A Welfarist Role for Nonwelfarist Rules: An example with envy

Matthew Weinzierl*

July 7, 2017

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 provide evidence of an example in which real-world policy judgments are consistent with this theoretical argument. Results from a novel U.S. opinion survey show that approximately half of respondents reject redistribution driven by envy even though it generates direct utilitarian gains. That share rises as the role of envy is made more salient, consistent with respondents using nonwelfarist principles to encode concerns about the unobservable consequences of policy.

*Harvard Business School and NBER, mweinzierl@hbs.edu. Thanks to Louis Kaplow, Benjamin B. Lockwood, Francois Maniquet, Hannah Shaffer, Itai Sher, Lucas Stanczyk, and Stefanie Stantcheva for insightful discussions and to Jiafeng Chen and Molly Wharton for both exceptional comments and research assistance.

Introduction

In this paper, my first goal is to propose and formalize an argument for why economists working in the welfarist normative tradition should include nonwelfarist principles in how they judge economic policy. If accepted, this argument would have broad implications for the dominant modern approach to policy analysis.

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. To see why this limitation may justify the use of nonwelfarist principles, start with opposite (conventional) case in which a policy designer is able to predict a given policy's full set of consequences or, more generally, the probability distribution across full sets of consequences. This is the setting Louis Kaplow and Steven Shavell (2001) assume, following a vast related literature, and in it they derive their well-known, important, and powerful result: namely, "any non-welfarist method of policy assessment violates the Pareto principle." The intuition behind their result is that an omniscient, welfarist policy designer will unavoidably risk sacrificing welfare by giving weight to anything other than welfare, including nonwelfarist principles, in policy assessments. In reality, however, even the most sophisticated policy designer falls far short of omniscience.

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. In that case, ignoring nonwelfarist principles means ignoring useful information about welfare-maximizing policy design. The implication is clear: to select policies that maximize a welfarist objective, our methods of policy assessment may at times, and perhaps regularly, need to include nonwelfarist principles. Kaplow and Shavell (2001) themselves acknowledge[1] this possibility in principle, but modern optimal tax research ignores it in practice. In fact, Liam Murphy and Thomas Nagel (2002) praise modern tax theory in the Mirrlees (1971) tradition for its purely welfarist approach.[2]

The second goal of this paper is to provide evidence of an example in which real-world policy judgments are consistent with this theoretical argument.[3] I focus on the tax policy response to the existence of envy, a classic area of tension for welfarism. When individuals suffer psychological losses from the economic gains of others, those gains yield less welfare to society as a whole than they do to the envied. In a welfarist framework, envy thus acts as a negative externality from economic advantage, and optimal tax policy will respond with greater redistribution.[4]

Using a novel opinion survey, I show that the American public is skeptical of envy-based redistribution, even when its direct effect would be to increase welfare. More than half of survey respondents reject envy-based redistribution when presented with a hypothetical scenario in which its direct utilitarian appeal is made clear.[5] Moreover, that share rises as the role of envy is made more salient, and respondents who express more support for libertarianism–a classic alternative

[1] For example, they write: "a notion of fairness might be useful for policymaking if it serves as a good proxy principle for raising utilities when direct assessment of policies cannot be undertaken."

[2] See also the extensive defense of the social welfare function as a criterion for policy evaluation in Matthew Adler (2011). Adler describes (in his Chapter 4) a way in which a policy designer can extend a limited set of information. My argument suggests an alternative approach.

[3] See Weinzierl (2014, 2016a, 2016b) for additional examples.

[4] See Boskin and Sheshinski (1978, result 11) as well as Oswald (1995, result 18).

[5] As I discuss below, public opinion thereby echoes the responses to envy-based redistribution of even staunchly welfarist scholars. For example, Kaplow (2008) is characteristically clear: "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."

to welfarism–are less likely to accomodate envy's welfare implications. These patterns are consistent with respondents using nonwelfarist principles to encode concerns about the indirect welfare consequences of policy.

One way to see the argument of this paper is as extending the reach of a logic for so-called rule utilitarianism from the the realm of individual ethics to society's policymaking decisions. 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. John Stuart Mill (1871) calls these nonwelfarist rules "secondary principles," and like many others he emphasizes their importance in achieving an underlying welfarist goal.[6] John Harsanyi (1992) 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." The argument of this paper is that this logic for why individuals making ethical choices ought to use principles based on what Mill called "the experience of human life" applies just as well to the difficult normative judgments societies must make.

The general version of this argument–namely, that policy design ought to be influenced by principles reflecting society's accumulated wisdom–is far from original to this paper.[7] Most prominently, Friedrich Hayek (1973) warns that "irremediable ignorance of most of the particular facts which determine the processes of society" prevents us from anticipating "all the costs of achieving particular results" and compels us to give weight to such principles. Related observations can be found in the writings of eminent thinkers such as David Hume, Adam Smith, and Edmund Burke.[8]

Despite its empirical plausibility and intellectual pedigree, this general argument exerts essentially no influence on the modern welfarist approach to optimal policy analysis, with predictable consequences. As I and others have discussed elsewhere, and as the evidence below on the response to envy suggests, that standard approach yields policy recommendations that appear to be at odds with prevailing public priorities for policy, priorities that in turn appear to be consistent with historically prominent nonwelfarist principles.[9] By advancing an argument for why such nonwelfarist principles may further, rather than displace, welfarist goals, this paper hopes to engage researchers who embrace the standard approach.

This paper proceeds as follows. Section 1 formalizes, in a setting based on Kaplow and Shavell (2001), how incorporating nonwelfarist principles may be optimal in the face of limits to the information society has on a policy's consequences. Section 2 turns to an example of this argument in action, reviewing the controversial implications of envy for welfarist-optimal taxation and using new survey results to show how public opinion against envy-based redistribution appears to reflect the influence of nonwelfarist principles on policy judgments. Section 3 links this argument to the vast literature on the use of rules in individual utilitarian ethics and earlier, more general versions of this argument as a response to intractable complexity. Section 4 concludes.

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

[7] In addition to the authors noted below, the current work of Eric Nelson (2016) on the link between egalitarianism and the problem of theodicy is closely, if subtly, related to the argument of this paper.

[8] Given this list of authors, it may be important to note that the logic for using nonwelfarist principles to achieve welfarist goals is not exclusively relevant to those on the political "right." On the contrary, the same logic applies for those on the political "left" whose welfarism gives priority to the worst off and whose resulting policy preferences 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.

[9] See Weinzierl (2014, 2016a, 2016b), in which I present evidence on the influence of Mill's (1871) Equal Sacrifice principle and Smith's (1776) view, described by Musgrave (1959) as Classical Benefit-Based Taxation. Work by Saez and Stancheva (2015) and Charité, Fisman, and Kuziemko (2016) identifies related departures of policy and public opinion from common normative assumptions.

1 A formal illustration of the role for nonwelfarist rules

To provide a formal statement of this paper's main argument, I generalize the framework in which Kaplow and Shavell (2001) analyze the conflict between welfarism and nonwelfarist principles.

1.1 Full information

Start by assuming, as in Kaplow and Shavell (2001), that the policy designer is able to predict (perhaps probabilistically) the complete set of a policy's consequences. 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 only through individual welfare levels: 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 $U_i(x^s) \geq U_i\left(x^t\right)$ for all i and $U_i(x^s) > U_i\left(x^t\right)$ 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.

1.2 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.[10] Recall that $U(x)$ is unobservable, so the designer cannot directly estimate these effects.[11] Let Ξ be the set of all χ. Denote the individual welfare function defined on this limited information $V : \Xi \to \mathbb{R}$. 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 $\chi \in \Xi$.

[10] I do not take a stand on which of these reasons is at play in any given circumstance. At times, society may be truly unaware of elements of the state space that matter to individual welfare. At other times, society may be unable to model the channels through which those elements matter.

[11] This point intersects with the limitations of econometrics. In many areas of tax policy, for example, many outcomes of interest or determinants of such outcomes are functionally unobservable.

1.2.1 A misleading observable state

Consider the specific but plausible case in which the observable state χ is 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 information, 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.

1.2.2 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.[12] 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.[13]

1.3 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, as those authors emphasize.

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

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

[12] An ideal $\hat{F} : \Xi \to \mathbb{R}$ would yield recommendations "close to" the true social objective function $F : X \to \mathbb{R}$. The process by which society decides on such an $\hat{F}$, including how it chooses α, is an important topic for further research. That task may be facilitated by the fundamentally welfarist social objective assumed here, as welfarism makes the ranking of options relatively straightforward. In the specific case of tax policy, work on principles such as Mill's equal sacrifice or Smith's classical benefit-based taxation suggest they are promising candidates for nonwelfarist principles implicitly chosen by (at least U.S.) society (see Scheve and Stasavage 2016, Weinzierl 2014, 2016a, 2016b). Thanks to Jiafeng Chen for suggesting this discussion.

[13] Note that 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. 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.

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.[14] 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:

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

Conditions such as (8) and (9) lead Kaplow and Shavell, in the full-information analogue to this example, to conclude 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.

1.4 Examples

To make this formalization more concrete, suppose that the distinction between χ and x is due to time horizon. That is, the policy designer cannot see infinitely far into the future, so the effects on welfare of elements in x but not in χ are impossible to model. Fortunately, past human experience has found that those hidden effects are systematically related to $\alpha \in \chi$, a feature of the state that is observable far before the effects occur, perhaps due to inertia or other complex channels of causality. Society has therefore developed, and passed from generation to generation, a principle regarding α. A welfarist planner will, in this situation, want to put weight on that principle.

To be even more concrete, suppose that an important welfare-relevant element of x not included in χ is the degree of interpersonal trust in a society, and α is the extent to which envy is acknowledged as a morally legitimate emotion in that society in a preceding generation. A policy designer may be unable to model how welfare generations in the future will be impacted by a current policy's effect on future levels of interpersonal trust. But, that policy designer may be able to choose a policy that supports or undermines the moral legitimacy of envy in the current generation. If

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

accommodating envy has tended to lower interpersonal trust–and thus overall welfare–in the past, society is likely to have developed a policymaking principle that discourages it. In that case, a policy designer will correctly attend to that nonwelfarist principle when making policy assessments.

As I will discuss in the next section, this last example fits well with both scholarly and popular resistance to envy-based redistribution in reality. It may be hard to explain precisely why many of us reject a policy that accommodates envy, though we may be confident in that judgment. In fact, many of us would agree that a policy accommodating envy is less likely to produce a Pareto-superior outcome than one that does not, so that the appearance of a Pareto-improving accommodation of envy would be misleading. We believe it would be missing something relevant to overall welfare even if we cannot say exactly what that something is.

2 A welfarist role for nonwelfarist rules in action: resistance to welfare-increasing, envy-based redistribution

In this section, I provide evidence of an example in which it appears that policy preferences are consistent with this paper's theoretical argument for the use of nonwelfarist rules. I focus on the role of envy in justifying income redistribution. First, I briefly review how tax theorists have responded to the complications introduced by the presence of envy. Then, I present the results of a new survey showing popular skepticism toward envy-based redistribution that appears to reflect the influence of nonwelfarist principles.

2.1 The welfarist's dilemma with envy in standard optimal tax theory

Given a standard welfarist social objective function, the direct implications of envy for optimal tax policy are to increase the extent of redistribution. Boskin and Sheshinski (1978) and Oswald (1992) show this result in linear and nonlinear tax environments, respectively. For the purposes of this paper, a simpler setup is sufficient. I focus directly on allocations of resources, broadly construed, and show that the acceptance of envy may cause a reversal in the welfarist preferences over two allocations.

Consider the following utility function for individual i comparing himself or herself to individual j, where y_i is a measure of economic resources and β, γ, and θ are scalars:

$$U_i\left(y_i, y_j\right) = \frac{\beta}{1-\gamma}\left[\left(\frac{y_i}{\left(\max\left\{y_i, y_j\right\}\right)^{\theta}}\right)^{1-\gamma} - 1\right]. \tag{11}$$

Note that utility in (11) depends on envy through the ratio of individual i's resources to the maximum level of individual resources in the population (here, of only two individuals).[15] The parameter θ controls the degree of envy: $\theta = 0$ means envy exerts no influence on utility; $\theta = 1$ means that only i's income relative to the maximum matters for i's utility.[16]

[15] This utility function is a standard representation of envy, but it could capture nonenvious concerns about the effects of relative status on wellbeing.

[16] This utility function's form of envy follows Rawls' (1971) definition: "We envy persons whose situation is superior to ours...and we are willing to deprive them of their greater benefits even if it is necessary to give up something ourselves." (pp. 532-5) Rawls made an important distinction, summarized in Fleurbaey et al. (2008), in that he "did accept legitimate forms of resentment induced by inequalities, while at the same time rejecting feelings of envy as irrelevant for social evaluation." Adam Smith (1759) similarly defines envy as "that passion which views with malignant dislike the superiority of those who are really entitled to all the superiority they possess."

The following table displays two allocations of resources and the resulting well-being levels for different values of θ.[17]

Table 1: Envy and well-being

	Resources		Well-being, $\theta = 0$		Well-being, $\theta = 0.65$	
	Alloc. 1	Alloc. 2	Alloc. 1	Alloc. 2	Alloc. 1	Alloc. 2
Person i	5,000	6,000	139.3	139.8	120	110
Person j	10,000	16,000	140.7	141.2	160	166
Total	15,000	22,000	280.0	281.0	280	276

Note that Allocation 2 provides both individuals with greater resources but increases inequality between them. One way to interpret these allocations is that Allocation 1 can be obtained from Allocation 2 through a policy that destroys resources to achieve greater equality.

This example illustrates the power of envy to make welfarist-optimal policy more redistributive: a utilitarian will prefer Allocation 1 if envy's effects on well-being are considered normatively relevant, but the same utilitarian will prefer Allocation 2 if not. That is, as Table 1 shows, Allocation 2 provides a greater sum of well-being if each individual's utility function is independent of the other's allocation (i.e., when $\theta = 0$). However, in the case of $\theta = 0.65$, where envy of the most advantaged lowers Person i's well-being, total well-being is greater in Allocation 1.

The welfarist logic for redistribution based on envy has been endorsed, perhaps grudgingly, by some economists and philosophers. They argue that, if envy reflects a real factor in well-being, it should count in our welfarist policy evaluations. Hare (1981), for example, writes: "...Envy has been accounted a sin; but its existence is a fact relevant to moral thinking. If there were good arguments in favour of inequalities, there would be equally good arguments in favour of suppressing envy. But if there are already arguments against the inequalities, it is likely to be better to remove the envy by removing the inequalities than by preaching against it as a sin." Rakowski (1991) makes the broader point: "utilitarianism forsakes its motivating idea and produces unattractive consequences if it excludes external preferences from its calculus."

On the other hand, many leading economists and philosophers (welfarist and otherwise) flatly reject policy that accommodates envy. Some argue that accommodating envy would not yield even directly welfarist gains.[18] But most take a more indirect route. For example, Harsanyi (1982) strongly argues for omitting envy from welfare calculations because it is an example of "clearly antisocial preferences" while "the basis of utilitarianism is benevolence." Mirrlees's (1971) model assumes away other-regarding preferences, consistent with his later arguments that they can be generally disregarded (Mirrlees 1982, p. 76). Hayek's opposition is vehement: "However human, envy is certainly not one of the sources of discontent that a free society can eliminate. It is probably one of the essential conditions for the preservation of such a society that we do not countenance envy, not sanction its demands by camouflaging it as social justice, but treat it, in the words of John Stuart Mill, as 'the most anti-social and evil of all passions.'" Thomas Nagel (1995) recognizes that

[17] I set $\gamma = 2$ for each situation and let β vary to scale total well-being to 280 in both instances of Allocation 1. These choices are unimportant for the results.

[18] For example, Rakowski (1991) writes that envy "tends to create much petty dissatisfaction and little genuine happiness." Or, following Banerjee (1990), if envy is driven by ordinal, not cardinal differences, it cannot be ameliorated by reducing inequality and anything short of full redistribution–which may be impossible for incentive reasons–is counterproductive. Nozick (1974) asserted that envy is inherently multidimensional and that a "conservation of envy" law applies such that removing disparities of one sort, say in income, will simply heighten envy along other dimensions. And Elizabeth Anderson (1999) noted that acknowledging envy makes the envious feel worthless, and any envy-based redistribution is pointless. As I explain below, my survey is designed to avoid these possibilities and isolate the case in which accommodating envy would directly raise overall welfare.

"...it must be acknowledged that such inequalities [of status] can cause a good deal of pain...But in the end, the unhappiness of unsuccessful contenders, or the low self-esteem of those who cannot even make the attempt, are not evils that a decent society should be asked to weigh in the balance against the concentrated support of what is best. Such inequalities are inextricable from the recognition and pursuit of certain values too important to be compromised."

One interpretation of this brief review is that the scholarly debate over tax policy's response to envy pits pure, or direct, welfarists against those who look past envy's direct welfare implications and rely on indirectly welfarist reasoning that is encoded into nonwelfarist principles. I now turn to evidence on how the public wrestles with that same dilemma.

2.2 New evidence on popular resistance to envy-based redistribution

First, I describe the survey.[19] I listed the survey on Amazon's Mechanical Turk (M-Turk) platform[20] in mid-2016, and approximately 750 respondents were paid $2.50 or $3.00 to complete the survey, a task that took less than ten minutes on average. The results of interest relate to a set of questions asked early in the survey, where each respondent was randomly allocated to one of seven versions of the main survey module. The respondents were also asked to self-report a set of demographic characteristics, answer three mathematical exercises, and choose responses to several questions on their political views.

2.2.1 Baseline setup and results

At the start of the main module of the survey, respondents are introduced to a hypothetical scenario that includes two unequal allocations of money across two individuals. (Readers will recognize the scenarios from Table 1 above). Specifically, in the baseline version respondents see the following screen after agreeing to take the survey:

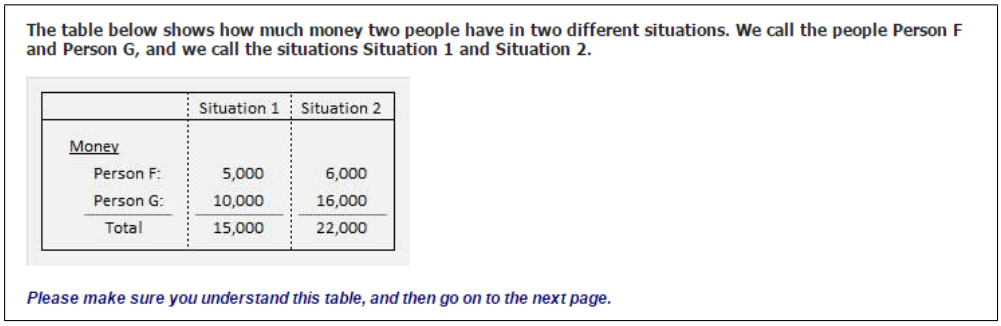


Figure 1. The first step of the main survey module in the baseline case.

Figure 1 shows that a shift from Situation 1 to 2 increases the amount of money each individual has but widens the gap between them. If an individual's welfare is independent of the other's resources, both should have greater well-being in Situation 2 than in 1.

Respondents then advance to a second screen with additional information. In particular, they are given a measure of each individual's well-being (for which a definition is given) in each situation.

[19] Available at https://hbs.qualtrics.com/SE/?SID=SV_80niBI5g6Fzhgln

[20] See Horton, Rand, and Zeckhauser (2010) for an analysis of online labor markets for economic research and see Saez and Stantcheva (2015), Kuziemko, Norton, Saez, and Stantcheva (2015), and Weinzierl (2014, 2016b) for examples of its use in gauging distributive preferences.

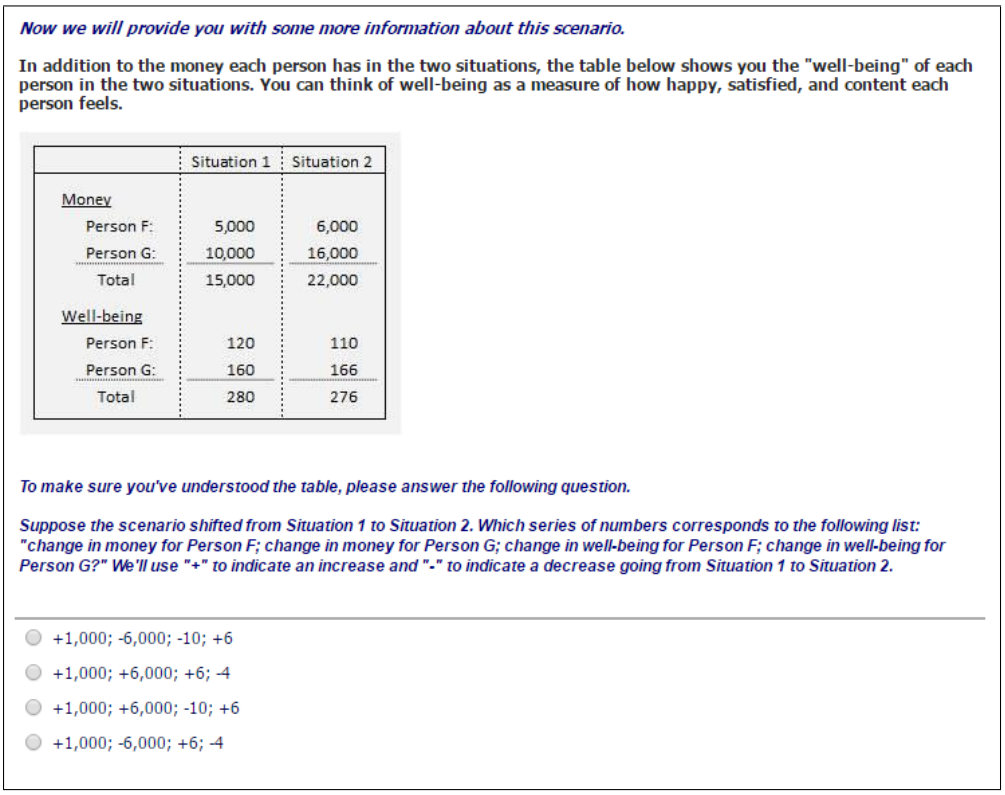


Figure 2: The second step of the main survey module in the baseline case.

As Figure 2 shows, a shift from the first to the second allocation increases the well-being of the better-off individual but lowers the well-being of the worse-off individual. One interpretation of the well-being decrease for Person F from Situation 1 to Situation 2 is that F suffers from envy toward Person G.[21] Note that, to increase engagement with the question, I require respondents to calculate the changes in money and well-being for each individual. If they choose an incorrect answer, they receive an error message and must try again.

Importantly, the gain in well-being for the better-off individual is smaller than the loss for the worse-off individual in Figure 2, so the second allocation is worse from a utilitarian (or more egalitarian) perspective than the first. This fact is highlighted by showing the total levels of well-being in each situation. Being explicit on this point with respondents is important for avoiding concerns, raised by some scholars as mentioned above, that accommodating envy would be ineffective from a directly (i.e., observable) welfarist perspective.

[21] Though this interpretation is not the only possibility (for example, respondents may think F has reduced access to scarce goods in Situation 2), variations on this baseline setup discussed below that explicitly suggest a role for envy reinforce the baseline results.

Finally, as shown in Figure 3, respondents are asked to consider the following screen:

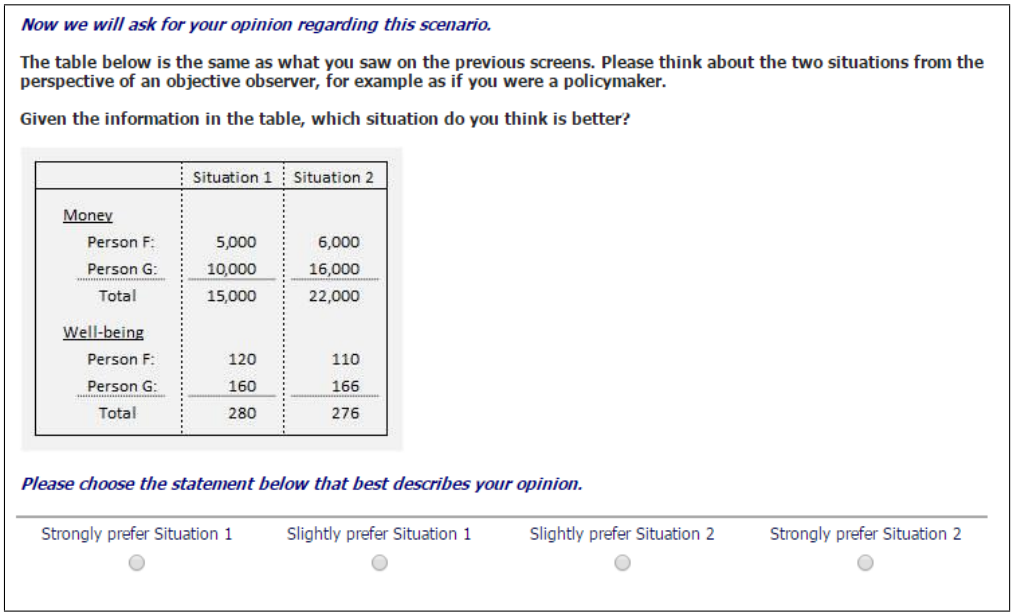


Figure 3: The third step of the main survey module in the baseline case.

Note that respondents are not given a specific criterion by which to judge, from the perspective of an objective observer, "which situation do you think is better?"

Resistance to envy-based redistribution In response to this question, 54% of respondents prefer Situation 2. In other words, by not choosing Situation 1, more than half of the respondents reject a potential utilitarian gain when that gain is due to a negative welfare externality imposed by envy. I will refer to this choice as a "rejection of envy-based redistribution" throughout the remainder of the paper.

This result stands in sharp contrast to what would be chosen by the welfarist normative objective used in conventional optimal tax analysis. But, to reiterate this paper's main argument, if respondents put weight on nonwelfarist principles that capture society's accumulated knowledge of envy-based redistribution's negative, unobservable welfare effects, a preference for Situation 2 is consistent with an underlying commitment to welfarism.

2.2.2 Robustness to variations on the baseline setup

Several variations on the survey support this baseline finding and provide insights into what drives it. As shown in Figure 4, the share of respondents preferring Situation 2, and thereby resisting envy-based redistribution, ranges from 40% to 73% across these variations.

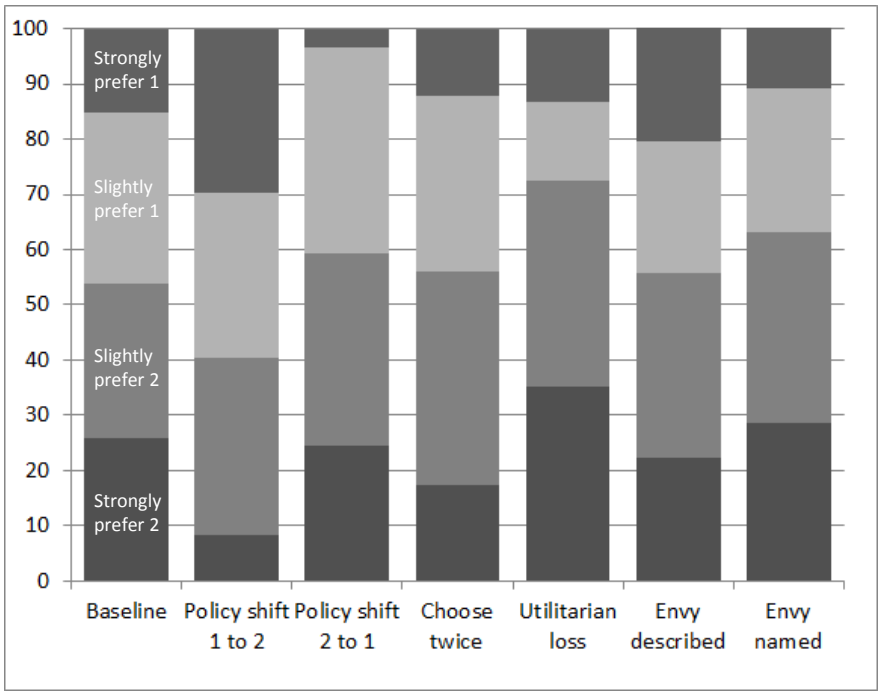


Figure 4: Responses to the baseline and six variations of the main survey module.

The second and third columns of Figure 4 show the results for two variations that present the choice between the situations as an active policy intervention. In the "Policy shift 1 to 2" variation the respondent sees only Situation 1 at first (both the money and well-being data). On the next screen, the respondent is told that a policy is available that would cause a shift to Situation 2. Then, the respondent chooses whether to support that policy intervention. The "Policy shift 2 to 1" variation is the same, but with Situation 2 presented as the starting point. If respondents generically dislike policy intervention, we would expect a smaller share of respondents to choose Situation 2 in the first of these variations and a larger share to choose it in the second. Figure 4 confirms these expectations, with 41% and 59% of respondents choosing Situation 2 compared to 54% in the baseline.

The fourth column gives the results when I ask respondents to give their opinion on the situations twice: first after seeing only the allocations of money (as in Figure 1) and then again after seeing the levels of well-being (as in Figure 3). This variation potentially increases respondents' engagement with and comprehension of the scenario. It also tests the extent to which having them focus on the allocations of money without information on well-being affects their eventual judgments. In this variation, 70% of respondents choose Situation 2 in the first step. With the information on well-being, that share is 56%, close to the baseline result. I return to this variation in Section 2.3.2.

The fifth column of Figure 4 corresponds to a variation in which I change the well-being calculations so that the shift from Situation 1 to Situation 2 generates a gain to Person F that is smaller than the loss to Person G. That is, total well-being is greater in Situation 2 than in Situation 1. If a respondent chooses Situation 1 in this case, he or she would appear to put greater weight on a prioritarian, or even maximin objective. In this variation, 73% of respondents choose Situation 2, suggesting that only a small share of respondents find such an objective determinative for their

distributive justice judgments.

In sum, these variations support the baseline results. So, too, do the variations behind the final two columns of Figure 4 to which I turn in the next subsection.

2.3 Suggestive evidence on use of nonwelfarist principles

Further examination of the survey results, including from two additional variations on the baseline setup, provide suggestive evidence that nonwelfarist principles may play the role suggested by the argument proposed in this paper, at least in the case of envy-based redistribution.

2.3.1 Increasing envy's salience

First, two variations increase the salience of envy's role in the hypothetical scenario. If respondents are weighing the utilitarian advantage of Situation 1 over Situation 2 against a nonwelfarist principle that rejects accommodating envy, priming their attention to that principle ought to sway preferences toward Situation 2.

The first variation highlights the role of envy, but avoids explicitly mentioning the word "envy." In particular, after the same screens as in Figure 1 and Figure 2, respondents assigned to this variation see the following screen[22].

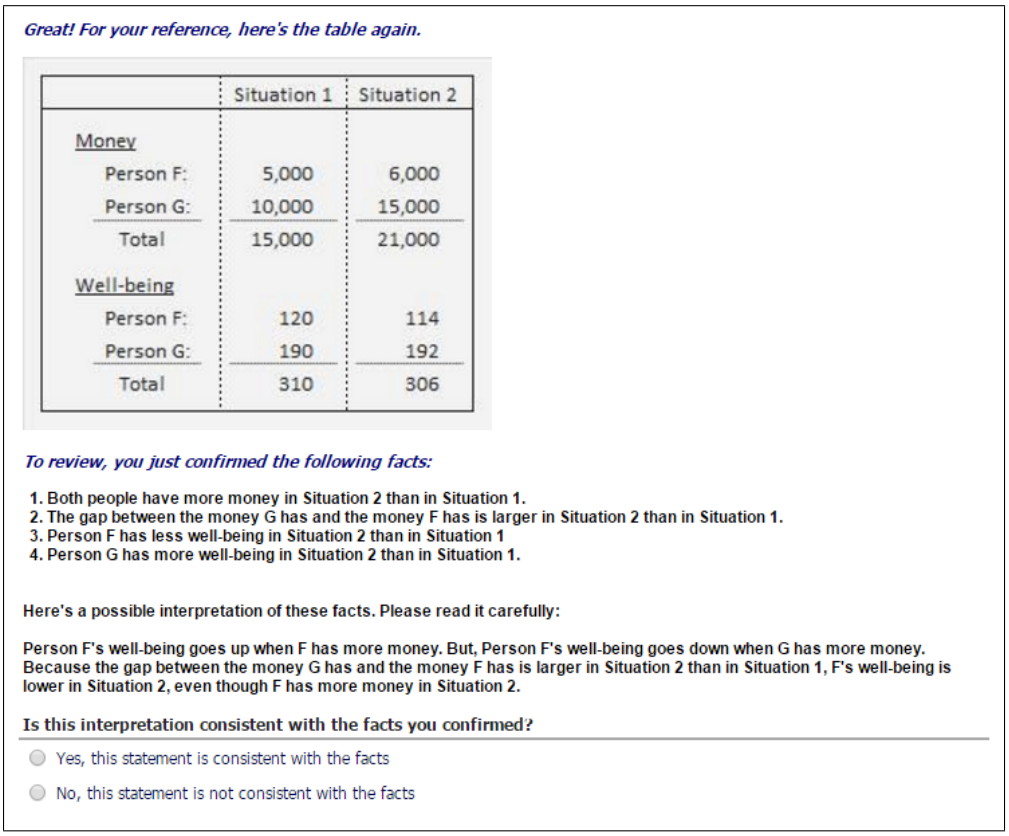


Figure 5: The third step of the main module in the Envy Salient variation.

[22] The well-being numbers are slightly different in this variation, but they play the same role as in the baseline specification.

The key variation in this version of the survey is that respondents are shown a "possible interpretation" of the scenario that describes the effect of Person G's allocation on Person F's well-being, making clear the negative externality envy exerts.[23]

The second variation more aggressively raises the salience of envy by explicitly naming it, replacing the last paragraph in Figure 5 with the following:

> Person F feels envy toward Person G. Person F's well-being goes up when F has more money. But, due to envy, Person F's well-being goes down when G has more money. Because the gap between the money G has and the money F has is larger in Situation 2 than in Situation 1, F's well-being is lower in Situation 2, even though F has more money in Situation 2.

As the final two columns of Figure 4 show, the share of respondents choosing Situation 2 in these variations rises to 55% and 63%, respectively. The latter of these is significantly greater than the baseline result at the 10% confidence level. The effect of increasing the salience of envy appears, therefore, to be to increase the share of respondents rejecting envy-based redistribution.[24]

These results are consistent with respondents using nonwelfarist principles to reject envy-based redistribution. Recall that the direct implications of envy for a welfarist criterion are made explicit and are constant across these variations, suggesting that something other than information on these direct welfare effects is driving the respondents' opinions. That is, if individuals give weight to nonwelfarist principles that resist the accommodation of envy, raising their level of awareness that Situation 1 violates those principles–while the utilitarian advantage of Situation 1 remains the same–would predict the increased support for Situation 2 that these variations generate.

2.3.2 Relation to political views

We can utilize results from the "Choose twice" variation to provide further suggestive evidence that nonwelfarist reasoning may be behind the general rejection of envy-based redistribution. In that variation, respondents first choose between the situations knowing only the allocations of money across Person F and Person G. Then, the well-being data are revealed, and the respondents choose a second time.

In this variation, 70% of respondents choose Situation 2 when only information on money is shown, and 36% of those respondents report a shift in their preferences toward Situation 1 after the well-being data are revealed. One interpretation of such a shift is that the respondents give sufficient weight to the potentially surprising facts they learn from the well-being data, namely that envy has caused both Person F's well-being and total well-being to be lower in Situation 2 than in Situation 1. Those respondents who shift from Situation 2 to Situation 1 are, in other words, more supportive of envy-based redistribution.

Respondents who express support for libertarianism, a prominent alternative to welfarism, are less likely to change their preferences toward Situation 1 after the well-being data are revealed. Only 21% of supporters of libertarianism shift, compared to 46% of those who oppose libertarianism (a statistically significant difference) Table 2 gives the results of a simple logit regression analysis using the self-reported demographic and opinion data from the survey as explanatory variables for the

[23] To increase the likelihood that respondents perceive this effect, I require them to answer whether this interpretation is consistent with the calculations they made (it is). If they choose an incorrect answer, they receive an error message and must try again.

[24] An alternative interpretation is that respondents were unable to understand the mechanism behind the changes in well-being without these explanations. In that case, the higher numbers choosing Situation 2 in these variations suggest greater baseline opposition to redistribution based on envy.

likelihood the respondents report a stronger preference for Situation 1 relative to Situation 2 after the well-being data are presented than before.

Table 2: Explaining support for envy-based redistribution

	Coefficient	Standard error
Political perspective on economic issues (L to R)	−0.17	(0.35)
Support for Libertarianism	−1.32	(0.63)**
Belief that individuals are responsible for their basic needs	−0.09	(0.64)
Household's income status at age 45	−0.04	(0.18)
Household's income status at age 10	0.08	(0.17)
Gender (M=0,F=1)	0.88	(0.54)
Age	−0.18	(0.47)
Race (White=0, Black=1)	0.41	(0.35)
Education level	0.03	(0.29)
Score on math questions	−0.23	(0.54)
N	95	

Notes: The dependent variable in this logit regression is the 0-1 indicator of whether the respondent expresses more support for Situation 1 relative to Situation 2 when the well-being allocations are revealed in the "Choose twice" variation of the baseline survey setup. A positive coefficient on an explanatory variable can be interpreted as meaning that a higher value for it is related to a higher likelihood that a respondent will support envy-based redistribution. The symbol ** denotes significance at the 5% level.

Table 2 shows that support for libertarianism is significantly and negatively related to this outcome, with an estimated marginal effect of shifting from opposing to supporting libertarianism equal to a 27 percentage point increase in the likelihood of changing preferences (consistent with the gap described above).

This result provides further suggestive evidence of the role of nonwelfarist principles in the public's resistance to envy-based redistribution. While some respondents respond to the welfare information in the second stage of this variation in a way that acknowledges (and offsets) the effects of envy on well-being, fewer libertarians–who generally reject welfarist reasoning for policy–respond in this way. This is the pattern we would expect if nonwelfarist rules have more normative significance for the policy judgments of some (i.e., libertarian) individuals than others (i.e., welfarists).

3 Rule utilitarian policy design

The argument presented in this paper can be seen as extending a familiar logic for rule utilitarianism beyond individual ethics and as a specific version of a more general argument found in the writings of prominent thinkers, especially Hayek. In this section, I briefly review these linkages.

3.1 An informational logic for rule utilitarianism

A vast philosophical literature on rule utilitarianism establishes the importance of nonutilitarian principles for utilitarian individual ethics. Utilitarianism is the idea that questions of morality are to be decided based on the criterion of maximum overall welfare. One way for an individual to implement utilitarianism is to judge each of his or her actions by its impact on overall welfare, a practice philosophers name act, or direct, utilitarianism. An alternative implementation is to judge

each action according to a rule that, when followed, will yield maximum overall welfare. Because this rule utilitarianism avoids the explicit calculation of an act's welfare impact, it is sometimes called indirect utilitarianism. As quoted in the Introduction, Harsanyi gave an emphatic endorsements of rule over act utilitarianism for individuals. At a formal level, rule utilitarianism dominates act utilitarianism because a rule utilitarian can choose adherence to act utilitarianism as his or her rule. But Harsanyi and others who argue for rule utilitarianism in individual ethics have a more substantive logic, as well.[25]

I will emphasize one aspect of the substantive logic for why individuals ought to follow rule utilitarianism: namely, that limits to our knowledge make it unrealistic and unwise for an individual to be an act utilitarian.[26] 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.[27]

In a well-known example of this logic, R.M. Hare (1981, 1982) stresses 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, 1982, 1992) 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." As summarized by Marc Fleurbaey, Maurice Salles, and John Weymark (2008), Harsanyi concluded that: "an optimal rule utilitarian code provides desirable incentives to take actions (to work hard, to save and invest, etc.) that benefit society as a whole, and it provides assurances that individual interests will be protected (e.g., the interest of a lender in having a loan repaid). These effects are typically fairly small when considering individual acts in isolation, but can be considerable when the complete set of rights and obligations that constitute a rule utilitarian's optimal moral code is considered."

Philosophers endorsing this logic for rule utilitarianism have long pointed out that it implies an individual may need to adhere to nonutilitarian principles to achieve utilitarian-optimal outcomes. Hare cites Plato and Aristotle as precursors for his two-level thinking, and Mill forcefully argues this

[25] Jonathan Riley (2008) argues–contrary to Harsanyi–that certain nonconsequentialist priorities (e.g., rights) should be treated as lexicographically prior to utilitarianism. Also see Rawls (1971) on the priority of maximal equal liberty over his difference principle. I follow Harsanyi's fundamentally consequentialist view.

[26] An additional argument (see Hare 1981, chapter 3 and Sen 2011) is that humans are not sufficiently morally disciplined to implement act utilitarian calculations rather than serve their own self-interest. This concern is not a component of my argument because economists routinely include that form of behavioral constraint in optimal policy analysis (i.e., through incentive compatibility constraints).

[27] 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 Dasgupta's (1982) insightful discussion of the optimal policy response to private information.

point.[28] More recently, Derek Parfit (1984) writes: "It is not the aim of a theory to be believed." Instead, at least in part because "[w]e often do not know what the effects of our acts would be," a moral theory may be most effective when individuals adopt principles for their actions that do not directly align with that theory; that is, when a theory is (in Parfit's terminology) "self-effacing."

The main thesis of this paper is that this 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 "general equilibrium" effects of policy choices, for example on unobservable social norms and attitudes.[29]

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."[30]

3.2 Ignorance and policy rules

This paper's argument is a specific version of a more general point made by eminent thinkers since policy design became a topic of study. Perhaps the most prominent development of that point is due to Friedrich A. Hayek, who stressed the impact of ignorance on optimal policymaking along a number of dimensions. Directly relevant to this paper, Hayek (1960) states the informational logic described above for the use of rules in individual ethics: "quite generally, the reliance on abstract rules is a device we have learned to use because our reason is insufficient to master the full detail of complex reality."[31] Moreover, he thinks that the limits to individual knowledge naturally extend to society, as "Knowledge exists only as the knowledge of individuals...The sum of the knowledge of all the individuals exists nowhere as an integrated whole." These observations lead Hayek to a deep skepticism of what he calls 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

[28] "...to argue as if no such secondary principles could be had, and as if mankind had remained till now, and always must remain, without drawing any general conclusions from the experience of human life, is as high a pitch, I think, as absurdity has ever reached in philosophical controversy." (U, II, pp. 296-297).

[29] As I mentioned in an earlier note, 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."

[30] Mill 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." Also see Harsanyi (1982), who discusses promise-keeping and writes: "rule utilitarianism comes fairly close to traditional morality in recognizing the importance of social institutions which establish a network of moral rights and of moral obligations among different people in society."

[31] Also see: "Certainty we cannot achieve in human affairs, and it is for this reason that, to make the best use of what knowledge we have, we must adhere to rules which experience has shown to serve best on the whole, though we do not know what will be the consequences of obeying them in the particular instance."

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."

Long before Hayek, eminent thinkers such as David Hume, Adam Smith, and Edmund Burke emphasized the importance of experience, not just abstract reason, in forming our individual moral judgments and the related importance of respecting social institutions and principles that arose from a social evolutionary process of historical trial and error. Smith (1759), for example, writes: "Some general, and even systematical, idea of the perfection of policy and law, may no doubt be necessary for directing the views of the statesman. But to insist upon establishing, and upon establishing all at once, and in spite of all opposition, every thing which that idea may seem to require, must often be the highest degree of arrogance. It is to erect his own judgment into the supreme standard of right and wrong. It is to fancy himself the only wise and worthy man in the commonwealth, and that his fellow-citizens should accommodate themselves to him and not he to them."[32]

Hayek (1973) laments that political theory in his time had moved away from his arguments and adopted, instead, what he called "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 optimal policy analysis.

4 Conclusion

Evidence that nonwelfarist principles are used in real-world policy evaluation is not limited to this paper. In brief, a large body of research across the social sciences has shown that most people consider multiple, often mutually contradictory principles when making normative judgments on economic policy. I have previously referred to this phenomenon as normative diversity. This paper's argument closely relates to the idea of normative diversity, as it offers a logic for principles not directly related to society's ultimate objective to nevertheless be included in that society's method of policy assessment. If this argument is accepted, the next task for research is to identify the diverse set of rules that act as informational proxies.[33] Potential sources of evidence on this include public opinion, robust features of policy that help delineate between principles, political rhetoric, and scholarly analysis.

Though the argument for a welfarist role for nonwelfarist rules has a long historical reach, a foundation in familiar theories of individual ethics, and growing empirical support, it has been discounted in modern normative economic policy research. This paper hopes to renew interest in it and spur further work on it.

[32] See also, for example, Hume (1738, vol 2, 3.2.2; 1752, p. 60) and Hayek (1960, note 37 p. 436); Smith (1759, 7.3.2); and Burke (1790, p. 120).

[33] For evidence and analysis of normative diversity and two prominent nonwelfarist principles that may exert influence in U.S. policymaking, see Weinzierl, 2014, 2016a, 2016b.

References

[1] Adler, Matthew (2011). *Well-being and Fair Distribution.* Oxford.

[2] Anderson, Elizabeth (1999). "What is the Point of Equality?" Ethics (109), pp. 287-337.

[3] Banerjee, Abhijit (1990). "Envy," in *Economic Theory and Policy*, Bhaskar Dutta, Shubhashis Gangopadhyay, Dilip Mookherjee, and Debraj Ray (eds), Oxford.

[4] Boskin and Sheshinski (1978)."Optimal Income Redistribution when Individual Welfare Depends upon Relative Income," *Quarterly Journal of Economics*, pp. 589-601.

[5] Brennan, Geoffrey. (1973), "Pareto Desirable Redistribution: The Case of Malice and Envy." *Journal of Public Economics* (2), pp. 173-183.

[6] Burke, Edmund. (1790). *Reflections on the Revolution in France*, Penguin.

[7] Charité, Jimmy, Raymond Fisman, and Ilyana Kuziemko (2015), "Reference Points and Redistributive Preferences: Experimental Evidence," NBER Working Paper 21009.

[8] Dasgupta, Partha (1982), "Utilitarianism, information and rights," in Sen and Williams.

[9] Dworkin, Ronald (1981). "What is Equality? Part 2: Equality of Resources," *Philosophy and Public Affairs*, 10(4), pp. 283-345.

[10] Fleurbaey, Marc, Maurice Salles, and John A. Weymark (2008), eds. *Justice, Political Liberalism, and Utilitarianism,* Cambridge.

[11] Hammond, Peter (1982), "Utilitarianism, uncertainty, and information," in Sen and Williams.

[12] Hare, R.M. (1981). *Moral Thinking*, Oxford.

[13] Hare, R.M. (1982), "Ethical theory and utilitarianism," in Sen and Williams.

[14] Harsanyi, John C. (1977), "Rule Utilitarianism and Decision Theory," Erkenntnis 11(1), 25-53.

[15] Harsanyi, John C. (1982), "Morality and the theory of rational behaviour," in Sen and Williams.

[16] Harsanyi, John C. (1992), "Game and Decision Theoretic Models in Ethics," chapter 19 in *Handbook of Game Theory*, R.J. Aumann and S. Hart, (eds.), Elsevier.

[17] Hayek, Friedrich A. (1945). "The Use of Knowledge in Society," *American Economic Review*, 35(4), pp. 519-530.

[18] Hayek, Friedrich A. (1960). *The Constitution of Liberty.* Chicago.

[19] Hayek, Friedrich A. (1973). *Law, Legislation, and Liberty.* Chicago.

[20] Horton, John J., David G. Rand, and Richard J. Zeckhauser (2010). "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics* 14.3, pp. 399-425.

[21] Hume, David (1738) *A Treatise of Human Nature.* Oxford (Clarendon 2007).

[22] Hume, David (1752), "Of the Original Contract," in *Political Essays*, Liberal Arts Press (1973).

[23] Kaplow, Louis (2008). *The Theory of Taxation and Public Economics*, Princeton.

[24] Kaplow, Louis and Steven Shavell (2001). "Any non-welfarist Method of Policy Assessment Violates the Pareto Principle," *Journal of Political Economy* 109(2).

[25] Kuziemko, Ilyana, Michael Norton, Emmanuel Saez, and Stefanie Stantcheva (2015). "How Elastic are Preferences for Redistribution: Evidence from Randomized Survey Experiments," *American Economic Review* 105(4), pp. 1478-1508.

[26] Mill, John Stuart (1871). *Principles of Political Economy*. Oxford University Press (1994).

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

[28] Mirrlees, J.A., (1982). "The economic uses of utilitarianism." In Sen and Williams (1982).

[29] Murphy, Liam and Thomas Nagel. (2002). *The Myth of Ownership*, Oxford.

[30] Musgrave, Richard (1959). *The Theory of Public Finance,*. McGraw-Hill.

[31] Nagel, Thomas (1995). *Equality and Partiality*, Oxford.

[32] Nelson, Eric (2016). Personal communication.

[33] Oswald, Andrew (1983). "Altruism, Jealousy, and the Theory of Optimal Non-linear Taxation," *Journal of Public Economics*, 20, pp. 77-87.

[34] Parfit, Derek (1984). *Reasons and Persons*, Oxford.

[35] Rakowski (1991). *Equal Justice*, Oxford.

[36] Rawls, John (1971). *A Theory of Justice*, Harvard.

[37] Riley, Jonathan (2008). "Rule Utilitarianism and Liberal Priorities," in *Justice, Political Liberalism, and Utilitarianism,* Fleurbaey et al. (eds.), Cambridge.

[38] Saez, Emmanuel and Stefanie Stantcheva (2015). "Generalized Social Marginal Welfare Weights for Optimal Tax Theory." American Economic Review, 106(1), pp. 24-45.

[39] Scheve, Kenneth, and David Stasavage (2016). *Taxing the Rich*, Princeton.

[40] Sen, Amartya and Bernard Williams (1982), eds. *Utilitarianism and Beyond*, Cambridge.

[41] Smith, Adam (1759). Theory of Moral Sentiments.

[42] Smith, Adam (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*, New York: Prometheus Books.

[43] Varian, Hal. (1974). "Equity, Envy, and Efficiency," *Journal of Economic Theory*, 9(1).

[44] Weinzierl, Matthew (2014). "The Promise of Positive Optimal Taxation: Normative Diversity and a role for Equal Sacrifice," *Journal of Public Economics*, 118, pp. 128-142.

[45] Weinzierl, Matthew (2016a). "Revisiting the Classical View of Benefit Based Taxation," Economic Journal (forthcoming), also NBER Working Paper 20735.

[46] Weinzierl, Matthew (2016b). "Popular Acceptance of Inequality due to Brute Luck and Support for Classical Benefit-Based Taxation," NBER Working Paper 22462