A Welfarist Role for Nonwelfarist Rules
[*plus a preview of* The Golden Rule of Taxation]

Matthew Weinzierl*

March 18, 2019

Abstract

I propose and formalize an argument for why economists working in the welfarist normative tradition should include nonwelfarist principles in how they judge economic policy. The key idea behind this argument is that the world is too complex, and our ability to model it too limited, for us to fully trace a policy's effects on welfare. Nonwelfarist principles can be valuable to a welfarist facing this limitation if they act as informational proxies, carrying accumulated knowledge about the effects of policy that otherwise cannot be considered. This argument can be seen both as extending a familiar logic for rule utilitarianism beyond the realm of individual ethics and as a specific version of a broader argument made for centuries by theorists from Hume to Hayek. I also propose the rudiments of a method by which a welfarist can assign weights to specific nonwelfarist principles, focusing on the tradeoff between the predictable welfare costs of using such a principle and the benefits implied by its endorsement in other relevant contexts.

1 Policy analysis with limited models

In this paper, I propose and formalize an argument for why economists working in the standard welfarist normative tradition should include nonwelfarist principles in how they judge economic policy. If accepted, this paper's argument would have broad implications for the dominant modern approach to policy analysis that it challenges.

The key idea behind this argument is that the world is too complex, and our ability to model it too limited, for us to fully trace a policy's effects on welfare. This idea stands in sharp contrast to the conventional modern setting for policy evaluation, where the policy designer is able to predict (perhaps probabilistically) the complete set of a policy's consequences. That setting allows Kaplow and Shavell (2001) to derive their justly famous benchmark result: namely, "any non-welfarist method of policy assessment violates the Pareto principle." In reality, however, even the most sophisticated policy designer falls far short of this ideal.

Nonwelfarist principles can be valuable to a welfarist policy designer facing this limitation if they act as informational proxies, carrying accumulated knowledge about the effects of policy that otherwise cannot be considered.[1] The implication is clear: to select policies that maximize a welfarist objective, our methods

*Harvard Business School and NBER, Soldiers Field, Boston MA 02163, mweinzierl@hbs.edu. Thanks to Amy Finkelstein, Nathan Hendren, Louis Kaplow, Benjamin B. Lockwood, Francois Maniquet, Eric Nelson, John Roemer, Hannah Shaffer, Itai Sher, Lucas Stanczyk, Stefanie Stantcheva, and Glen Weyl for insightful discussions and to Jiafeng Chen and Molly Wharton for both exceptional comments and research assistance.

[1] Kaplow and Shavell (2001) note this possibility, and they also discuss it at a number of places in and Kaplow and Shavell (2009), but in practice it has been ignored in optimal policy analysis.

of policy assessment may at times, and perhaps regularly, need to include nonwelfarist principles. Toward the end of this paper, I briefly describe three real-world examples of the influence of nonwelfarist principles on policy–the resistance to envy-based redistribution, the insistence on horizontal equity, and the use of Smith's Classical Benefit-Based Taxation principle for tax design–to illustrate the practical relevance of this conclusion.

The consequences of ignoring the information that may be found in such nonwelfarist principles are both predictable and (even from the perspective of a committed welfarist) potentially problematic. As I and others have discussed elsewhere, especially with regard to optimal tax theory in the tradition of Mirrlees (1971), the standard purely welfarist approach yields policy recommendations that appear to be at odds with prevailing–and not obviously unreasonable–public priorities for policy, priorities that in turn appear to be consistent with historically prominent nonwelfarist principles.[2] By proposing a mechanism through which using such nonwelfarist principles may further, rather than displace, welfarist goals, this paper hopes to engage researchers who wish to retain the advantages of the welfarist approach while avoiding these uncomfortable consequences.[3]

This paper also lays out the rudiments of a method by which a welfarist can rigorously assign weights to specific nonwelfarist principles. No matter its conceptual and empirical appeal, the idea of using nonwelfarist principles understandably generates resistance. After all, we might worry that any departure from the rigor of direct welfarist calculations would leave behind our ability to discern among potential policy objectives. The method I propose focuses on the tradeoff between the predictable welfare costs of using a nonwelfarist principle and the benefits implied by its endorsement in other relevant contexts (where policymakers would have predicted direct welfare costs, as well). In brief, the weight given to such a principle should decrease in its predicted direct costs but increase in the weight it was given in those other contexts, and it should respond to changes in our information about those two factors. Having a method such as this allows us to discern among nonwelfarist principles, identifying those most likely to produce welfarist gains.

As I describe below, the argument of this paper connects to two important precursors that bolster its claim to being given careful consideration. First, it extends the reach of a logic for so-called rule utilitarianism (see Mill, 1871 among others) from the the realm of individual ethics to society's policymaking decisions. Second, it is a specific version of a more general claim (see Hume 1638, among others) that policy design ought to be influenced by principles reflecting society's accumulated wisdom.

It may be important to clarify that this paper is advocating neither an unthinking policy conservativism nor a wholesale retreat from rigorous cost-benefit analysis. Many of the nonwelfarist principles that might meaningfully contribute to a more comprehensive method of policy evaluation are by nature progressive. And those few–likely very few–nonwelfarist principles whose value could be convincingly demonstrated to welfarists would complement, rather than replace, the dominant methodology that currently ignores them and their embedded wisdom.

[2]Examples include the discussions of the theoretical case for height taxation in Mankiw and Weinzierl (2010), attitudes toward the proper response to innate luck in Weinzierl (2016), implied social welfare weights that depend not just on outcomes in Saez and Stantcheva (2016), a historical role for the principle of equal sacrifice in forming U.S. tax policy in Scheve and Stasavage (2016), and–in a companion article to this paper–popular and scholarly hesitancy toward envy-based redistribution in Weinzierl (2017b)

[3]A more straightforward, though surely controversial, reaction to the prevalence of nonwelfarist reasoning would be to abandon the purely welfarist approach and incorporate nonwelfarist principles *for their own sake* into the method of policy evaluation. I have explored this option for optimal tax theory with regard to the Classical Benefit-Based principle of Smith (1776 (1991)) in Weinzierl (2017a) and the Equal Sacrifice principle of Mill (1871 (1994)) in Weinzierl (2014).

2 A formalization of a welfarist role for nonwelfarist rules

In this section, I provide a formal statement of this paper's main argument.

2.1 Full information

Start by considering the standard setting where the policy designer can foresee (perhaps probabilistically) all welfare effects of policy. Notation follows Kaplow and Shavell (2001).

Let x denote a vector of length n of data characterizing the state of the world: that is, a comprehensive account of anything that might be relevant to policy judgments, including but not limited to aspects of the economic status of individuals. For simplicity, I will refer to x as the state. Let X be the set of all possible states. The function $F : X \to \mathbb{R}$ is what I will call the social objective function and is used to make policy judgments. The function $U_i : X \to \mathbb{R}$ gives the welfare of individual i, though welfare is not directly observable. A social objective function will be called welfarist if it depends on the state x through individual welfare levels only: that is if it takes the form $F(x) = W(U_1(x), U_2(x), ..., U_I(x)) \equiv W(U(x))$. Assume that, for all i and U_i, $\frac{\partial W}{\partial U_i} \geq 0$. Note that a social objective function may depend on the state in other, i.e., nonwelfarist, ways. Define the Pareto social preference relation $\succ_P$ over two states $x^s, x^t \in X$ as usual: $x^s \succ_P x^t$ iff $U_i(x^s) \geq U_i(x^t)$ for all i and $U_i(x^s) > U_i(x^t)$ for some i.

Consider two states, $x^1, x^2 \in X$, for which

$$x^1 \succ_P x^2. \tag{1}$$

Then,

$$W\left(U\left(x^1\right)\right) > W\left(U\left(x^2\right)\right), \tag{2}$$

and a policy designer choosing between states based on maximizing a welfarist social objective will choose the Pareto-preferred state, x^1.

2.2 Limited information

Suppose, instead, that the policy designer understands the welfare implications of only the first $m < n$ elements of x. For simplicity, I will denote this vector of data χ and refer to it as the observable state. For each state x^s, I will use χ^s to denote its associated observable state. Specifically, the elements of x^s not in χ^s are aspects of the state of the world that may affect individual welfare, but the policy designer is either unaware of those effects, unable to model them, or in some other way cannot form beliefs about how they affect individual welfare.[4] Recall that $U(x)$ is unobservable, so the designer cannot directly estimate these effects. Let Ξ be the set of all χ. Denote the individual welfare function defined on this limited information $V : \Xi \to \mathbb{R}$ (again, the values of V are unobservable). Retain the same definitions and notation for the social objective function and Pareto social preference relation as above, but now $F : \Xi \to \mathbb{R}$, the welfare arguments of any F are the functions V, and $\succ_P$ is defined using the functions V over observable states $\chi \in \Xi$.

[4] A simple example of what might separate χ and x is the planner's inability to model a policy's effects into the far future.

2.2.1 Misleading observable states

Consider the specific but plausible case in which the observable states are what I will call *misleading*. For example, consider the case in which, for the pairs $\left(\chi^1, x^1\right)$ and $\left(\chi^2, x^2\right)$,

$$x^1 \succ_P x^2, \text{ but } \chi^1 \prec_P \chi^2. \tag{3}$$

That is, while the superiority of state x^1 is clear under full information, under limited information the Pareto-ranking of the states reverses and x^2 appears, through χ^2, to be the better choice. Then,

$$W\left(U\left(x^1\right)\right) > W\left(U\left(x^2\right)\right), \text{ but } W\left(V\left(\chi^1\right)\right) < W\left(V\left(\chi^2\right)\right). \tag{4}$$

In words, with misleading states, the policy designer choosing between states based on a welfarist social objective will mistakenly choose x^2 (by choosing χ^2) even though x^1 is Pareto preferred.

2.2.2 A welfarist role for nonwelfarist rules

In this case, information from the observable states χ^1 and χ^2 may help the policy designer avoid this mistake. Consider an element α of χ that takes different values in states x^1 and x^2. A social objective function may depend on α, and in so doing capture omitted, welfare-relevant information about the state x.

In particular, define a social objective function $\hat{F}$ such that, despite the misleading information,

$$\hat{F}\left(V\left(\chi^1\right), \alpha^1\right) > \hat{F}\left(V\left(\chi^2\right), \alpha^2\right). \tag{5}$$

That is, if the policy designer uses this nonwelfarist (or more precisely, not purely welfarist) social objective function, state x^1 will be preferred, consistent with the Pareto criterion. The intuition for this result is that α carries relevant information about x that is understandable even when other, welfare-relevant components of x are not.

Before going on, two notes of clarification on the meaning of (5) may be helpful. First, V cannot depend directly on α. By assumption, the planner cannot know how the omitted welfare-relevant aspects of the state directly affect utilities. That assumed limitation is why $\hat{F}$ gives weight to the proxy α directly.[5] Second, a Bayesian learning approach to specifying the social objective function cannot resuscitate pure welfarism in this setting. One might imagine the planner could make a best guess about how factors such as α affect individual utilities, refine those guesses as information is received, and thereby retain a welfarist approach while accommodating limited information. Such a learning process would make the α factors–which by definition serve a welfarist end–unnecessary if it could converge on the full-information state. I am making the epistemological assumption, therefore, that such a learning process cannot converge in this way.[6]

[5] A different case, which I do not consider, is if individuals internalize a direct concern for the information in α, in which case α would be an appropriate argument of V. The policy designer will nevertheless retain an incentive to put weight on α in $\hat{F}$ in that case, as the individuals' internalized concerns for α may fail to match society's optimal consideration of α. For a related discussion, see Kaplow (2008), p. 358.

[6] This discussion relates closely to R.M. Hare's idea of two-level thinking described below and to discussions of model misspecification in Hansen and Sargent (2001) and ignorance in Hansen and Sargent (2015). My assumption that planners cannot learn how factors such as α affect welfare could be justified in a number of ways. One option, which I do not prefer, is that such knowledge is somehow unobtainable *in principle*, that is even with infinite observations on the behavior of the economy in response to the planner's choices. Such an extreme case may be true, but it is unnecessary. I prefer the justification that such learning is impossible *in practice*, given the complexity of society, the dynamic nature of that complexity, the high dimensionality of policy actions, and the limits of econometric analysis.

2.3 Relationship to Kaplow and Shavell (2001)

The preceding analysis relates to Kaplow and Shavell's in an instructive way, in that it shows how limited information can create a situation in which a welfarist policy designer correctly (from its own perspective) using a nonwelfarist social objective *appears* to be violating the Pareto criterion.

Consider the pair $\left(\chi^0, x^0\right)$, where $\chi^0 \in \Xi$ and $x^0 \in X$. Suppose that, for all i,

$$V_i\left(\chi^1\right) = V_i\left(\chi^0\right) \tag{6}$$

but that the same social objective function as above, $\hat{F}\left(V\left(\chi\right), \alpha\right)$, yields:

$$\hat{F}\left(V\left(\chi^1\right), \alpha^1\right) > \hat{F}\left(V\left(\chi^0\right), \alpha^0\right). \tag{7}$$

In words, the observable welfare implications of the observable states χ^1 and χ^0 are the same, but the nonwelfarist $\hat{F}$ prefers χ^1 because of its value of α.

As Kaplow and Shavell show, the nonwelfarist nature of $\hat{F}\left(V\left(\chi\right), \alpha\right)$ may make it appear as though a policy designer using it is violating the Pareto criterion in this case. To see why, suppose that the state x^2, the same state as in the limited information case above, relates to x^0 in a specific way. Namely, all individuals have slightly more of a component x_k of x that they positively and continuously value in x^2 than in x^0 (i.e., $x_k^2 = x_k^0 + \delta$ for $\delta > 0$), but all other aspects of x^2 and x^0 are unchanged.[7] Assume $k < m$, so that x_k is included in the observable state χ. Then,

$$\chi^1 \prec_P \chi^2, \tag{8}$$

because for some $\delta > 0$ the observable state χ^2 yields greater welfare for all individuals than does χ^1. Recall, however, from the limited information case above, condition (5) relating these two observable states:

$$\hat{F}\left(V\left(\chi^1\right), \alpha^1\right) > \hat{F}\left(V\left(\chi^2\right), \alpha^2\right). \tag{9}$$

Kaplow and Shavell, in the full-information analogue to this example, conclude from conditions (8) and (9) that the nonwelfarist social objective function $\hat{F}$ causes the policy designer using it to violate the Pareto criterion (i.e., by choosing χ^1 over χ^2).

In this limited information setting, however, we know from condition (3) that:

$$x^1 \succ_P x^2, \tag{10}$$

because components of the states not included in the observable states make x^1 Pareto-preferred to x^2 (and x^0). That is, the observable states χ^0, χ^1, and χ^2 are misleading.

This example shows, then, that the conclusion opposite to Kaplow and Shavell's may be merited: if a nonwelfarist function takes into account otherwise unobservable, welfare-relevant information, it can help to *ensure* that the Pareto criterion is respected.

[7] I thus adopt Kaplow and Shavell's two mild technical assumptions.

3 Determining the weight on nonwelfarist rules

In this section, I suggest a method by which to determine and–critically–adjust the weights given to nonwelfarist rules. After all, it is likely that only some, even among those that have had sufficient appeal to become prominent in everyday discourse, can be relied upon to yield welfarist gains. Changes in society or our understanding of it will also change which (and to what extent) nonwelfarist rules are appealing components of our objective.

The core tradeoff captured by this method is between the direct, predictable welfare costs of using a nonwelfarist rule and the implicit benefits implied by its use in other contexts such as the past or other jurisdictions (where, importantly, predictable welfare costs are also considered by policymakers). The formalization relies on the notation used above.

Suppose, for simplicity, that the social objective function introduced in (5) is quasilinear:

$$\hat{F}\left(V\left(\chi\right),\alpha\right)=V\left(\chi\right)+\beta_{\alpha}\hat{f}\left(\alpha\right),$$

so that the value of β_{α} summarizes the weight society gives to the nonwelfarist component of the objective that depends on α, i.e., the "sub-objective" function $\hat{f}\left(\alpha\right)$. Specifically, if $\beta_{\alpha}=0$ the social objective function is the welfarist (utilitarian) benchmark.

The next step of the method is to specify how we determine the value of β_{α}. Consider the following formulation:

$$\beta_{\alpha}=B\left(\left[V^{*}\left(\chi|\beta_{\alpha}=0\right)-V^{*}\left(\chi|\beta_{\alpha}\right)\right],\eta_{\alpha}\right). \tag{11}$$

In (11), $V^{*}\left(\chi|\beta_{\alpha}=0\right)-V^{*}\left(\chi|\beta_{\alpha}\right)$ is the predicted loss in (maximized) aggregate welfare from giving weight β_{α} rather than 0 to the nonwelfarist component of the social objective function that depends on α. Note that this predicted loss depends on the observable state. The parameter η_{α} is the perceived "value" put on that component in whatever the relevant context being considered (for example, in the past or in other jurisdictions).

Assuming that the function $B\left(\cdot\right)$ is decreasing in its first argument and increasing in its second, then this functional form for β_{α} captures the central tradeoff in using nonwelfarist principles described above: a nonwelfarist rule receives less weight in the social objective function the greater are the estimable (direct) utility costs of its inclusion but more weight the greater is the value society places on it.

Applying this method across time or space substantially facilitates its practicality. For example, suppose we add time-dependence to the specification above. In particular, the value for β_{α} in period $t-1$ is:

$$\beta_{\alpha,t-1}=B\left(\left[V^{*}\left(\chi_{t-1}|\beta_{\alpha,t-1}=0\right)-V^{*}\left(\chi_{t-1}|\beta_{\alpha,t-1}\right)\right],\eta_{\alpha,t-1}\right). \tag{12}$$

Assuming we have the right model of policymaking, we can use the inverse-optimum technique in which prevailing policy choices imply parameter values (see Bourguignon and Spadaro (2012) and Lockwood and Weinzierl (2016)). In particular, we use chosen policy to estimate $\beta_{\alpha,t-1}$, assume a functional form for $B\left(\cdot\right)$, and predict $V^{*}\left(\chi_{t-1}|\beta_{\alpha,t-1}=0\right)-V^{*}\left(\chi_{t-1}|\beta_{\alpha,t-1}\right)$, leaving us with an inferred value of $\eta_{\alpha,t-1}$. Then, at time t, we can reverse the process. That is, we assume the same $B\left(\cdot\right)$, calculate $V^{*}\left(\chi_{t}|\beta_{\alpha,t}=0\right)-V^{*}\left(\chi_{t}|\beta_{\alpha,t}\right)_{i}$, and predict $\eta_{\alpha,t}$, perhaps with $\eta_{\alpha,t-1}$ as the baseline value, to yield a value for $\beta_{\alpha,t}$.

Note that this process delivers some rigor in the determination of η_{α}, a measure of social "value" that is likely hard to elicit directly, by inferring its value in a relevant context from policy choices. To the extent that we can choose a context similar to our own–that is, where our functional form assumptions are likely

to remain applicable–this inferred value of η_α will carry important information.

This method also has the advantage of clarifying how we would adjust the weights on principles to new information or preferences. Information can make its way into the calculation of β_α through either the estimated direct welfare costs of using the nonwelfarist principle or the estimated social value put on that principle. If the inference exercise uses past policymaking, for example, and we acquire new information that increases our prediction of the welfare costs of the principle (versus what policymakers were predicting before), the first argument of the function determining the current weight will increase, and β_α will decrease. Or, suppose an evolution of social norms changes preferences toward the principle, such that we can be confident η_α has fallen. This update will also cause the weight β_a to decrease through a decrease in the second argument of (11).[8]

Finally, note that we might apply this process not across time but across jurisdictions. Reinterpreting the subscripts to indicate place rather than period, we could estimate

$$\beta_{\alpha,s} = B\left(\left[V\left(\chi_s|\beta_{\alpha,s}=0\right) - V\left(\chi_s|\beta_{\alpha,s}\right)\right], \eta_{\alpha,s}\right) \tag{13}$$

for jurisdiction s and use it to predict values for our jurisdiction t, modifying our estimates of the arguments of $B(\cdot)$ accordingly.

Of course, this method has weaknesses and I have proposed only the rudiments of it. It requires specifying a number of abstract functional forms, identifying the nonwelfarist principles to analyze, and estimating a number of not directly observable parameters. Nevertheless, it provides a structure by which to work through the central tension in giving weight to these principles.

4 Two intellectual precursors

Here, I connect this paper's argument to two prominent antecedents with closely related justifications.

4.1 An informational logic for rule utilitarianism

A vast philosophical literature on rule utilitarianism establishes the importance of nonutilitarian principles for utilitarian individual ethics. Moral philosophers have long recognized that individuals, even if seeking to act so as to produce the best outcomes, must sometimes act according to rules that make no direct reference to the consequences of their actions. Mill (1871 (1994) calls these nonwelfarist rules "secondary principles," and like many others he emphasizes their importance in achieving an underlying welfarist goal.[9] Harsanyi (1982) is particularly clear: "rule utilitarianism is free to choose a moral code that judges the moral value of individual actions partly in terms of nonconsequentialist criteria if use of such criteria increases social utility."

One aspect of the substantive logic for rule utilitarianism is that limits to our knowledge make it unrealistic and unwise for an individual to be an act utilitarian. Unrealistic because an individual cannot know the full consequences of his or her actions; unwise because such an information-poor method of choosing actions will give the wrong answers in many cases.[10] In a well-known example of this logic, Hare (1981) stresses

[8] Changing norms may, instead, be captured by changes in the predicted welfare costs (the first term of $B(\cdot)$.

[9] "Whatever we adopt as the fundamental principle of morality, we require subordinate principles to apply it by..." (U, II, pp. 296-297).

[10] The formalization of the previous section shows that this paper's argument does not rest on there being risky consequences of actions (clarified by Hammond (1982)), but on ignorance of, or inability to consider, a subset of consequences. My argument is also separate from the insightful discussion in Dasgupta (1982) of the optimal policy response to private information.

the role of information in requiring that moral thought proceed at two levels: "level-1, the everyday moral thinking...in which information is sparse," and "level-2...in which there is time for unlimited investigation of the facts." He imagines an "archangel [with] superhuman knowledge of the facts" practicing level-2 thinking in order to provide a moral education to children. The archangel "will try to implant in them a set of good general principles...they will use in their ordinary level-1 moral thinking." As Hare writes, "The result will be a set of general principles, constantly evolving, but on the whole stable, such that their use in moral education, including self-education, and their consequent acceptance by the society at large, will lead to the nearest possible approximation to the prescriptions of archangelic thinking."

A specific way in which individuals not using rules will miss relevant information is that the scope of direct, provably causal consequences of any individual ethical choice is inevitably narrow. Harsanyi (1977) stresses that rule utilitarianism circumvents this narrowness by taking into account not only the direct consequences of the rule's verdict on a given ethical choice but also the causal consequences of the rule being adopted in general and the "noncausal logical implications of its adoption." See Mill (1871) and Parfit (1984).

One way to view the main argument of this paper is that this informational logic for rule utilitarianism, as well as its implied role for principles that only indirectly further the underlying utilitarian goal, can be readily extended to the public sphere, justifying the use of nonwelfarist principles in fundamentally welfarist policy evaluation. Hare's archangelic level-2 thinking requires a level of knowledge–about the functioning of the economy and society more broadly–that even the most sophisticated real-world policy designers lack, at least at any point in time. And Harsanyi's emphasis on the indirect effects of adopting moral rules can be translated immediately to concerns about the indirect effects of policy choices, for example on unobservable social norms and attitudes.[11]

Though it will not be the focus of this paper, a natural related conjecture is that the secondary principles useful to individuals and those useful to society share a source: accumulated human wisdom. As Mill (1871) writes in the context of individual ethics, "...mankind must by this time have acquired positive beliefs as to the effects of some actions on their happiness; and the beliefs which have thus come down are the rules of morality for the multitude, and for the philosopher until he has succeeded in finding better."[12]

4.2 Ignorance and policy rules

This paper's argument is a specific version of a more general point made by eminent thinkers such as David Hume, Adam Smith, and Edmund Burke, who emphasized the importance of experience, not just abstract reason, in forming our moral judgments and the related importance of social institutions and principles that arose from a process of historical trial and error in forming our policy judgments.[13]

[11] As I mention elsewhere in the text, such concerns are relevant to advocates across the political spectrum. The unobservable long-term effects of progressive tax and transfer policy on social norms is a classic concern of those on the "right," while the importance of progressive tax and transfer policy to social cohesion and the legitimacy of the state in a capitalist economy is a major concern of the "left."

[12] Mill (1871) reiterates the point later: "Again, defenders of utility often find themselves called upon to reply to such objections as this–that there is not time, previous to action, for calculating and weighing the effects of any line of conduct on the general happiness...The answer to the objection is, that there has been ample time, namely, the whole past duration of the human species. During all that time, mankind have been learning by experience the tendencies of actions; on which experience all the prudence, as well as all the morality of life, are dependent."

[13] See, for example, Hume (1738 (2007), vol 2, 3.2.2); Hume (1752 (1973), p. 60); Smith (1759 (2009), 7.3.2); and Burke (1790 (2004), p. 120). Given this list of authors, it may be important to note that the logic for using nonwelfarist principles is not exclusively relevant to those on the political "right." On the contrary, the same logic may apply to those on the political "left" who support, among other things, progressive taxation. An empirical example, explained at length by Scheve and Stasavage (2016), is the use of Mill's principle of Equal Sacrifice to justify the estate tax in developed economies throughout the twentieth century. Also see the work of Roemer (1998) that emphasizes and derives the policy implications of a commitment to equality

Perhaps the most prominent modern development of that point is due to Friedrich A. Hayek. In Hayek (1960), he expresses a deep skepticism of the "idea of intelligent men coming together for deliberation about how to make the world anew." Instead, he endorses the idea that "civilization was the accumulated hard-earned result of trial and error; that it was the sum of experience, in part handed from generation to generation as explicit knowledge, but to a larger extent embodied in tools and institutions which had proved themselves superior–institutions whose significance we might discover by analysis but which will also serve men's ends without men's understanding them."

Hayek (1973) goes on to lament that political theory in his time had moved away from his arguments and adopted "constructivist rationalism" based on the belief that "all the relevant facts are known to some one mind, and that it is possible to construct from this knowledge of the particulars a desirable social order." He would no doubt raise the same concern about the standard modern approach to policy analysis.

5 Practical examples

In this section, to illustrate the practical relevance of this paper's argument, I briefly describe three examples of the influence of nonwelfarist rules on policy.[14]

5.1 Resistance to envy-based redistribution

The existence of envy has long posed a puzzle for tax theorists. On one hand, envy can be interpreted as the imposition of externalities on society by high-income earners, in which case taxation that reduces their earnings may be justified on efficiency grounds. On the other hand, most tax theorists–some instrumental in the rise of welfarism–are uncomfortable with the idea of accommodating envy through redistributive taxation and argue against it on indirectly welfarist grounds. Harsanyi (1982), arguably the economist most responsible for the utilitarian foundation of modern public economics, emphasized that "the basis of utilitarianism is benevolence" so that envy, as one of the "clearly antisocial preferences," should not be a part of welfarist calculations. But Kaplow (2008) gives the justification closest to that elaborated in this paper: "In sum, it seems difficult to articulate the actual meaning of preference censoring or to identify a convincing rationale for ignoring, as a matter of first principles, certain sorts of preferences. Nevertheless, it may be good social policy to set aside certain negative preferences, such as envy."

This philosophical debate has real-world parallels. In Weinzierl (2018), I show survey evidence that a majority of American survey respondents share this discomfort, resisting welfare gains based on taxing the envied. Consistent with these results, Mitt Romney, the Republican nominee for U.S. President in 2012, made the following comments on some of President Barack Obama's statements: "You know, I think it's about envy. I think it's about class warfare. When you have a President encouraging the idea of dividing America based on the 99 percent versus 1 percent—and those people who have been most successful will be in the 1 percent—you have opened up a whole new wave of approach [sic] in this country... The American people, I believe in the final analysis, will reject it." While Romney lost the election, his skepticism toward policy based on accommodating envy is shared by many. Perhaps even more important, those who criticized his rhetoric did so by (quite defensibly) rejecting the idea that envy was the root of Americans' dissatisfaction

of opportunity, a nonwelfarist (or at least indirectly welfarist) priority.

[14] A fourth potential, though debatable, example is the widespread enthusiasm for equal opportunity, which is especially striking because that principle is *defined* by its focus on something other than the outcomes that a welfarist typically views as the only things that matter for policy evaluation.

with economic inequality, not by embracing the idea of envy as a justification for redistributive taxation (see Yglesias (2012)).

5.2 Insistence on horizontal equity

Horizontal equity–the "equal treatment of equals"–is one of the core principles of tax design long taught to undergraduate economics students, but as Kaplow (2000) writes: "When one carefully examines the concept of HE and what its pursuit entails, one discovers there is no normative basis for deeming it to be important and, in fact, it conflicts with the basic foundations of welfare economics. That is, HE stands in opposition to the advancement of human welfare. Indeed, consistent pursuit of HE can conflict with the Pareto principle."

To highlight this tension, Mankiw and Weinzierl (2010) focus on the possible welfare gains from the taxation of height. Height is correlated with income-earning ability, and it is fixed, so the taxation of height would yield equality gains without efficiency costs and, thereby, welfare gains (more technically, height reduces the information problem facing the tax authority who cannot observe income-earning ability, so its use can only help that authority pursue its welfare-maximization objective). The apparent gains from taxing other "tags," such as race or gender, would be much greater, as I show in Weinzierl (2014).

Of course, in contrast to these welfarist analyses, no society taxes personal traits such as height, instead insisting on horizontal equity. Why? One plausible answer is that the costs of violating horizontal equity–in social cohesion, the legitimacy of the state, or other unforeseen areas–can be expected to far outweigh its apparent benefits.

5.3 Classical Benefit-Based Taxation: The Golden Rule of Taxation [preview of broader paper]

In Weinzierl (2017a) and Weinzierl (2016), I argue that Adam Smith's first maxim of taxation, what Musgrave (1959) called Classical Benefit-Based Taxation (CBBT), is an intellectually coherent rule of optimal tax design that enjoys substantial popular support. In sharp contrast, CBBT is almost entirely disregarded by modern optimal tax theory. Here, I propose an argument for CBBT based closely on the logic of this paper, namely that CBBT can be seen as a Golden Rule of Taxation–an analogy to the familiar Golden Rule of Ethics–that even committed welfarists ought to consider as part of their objective.

The Golden Rule of Ethics suffers from a paradox similar to that of CBBT: while "do unto others as you would have others do unto you" is a core element of modern common morality, it is largely dismissed by philosophers as an ethical guide. Using the terminology of Hare (1981) discussed above, the Golden Rule of Ethics is Level 1 thinking not generally used by those (professional ethicists) who practice Level 2 thinking (except, perhaps, when teaching practical ethics). If Level 2 thinking is, however, rarely possible even among the most sophisticated persons, then the Golden Rule may be more than instrumentally useful.

As an ethical rule, the Golden Rule of Ethics has at least three strengths. It decentralizes and localizes ethical judgments, eliminating the need to know all the effects of actions to judge them; it includes both sides affected by ethical judgments, ensuring a holistic view of the actions being judged; and its vagueness–which is a source of criticism among Level 2 practitioners–protects us from grave errors of judgment based on too narrow a normative framework.

If Level 2 thinking is just as unlikely in tax theory as in ethics, we might expect a Golden Rule of Taxation to be both merited and prevalent, and Smith's CBBT is a promising candidate. Smith wrote, "The subjects of every state ought to contribute toward the support of the government, as near as possible, in proportion

to their respective abilities; that is in proportion to the revenue which they respectively enjoy under the protection of the state." In other words, Smith thought people should pay taxes based on the benefit they obtain from the activities of the state (the standard benefit-based logic), and his CBBT specifies how to calculate benefit: through the magnification of innate ability. Therefore, people should pay taxes based on the income they can earn due to the activities of the state.

CBBT is a promising candidate for the Golden Rule of Taxation in part because it shares analogues to the three strengths of the Golden Rule of Ethics.

First, CBBT decentralizes and localizes tax judgments, eliminating the need to know all the effects of a given policy. Taxes are justified based on the individual's direct benefit from government, not as a way to achieve broader social objectives. If set fairly, therefore, taxes are justifiable at individual level, and benefit-based taxes' great appeal, in fact, is that they mimick the market mechanism (Lindahl (1919 (1958)).

Second, CBBT includes both sides affected by tax judgments–revenue and expenditure–ensuring a holistic view of taxation. In contrast, modern Mirrleesian optimal tax theory assumes we can separate decisions on government spending from taxation. If people judge taxes based on how revenue is used, this assumption is faulty, and given the high levels of popular frustration with governments today, this assumption may be quite damaging. CBBT explicitly links the revenue and spending side of tax policy, taxpayer by taxpayer.

Third, CBBT's vagueness protects us from grave errors based on more specific tax principles. Utilitarian tax theory's results are notoriously fragile to small parameter variation or model features, such that if a model specification is incorrect, tax policy can be substantially sub-optimal. Instead, CBBT is self-regulating, in that if benefits do not align with taxes, people will demand reform. We need not know how, in particular, dissatisfaction arises or how that dissatisfaction relates to overall (unobservable) well-being, we need only pay attention to how satisfied individuals are with the activities of the state.

Consistent with these arguments for CBBT is the statement of principle repeatedly made by President Barack Obama: "As a country that values fairness, wealthier individuals have traditionally borne a greater share of this [tax] burden than the middle class or those less fortunate. Everybody pays, but the wealthier have borne a little more. This is not because we begrudge those who've done well — we rightly celebrate their success. Instead, it's a basic reflection of our belief that those who've benefited most from our way of life can afford to give back a little bit more." Evidence suggests that Obama is not alone, with a substantial share of the public supporting CBBT as a guide to tax policy.

6 Conclusion

Though the argument for a welfarist role for nonwelfarist rules has a long historical reach, a foundation in familiar theories of individual ethics and policymaking under limited information, and apparent empirical relevance, it has been discounted in modern normative economic policy research. This paper hopes to renew interest in it and spur further work on it, including in the elaboration of the method by which to determine the weights given to nonwelfarist principles. In such work, critical tasks will include identifying (through public opinion, robust policy features, political rhetoric, and scholarly analysis) the diverse set of rules that may act as informational proxies, rigorously evaluating them to determine which justify inclusion, and developing–perhaps based on the rudiments proposed here–a formal means through which to incorporate them in our analysis.

References

Bourguignon, Francois, and Amedeo Spadaro. 2012. ""Tax-benefit revealed social preference"." *The Journal of Economic Inequality.*

Burke, Edmund. 1790 (2004). *Reflections on the Revolution in France.* Penguin.

Dasgupta, Partha. 1982. "Utilitarianism, information and rights." In *Utilitarianism and Beyond.* , ed. A. Sen and B. Williams. Cambridge.

Hammond, Peter. 1982. "Utilitarianism, uncertainty, and information." In *Utilitarianism and Beyond.* , ed. A. Sen and B. Williams. Cambridge.

Hansen, Lars Peter, and Thomas Sargent. 2001. "Acknowledging misspecification in macroeconomic theory." *Review of Economic Dynamics.*

Hansen, Lars Peter, and Thomas Sargent. 2015. "Four Types of Ignorance." *Journal of Monetary Economics.*

Hare, R. M. 1981. *Moral Thinking.* Oxford.

Harsanyi, John. 1977. "Rule Utilitarianism and Decision Theory." *Erkenntnis.*

Harsanyi, John. 1982. "Morality and the theory of rational behaviour." In *Utilitarianism and Beyond.* , ed. A. Sen and B.Williams. Cambridge.

Hayek, Friedrich. 1960. *The Constitution of Liberty.* Chicago.

Hayek, Friedrich. 1973. *Law, Legislation, and Liberty.* Chicago.

Hume, David. 1738 (2007). *A Treatise of Human Nature.* Oxford.

Hume, David. 1752 (1973). *Political Essays.* Liberal Arts Press.

Kaplow, Louis. 2000. "Horizontal equity: New measures, unclear principles." *National Bureau of Economic Research WP 7649.*

Kaplow, Louis. 2008. *The Theory of Taxation and Public Economics.* Princeton.

Kaplow, Louis, and Steven Shavell. 2001. "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle." *Journal of Political Economy.*

Kaplow, Louis, and Steven Shavell. 2009. *Fairness vs. Welfare.* Harvard.

Lindahl, Erik. 1919 (1958). "Just taxation–a positive solution, in Classics in the theory of public finance." 168–176.

Lockwood, Benjamin, and Matthew Weinzierl. 2016. "Positive and normative judgments implicit in US tax policy, and the costs of unequal growth and recessions." *Journal of Monetary Economics.*

Mankiw, N. Gregory, and Matthew Weinzierl. 2010. "The optimal taxation of height: A case study of utilitarian income redistribution." *American Economic Journal: Economic Policy.*

Mill, John Stuart. 1871 (1994). *Principles of Political Economy.* Oxford.

Mirrlees, James. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies.*

Musgrave, Richard. 1959. *The Theory of Public Finance.* McGraw-Hill.

Parfit, Derek. 1984. *Reasons and Persons.* Oxford.

Roemer, John. 1998. *Equality of Opportunity.* Harvard.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Marginal Welfare Weights for Optimal Tax Theory." *American Economic Review.*

Scheve, Kenneth, and David Stasavage. 2016. *Taxing the Rich.* Princeton.

Smith, Adam. 1759 (2009). *Theory of Moral Sentiments.* Penguin.

Smith, Adam. 1776 (1991). *Wealth of Nations.* Prometheus Books.

Weinzierl, Matthew. 2014. "The Promise of Positive Optimal Taxation: Normative Diversity and a role for Equal Sacrifice." *Journal of Public Economics.*

Weinzierl, Matthew. 2016. "Popular Acceptance of Inequality due to Brute Luck and Support for Classical Benefit-Based Taxation." *Journal of Public Economics.*

Weinzierl, Matthew. 2017a. "Revisiting the Classical View of Benefit Based Taxation." *Economic Journal.*

Weinzierl, Matthew. 2017b. "A Welfarist Role for Nonwelfarist Rules." *NBER WP 23587.*

Weinzierl, Matthew. 2018. "WelfarismâĂŹs Envy Problem Extends to Popular Judgments." *AEA P&P.*

Yglesias, Matthew. 2012. "Mitt Romney Says Concern About Inequality Is Just "Envy"." *MoneyBox.*