is

$$(13) \quad \max [p_x \cdot U_x(1, A) + (1 - p_x) \cdot U_x(0, A), p_x \cdot U_x(1, B) + (1 - p_x) \cdot U_x(0, B)].$$

With criterion (12), the maximized welfare for patients with covariates (x, z) is

$$(14) \quad \max [p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A), p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)].$$

In the latter case, the maximized welfare for persons with covariates x is the mean of (14) with respect to the distribution $P(z|x)$; that is,

$$(15) \quad E_{z|x} \{ \max [p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A), p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)] \}.$$

Jensen's inequality provides a simple proof that, conditional on x , maximum welfare using p_{zx} to predict y is at least as great as maximum welfare using p_x . Hence, decentralized decision making is preferable to mandating a treatment. However, Jensen's inequality does not quantify the extent to which criterion (12) outperforms (11). We can do this through direct comparison of the criteria.

Without loss of generality, let treatment A be optimal in (11). Let A be optimal in (12) for all $z \in Z_A$ and let Z_B be the complement of Z_A . Thus, inequality (12a) holds for $z \in Z_A$, some non-empty proper subset of Z , and does not hold for $z \in Z_B$, also a non-empty proper subset of Z . Criterion (12) yields better outcomes than (11) for persons with $z \in Z_B$ and the same outcomes as (11) for persons with $z \in Z_A$.

Use the decomposition of Z into (Z_A, Z_B) to rewrite (13) and (15) as

$$\begin{aligned}
 (16) \quad & \max [p_x \cdot U_x(1, A) + (1 - p_x) \cdot U_x(0, A), p_x \cdot U_x(1, B) + (1 - p_x) \cdot U_x(0, B)] \\
 &= p_x \cdot U_x(1, A) + (1 - p_x) \cdot U_x(0, A) \\
 &= P(z \in Z_A | x) \cdot E[p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)] | x, z \in Z_A] \\
 &\quad + P(z \in Z_B | x) \cdot E[p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)] | x, z \in Z_B].
 \end{aligned}$$

and

$$\begin{aligned}
 (17) \quad & E_{z|x} \{ \max [p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A), p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)] \} \\
 &= P(z \in Z_A | x) \cdot E[p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)] | x, z \in Z_A] \\
 &\quad + P(z \in Z_B | x) \cdot E[p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)] | x, z \in Z_B].
 \end{aligned}$$

Subtracting (16) from (17) yields

$$(18) \quad P(z \in Z_B | x) \cdot E\{ [p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)] - [p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)] | x, z \in Z_B \}.$$

The inequality $p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B) > p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)$ holds for all $z \in Z_B$. Hence, (18) is positive. This qualitative finding repeats the one obtainable using Jensen's inequality. What is new here is that (18) quantifies the extent to which criterion (12) outperforms (11). The magnitude of (18) is the product of two factors. One is the fraction $P(z \in Z_B|x)$ of persons for whom treatment B yields strictly larger expected utility than treatment A. The other is the mean gain in expected utility that criterion (12) yields for the subset Z_B of persons.

4.1.2. Optimal Treatment Choice with Bounded Rationality

As above, suppose without loss of generality that treatment A is optimal in (11). Again let A be optimal in (12) for $z \in Z_A$ and let Z_B be the complement of Z_A . However, suppose that some decentralized decision makers, having bounded rationality, do not use criterion (12) to choose treatments. They choose the worse of the two treatments rather than the better one.

For persons with covariate values (x, z) , let q_{xz} denote the choice probability for the better treatment and let $1 - q_{xz}$ be the choice probability for the worse treatment. If all of these persons have complete rationality, then $q_{xz} = 1$. If some have bounded rationality and choose sub-optimally, then $q_{xz} < 1$.

In this setting, mandating A continues to yield welfare (16). However, decentralization does not yield (17). Instead, decentralization yields this lower welfare:

$$\begin{aligned}
 (19) \quad & P(z \in Z_A|x) \cdot E\{q_{xz}[p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)] \\
 & \quad + (1 - q_{xz})[p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)] \mid x, z \in Z_A\} \\
 & + P(z \in Z_B|x) \cdot E\{q_{xz}[p_{xz} \cdot U_x(1, B) + (1 - p_{xz}) \cdot U_x(0, B)] \\
 & \quad + (1 - q_{xz})[p_{xz} \cdot U_x(1, A) + (1 - p_{xz}) \cdot U_x(0, A)] \mid x, z \in Z_B\}.
 \end{aligned}$$

Here, as in Sections 2 and 3, the best policy depends on the subtle interaction of population preferences and choice probabilities. Decentralized decisions yield higher welfare than mandating treatment A if (19) exceeds (16). The mandate is preferable if (16) exceeds (19). *Ceteris paribus*, the latter occurs if the choice probabilities $[q_{xz}, z \in Z]$ for the optimal z -specific treatments are sufficiently small.

4.2. Application to Medical Treatment

4.2.1. Clinical Practice Guidelines

Medical textbooks and training have long offered clinicians guidance in patient care. Such guidance has become institutionalized through issuance of clinical practice guidelines (CPGs). Institute of Medicine (2011) writes (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.” Although the recommendations made in CPGs are not legal mandates, clinicians often have strong incentives to comply, making adherence close to compulsory. A patient’s health insurance plan may require adherence to a CPG as a condition for reimbursement of the cost of treatment. Adherence may furnish evidence of due diligence that legally defends a clinician in the event of a malpractice claim. Adherence to guidelines provides a rationale for care decisions that might otherwise be questioned by patients, colleagues, or employers.

Adherence to a CPG cannot outperform decentralized care if guideline panels and clinicians are utilitarian and have complete rationality. If a CPG conditions its recommendations on all of the patient covariates that clinicians observe, it can do no better than reproduce clinical decisions. CPGs typically condition recommendations on a subset of the clinically observable covariates. Hence, adhering to a CPG may yield inferior welfare because the guideline does not personalize patient care to the extent possible.

Even though clinicians can personalize care beyond the capability of CPGs, the medical literature contains many commentaries exhorting clinicians to adhere to guidelines, arguing that CPGs developers have superior knowledge of treatment response than do clinicians. A prominent argument for adherence to

CPGs has been to reduce unwarranted variation in clinical practice. Institute of Medicine (2011) states (p. 26): “Trustworthy CPGs have the potential to reduce inappropriate practice variation.” Institute of Medicine (2013) states (p. 2-15): “geographic variation in spending is considered inappropriate or ‘unacceptable’ when it is caused by or results in ineffective use of treatments, as by provider failure to adhere to established clinical practice guidelines.” These and many similar quotations exemplify a widespread belief that adherence to guidelines is socially preferable to decentralized patient care.

4.2.2. Psychological Research Comparing Statistical Prediction and Clinical Judgment

A possible rationale for endorsement of CPGs by the medical establishment is a prevalent belief that guideline panels are close to complete rationality when they make recommendations, but clinicians have seriously bounded rationality. We have no basis to assess the rationality of guideline panels. However, we can cite ample evidence of bounded rationality among clinicians, specifically deviation from rational expectations. We summarize here, paraphrasing discussion in Manski (2018, 2019).

Psychological research comparing evidence-based statistical predictions with ones made by clinical judgment has concluded that the former consistently outperforms the latter when the predictions are made using the same patient covariates. The gap in performance persists even when clinical judgment uses additional covariates as predictors. This research began in the mid-twentieth century, notable early contributions including Sarbin (1943, 1944), Meehl (1954), and Goldberg (1968). To describe the conclusions of the literature, we quote the informative review article of Dawes, Faust, and Meehl (1989).

Dawes, Faust, and Meehl distinguish actuarial prediction and clinical judgment as follows (p. 1668):

“In the clinical method the decision-maker combines or processes information in her or her head. In the actuarial or statistical method the human judge is eliminated and conclusions rest solely on empirically established relations between data and the condition or event of interest.”

Comparing the two in circumstances where a clinician observes patient covariates that are not utilized in available actuarial prediction, they state (p. 1670):

“Might the clinician attain superiority if given an informational edge? For example, suppose the clinician lacks an actuarial formula for interpreting certain interview results and must choose

between an impression based on both interview and test scores and a contrary actuarial interpretation based on only the test scores. The research addressing this question has yielded consistent results Even when given an information edge, the clinical judge still fails to surpass the actuarial method; in fact, access to additional information often does nothing to close the gap between the two methods.”

Here and elsewhere, Dawes, Faust, and Meehl (1989) caution against use of clinical judgment to subjectively predict disease risk or treatment response conditional on patient covariates that are not utilized in evidence-based assessment tools or research reports. They attribute the weak performance of clinical judgment to clinician failure to adequately grasp the logic of the prediction problem.

4.2.3. Welfare Comparison of Adherence to Guidelines and Clinical Judgment

The psychological literature challenges the realism of assuming that clinicians have rational expectations. However, it does not per se imply that adherence to CPGs would yield greater welfare than decentralized decision making using clinical judgment.

One issue is that the psychological literature has not addressed all welfare-relevant aspects of clinical decisions. We showed in Section 4.1 that optimal decisions are determined by outcome probabilities and expected utilities. Psychologists have compared the accuracy of medical risk assessments made by statistical predictors and by clinicians, but they have not compared the accuracy of evaluations of patient preferences.

A second issue is that psychological research has seldom examined the accuracy of probabilistic risk assessments. It has been more common to assess point predictions. Study of the logical relationship between probabilistic and point prediction shows that data on the latter at most yields wide bounds on the former. For example, assume that a forecaster uses a symmetric loss function to translate a probabilistic risk assessment into a yes/no point prediction that a patient will develop a potential disease. Then observation that the forecaster states “yes” or “no” only implies that he judges the probability to be in the interval $[\frac{1}{2}, 1]$ or $[0, \frac{1}{2}]$ respectively. Thus, analysis of the accuracy of point predictions does not reveal much about the accuracy of statistical and clinical assessment of illness probabilities.

In light of these and other issues, psychological research does not suffice to conclude that adherence to CPGs is superior to decentralized decision making. Adherence to CPGs may be inferior to the extent that CPGs condition on fewer patient covariates than do clinicians. It may be superior to the extent that imperfect clinical judgment generates sub-optimal clinical decisions. How these opposing forces interplay depends on the specifics of the setting, as shown in Section 4.1.2.

5. Conclusion

The optimal paternalistic policy for a population with homogeneous preferences is a mandate. However, we showed in Section 2 that the optimal policy in a heterogeneous population may be complex, involving either or both a constraint on choice and a nudge to influence choice. The detailed analyses of Sections 3 and 4 showed that optimal paternalistic policy depends on the fine structure of the population distribution of preferences and bounded rationality. Thus, optimal policy may be highly context specific.

Unfortunately, detailed empirical knowledge of population distributions of preferences and bounded rationality is rare. Econometric research on identification of structural choice models shows that fine-grained empirical interpretation of behavior is difficult to achieve even when heterogeneous agents are rational (Manski, 2014a; Molinari, 2020). The identification problem is even more severe when agents may be boundedly rational. Hence, we caution against premature implementation of policies that attempt to ameliorate bounded rationality by constraining or influencing choice behavior. In the absence of firm empirical understanding of population behavior, such policies may do more harm than good.

As we see it, a utilitarian planner with limited knowledge of the population distribution of preferences and bounded rationality should not seek to optimize policy invoking assumptions that lack credibility. Instead, the planner should use a reasonable criterion for planning under ambiguity. One of us has performed analyses of this type in settings that assume complete rationality and focus on the problem of partial knowledge of the distribution of preferences and other primitives. In particular, Manski (2014b) studied choice of an income tax structure in a class of settings where the planner has partial knowledge of

population preferences and the productivity of public spending. The article analyzed the policy choices that result with several different decision criteria: maximization of subjective expected welfare, maximin, minimax-regret, or a Hurwicz criterion. Another example is a study of minimax-regret choice of climate policy under uncertainty (DeCanio, Manski, and Sanstad, 2022). Policy choice in populations with bounded rationality may similarly be studied as problems of planning under ambiguity.

References

- Atkinson, A. and J. Stiglitz (1980), *Lectures on Public Economics*, Princeton: Princeton University Press.
- Basu, A. and D. Meltzer (2007), “Value of Information on Preference Heterogeneity and Individualized Care,” *Medical Decision Making* 27, 112-27.
- Bernheim, B.D., A. Fradkin, and I. Popov (2015), “The Welfare Economics of Default Options in 401(k) Plans,” *American Economic Review*, 105, 2798-2837.
- Dawes, R., R. Faust, and P. Meehl (1989), “Clinical Versus Actuarial Judgment,” *Science*, 243, 1668-1674.
- 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>.
- Farhi, E. and X. Gabaix (2020), “Optimal Taxation with Behavioral Agents,” *American Economic Review*, 110, 298-336.
- Goldberg, L. (1968), “Simple Models or Simple Processes? Some Research on Clinical Judgments,” *American Psychologist*, 23, 483-496.
- Goldin, J. and N. Lawson (2016), “Defaults, Mandates, and Taxes: Policy Design with Active and Passive Decision-Makers,” *American Law and Economics Review*, 18, 438-462.
- Goldin, J. and D. Reck (2022), “Optimal Defaults with Normative Ambiguity,” *Review of Economics and Statistics*, 104, 17-33.
- Good, I. (1967), “On the Principle of Total Evidence,” *The British Journal for the Philosophy of Science* 17, 319-321.
- Handel, B., Kolstad, J. and J. Spinnewijn (2019), “Information Frictions and Adverse Selection: Policy Interventions in Insurance Markets,” *Review of Economics and Statistics*, 101, 326-340.
- Hotelling, H. (1929), “Stability in Competition,” *The Economic Journal*, 39, 41-57.
- Institute of Medicine (2011), *Clinical Practice Guidelines We Can Trust*, Washington, DC: National Academies Press.
- Kadane, J., M. Shervish, and T. Seidenfeld (2008), “Is Ignorance Bliss?” *Journal of Philosophy* 105, 5-36.
- Manski, C. (1975), “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 3, 205-228.
- Manski, C. (1977), “The Structure of Random Utility Models,” *Theory and Decision*, 8, 229-254.
- Manski, C. (2007), *Identification for Prediction and Decision*, Cambridge, MA: Harvard University Press.
- Manski, C. (2014a), “Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy,” *Quantitative Economics*, 5, 145-174.

- Manski, C. (2014b), “Choosing Size of Government under Ambiguity: Infrastructure Spending and Income Taxation,” *Economic Journal*, 124, 359-376.
- Manski, C. (2018), “Reasonable Patient Care under Uncertainty,” *Health Economics*, 27, 1397-1421.
- Manski, C. (2019), *Patient Care under Uncertainty*, Princeton: Princeton 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, <https://www.pnas.org/doi/10.1073/pnas.2303370120>.
- McFadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Behavior,” In P. Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press.
- Meehl, P. (1954), *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Minneapolis: University of Minnesota Press.
- Mirrlees, J. (1971), “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38, 175-208.
- Molinari, F. (2020), “Microeconometrics with Partial Identification,” *Handbook of Econometrics*, Vol. 7A, edited by S. Durlauf, L. Hansen, J. Heckman, and R. Matzkin, Amsterdam: Elsevier, 355-486.
- Moser, C. and P. Olea de Souza e Silva (2019), “Optimal Paternalistic Savings Policies,” Columbia Business School Research Paper No. 17-51.
- National Academies of Sciences, Engineering, and Medicine (2023), *Policy Impact and Future Directions for Behavioral Economics*, Washington, DC: The National Academies Press.
- O’Donoghue, T. and M. Rabin (2003), “Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes,” *American Economic Review (Papers and Proceedings)*, 93, 186-191.
- O’Donoghue, T. and M. Rabin (2006), “Optimal Sin Taxes,” *Journal of Public Economics*, 90, 1825-1849.
- Phelps, C. and A. Mushlin (1988), “Focusing Technology Assessment Using Medical Decision Theory,” *Medical Decision Making* 8, 279-289.
- Rees-Jones, A. and D. Taubinsky (2018), “Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment,” *Review of Economic Studies* 85, 2462-2496.
- Sarbin, T. (1943), “A Contribution to the Study of Actuarial and Individual Methods of Prediction,” *American Journal of Sociology*, 48, 593-602.
- Sarbin, T. (1944), “The Logic of Prediction in Psychology,” *Psychological Review*, 51, 210-228.
- Sheshinski, E. (1972), “The Optimal Linear Income-Tax,” *Review of Economic Studies*, 39, 297-302.

Sheshinki, E. (2012), "Socially Desirable Limits on Individual Choice," slides for Musgrave Lecture, Munich, April 2012,
https://scholars.huji.ac.il/sites/default/files/mseytan/files/the_musgrave_lectureapril_1012.pdf, accessed June 20, 2023.

Thaler, R. and C. Sunstein (2003), "Libertarian Paternalism," *American Economic Review (Papers and Proceedings)*, 93, 175-79.

Thaler, R. and C. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New Haven: Yale University Press.