ALL INDIVIDUALS MAY BE MADE
WORSE OFF UNDER ANY
NONWELFARIST PRINCIPLE

Louis Kaplow
Steven Shavell

Discussion Paper No. 350

02/2002

Harvard Law School
Cambridge, MA 02138

All Individuals May Be Made Worse Off
under Any Nonwelfarist Principle

Louis Kaplow and Steven Shavell[*]

Abstract

Nonwelfarist principles — notably, deontological principles — are often advanced to guide moral decisions. The types of choices addressed by such principles typically seem, on their face, to involve conflicts of interests among individuals. Nevertheless, it can be demonstrated that any nonwelfarist principle will, in some circumstances, favor choices that make all individuals worse off. For a variety of reasons, this conclusion has important implications for moral theories that are understood to support nonwelfarist principles.

All Individuals May Be Made Worse Off
Under Any Nonwelfarist Principle

Louis Kaplow and Steven Shavell

I. INTRODUCTION

A variety of deontological principles have been endorsed by moral philosophers. Examples include theories of retributive justice, the idea that it is wrong per se to break a promise, belief in the independent importance of rights, and acceptance of such notions as the act/omission distinction and the doctrine of double effect.[1] Although many do not favor positions as extreme as some of Kant's (1785), wherein certain moral rules may not be violated however adverse the consequences, a substantial portion of philosophers would accord weight to some nonconsequentialist notions. A fortiori, they would reject an insistence that social decisions be based exclusively on how they would affect each individual's welfare — the most commonly discussed type of purely consequentialist moral theory.

In this article, we explain that giving weight to any nonwelfarist principle — and thus to any deontological principle — entails the view that morality may sometimes require social choices that make *every* individual worse off. That is, all nonwelfarist theories are in conflict with the Pareto principle, which holds that if literally everyone would be better off under one social choice than under another, then the former choice is compelled.[2]

We suspect that many who endorse nonwelfarist principles will be bothered by the fact that, upon examination, the principles imply that it can be morally correct for society to make a choice that is to everyone's detriment. Although we ourselves endorse welfarism in general and the Pareto principle in particular, we wish to emphasize that we do not here offer any defense of the Pareto principle itself. Instead, we limit our analysis entirely to exploring the purely logical claim concerning the conflict between nonwelfarist principles and the Pareto principle.[3]

Our analysis proceeds as follows. We begin by making more precise what we mean by

---

[1]See, for example, Dworkin (1977), Nozick (1974), Rawls (1971), Ross (1930), and Williams (1981).

[2]The version of the Pareto principle stated in the text is formally referred to as the weak Pareto principle. Under the strong version of the principle, if everyone is at least as well off and at least one person is better off under one social choice than under another, then the former choice is deemed to be socially preferable. Our analysis below, showing that any nonwelfarist principle conflicts with the weak Pareto principle, implies a conflict with the strong version of the Pareto principle as well.

[3]We offer arguments in favor of welfarism in Kaplow and Shavell (2002).

welfarism and, accordingly, by identifying which principles are subject to our claim concerning conflict with the Pareto principle.[4] Next, we present two arguments that each establish the conflict. The first shows that in symmetric settings — in which each individual is identically situated — nonwelfarist principles favor social choices under which everyone is made worse off *whenever* the favored choices differ from those dictated by a welfarist approach. The second argument shows, more generally, that for any nonwelfarist principle, there sometimes will be cases in which everyone is made worse off.

After presenting these two demonstrations, we discuss some of their implications. We first emphasize that the requirement of logical consistency in moral argument implies that, if one accepts the Pareto principle, one must reject nonwelfarist principles generally, not merely in settings in which their application makes everyone worse off. Next, we argue that many moral frameworks — such as those employing the Golden Rule, the categorical imperative, or a veil of ignorance — should be understood as demanding that all moral principles be tested in a symmetric setting, where we demonstrate that nonwelfarist principles always make everyone worse off when they differ from welfarism. Then, we discuss the special significance of the Pareto principle under certain moral theories. Last, we comment on the instrumental relevance of nonwelfarist principles under welfarism, which helps to reconcile our analysis with the fact that our moral intuitions often favor nonwelfarist principles.

Before proceeding, we comment briefly on a respect in which our analysis may seem counterintuitive. In particular, it may appear odd — or even impossible — for us to examine nonwelfarist principles in settings in which they result in everyone being made worse off. The reason is that most such principles, on their face, apply to situations involving divergent interests: Parties breaking promises may gain, at the expense of promisees. Committing an act — in contrast to an omission — may save certain individuals, at the expense of others. Corrective justice holds that injurers must compensate their victims. Indeed a central role of moral principles is understood, for good reason, to focus on interpersonal relations involving conflicting interests, not matters in which self-interest is largely unproblematic because the acts in question do not affect third parties.[5]

Accordingly, it seems natural — indeed, necessary — to examine moral principles in settings in which none of the social choices, and thus none of the candidate moral principles, can make everyone better off than can a plausible alternative. Nevertheless, our analysis will demonstrate that this seemingly self-evident understanding is incorrect. Indeed, it turns out to be incorrect quite generally with respect to nonwelfarist moral principles.

---

[4]We use the distinction between welfarism and nonwelfarism rather than the more common distinction between consequentialism and nonconsequentialism simply because it is indeed the former distinction that defines the domain of our argument. That is, our argument that nonwelfarist principles conflict with the Pareto principle applies to consequentialist yet nonwelfarist principles as well as to nonconsequentialist principles.

[5]This is not to deny that there can be paternalistic moral principles and so forth. In such cases, it is usually recognized that it is important to offer a compelling justification if an individual's decisions or preferences are to be overridden.

II. DEFINITIONS

We begin by defining our terms because they indicate the domain of our argument in the sections to follow. Specifically, all principles that are nonwelfarist as we now define it can be shown sometimes to favor social choices under which everyone is made worse off.[6]

A. *Welfarist and Nonwelfarist Principles*

We define a principle to be welfarist if, under the principle, the preferred social choice depends *only* on how it affects individuals' well-being.[7] Accordingly, we define as nonwelfarist any principle under which the preferred social choice depends *in any way* on factors other than individuals' well-being. Thus, we include as nonwelfarist not only absolute deontological principles such as some espoused by Kant (1785) but also mixed or pluralistic principles. The latter group includes both principles that themselves depend in part on welfare and principles that, although independent of welfare, would not be given absolute weight in making social choices but rather would be combined in some manner with welfarist principles. Examples of mixed approaches are Ross's (1930) view under which rights have only prima facie validity, those theories of retributive justice that merely limit what punishment may be imposed on other (perhaps utilitarian) grounds, and views of the act/omission distinction that do not immunize omissions from moral scrutiny but rather insist only that they be judged differently from acts having the same consequences.

A logically equivalent definition of a welfarist principle is that, under it, it is always possible to determine the proper social choice knowing only the effects of possible choices on each individual's well-being. By contrast, to make a social choice under a nonwelfarist principle, it will sometimes be necessary to know more than the effects of possible choices on individuals' welfare.[8]

---

[6]It is not true, however, that no welfarist principle violates the Pareto principle: Consider, for example, the principle that one should *minimize* the sum of utilities. Nevertheless, most commonly discussed welfarist principles deem higher well-being to be desirable and thus do not entail any conflict with the Pareto principle, and in our consideration of welfarism we confine our attention to such principles.

[7]As is apparent from the statement in the text, we focus on the morality of social choices rather than personal morality. It will be clear that our analysis is in principle applicable in the latter domain, although the application is complicated by the two-level nature of moral reasoning under welfarism, as we note in section V.D.

[8]Using this second formulation makes it straightforward to state our definitions of welfarist and nonwelfarist principles formally. First, consider a purely welfarist method of social choice, which depends only on individuals' well-being. Let a possible social choice be denoted by x, interpreted as an exhaustive description of the state of the world that will prevail. Let the well-being (utility) of the first individual, given the choice x, be denoted $U_1(x)$, that of the second individual $U_2(x)$, and so forth for the n individuals in society. Then, under a welfarist principle, the assessment of a social choice x depends exclusively on the well-being of the n individuals under x, namely, exclusively on $U_1(x)$, $U_2(x)$, ..., $U_n(x)$. A welfarist principle can thus be represented using a social welfare function, $W(x)$, as follows:

$$W(x) = F(U_1(x), U_2(x), ..., U_n(x)).$$

Let us illustrate welfarist and nonwelfarist approaches by considering how criminal punishment would be assessed under each. A welfarist would need to know how a proposed policy would affect criminal activity, how the punishment would affect criminals' well-being (and that of any innocent individuals who might mistakenly be sanctioned), the cost of law enforcement under the scheme, the extent to which the public would be upset or pleased about the resulting level of crime and the system of punishment itself, and so forth. By contrast, under retributive theories of justice, this information would not be sufficient (and, under pure retributivist theories, it would not be relevant), for one would need to know the extent to which proposed punishments would fit corresponding crimes. This factor, however, has no relevance per se under welfarism. It would only be indirectly relevant, reflecting the manner in which fitting levels of punishment affect each of the aforementioned factors bearing on individuals' welfare.

B. Individuals' Well-being

Although our definitions of welfarism and nonwelfarism are in terms of individuals' well-being, our analysis to follow does *not* depend on how individuals' well-being is defined, as long as well-being is defined consistently. This can be seen from a brief preview of our arguments that follow. Our demonstrations begin with a nonwelfarist principle, which, as stated above, is one that makes some choices depend on factors other than well-being — which itself could be defined in any particular manner. Then we show that such a principle will favor some choices that reduce everyone's well-being — where well-being is defined in the same manner as when determining that the initial principle was nonwelfarist. Thus, if well-being were taken to be personal happiness, our claim is that a principle that does not make choices depend solely on personal happiness will sometimes reduce everyone's personal happiness. If, instead, well-being were taken to embody a certain objective theory of welfare, our claim is that a principle that does not make choices depend solely on how they promote individuals' welfare thus construed will sometimes reduce everyone's welfare in that sense.

The case in which $W(x_1) > W(x_2)$ is taken to mean that choice x_1 is socially preferred to choice x_2 under the posited welfarist principle, $W(x)$. (We note that economists often call W an individualistic social welfare function because social welfare depends only on each individual's well-being. We also remind the reader that W is ordinarily taken to be an increasing function of the U_i.)

By contrast, under a nonwelfarist principle, the social choice does not depend exclusively on the well-being of the n individuals under x, namely, exclusively on $U_1(x)$, $U_2(x)$, ..., $U_n(x)$. Instead, the social choice depends in some way on x other than regarding the manner in which x influences each individual's well-being. To represent a nonwelfarist principle, therefore, we can employ a different sort of function, say $Z(x)$, which differs from the $W(x)$ function just defined in that it cannot be expressed solely as a function of the n individuals' levels of well-being:

$$Z(x) = F(U_1(x), U_2(x), ..., U_n(x), x).$$

Here, Z may depend not only on each individual's well-being, but also directly on x, which includes all characteristics of the situation that will prevail under a social choice. Thus, it is possible that a characteristic of the situation x that affects no one's well-being nevertheless affects Z. Moreover, it may be that a characteristic of a situation x that does affect individuals' well-being influences Z but in a manner that does not solely reflect the effects on individuals' well-being.

We emphasize the independence of our analysis from the definition of well-being because welfarism (utilitarianism in particular) is often associated with narrow or otherwise problematic views about well-being and more generally because of the general disagreement about what concept of welfare is normatively compelling.[9] For purposes of the present analysis, it is helpful to set aside these issues as well as concerns about whether individuals are able to perceive accurately what will contribute to their own well-being (however construed) and to act accordingly.[10] Concretely, it is helpful simply to imagine cases in which what advances individuals' well-being is uncontroversial and is accurately perceived by each individual. After all, any good welfarist theory or nonwelfarist principle should be applicable to such a case.

C. Welfarism and the Distribution of Income or Well-being

To avoid any misunderstanding, we wish to state how various theories concerning distributive justice should be classified for present purposes. As we will now explain, many such notions can be maintained entirely within a welfarist view. Accordingly, our argument about the conflict between nonwelfarist principles and the Pareto principle is simply inapplicable to such distributive theories.

To elaborate, under a welfarist theory social choices depend only on how individuals' well-being is affected. Many distributive theories, in turn, also depend only on well-being. For example, the view that it is socially desirable to redistribute income in favor of the poor because of the diminishing marginal utility of income — a view associated with utilitarianism — is one that is entirely welfare-based. Similarly, according greater weight to the poor than implied by utilitarianism — even extreme weight, in the spirit of Rawls (1971)[11] — is consistent with welfarism as long as the theory is concerned only with individuals' well-being. In contrast, a theory favoring equality of income or some other measure of resources as an end in itself and thus without regard to effects on individuals' well-being is nonwelfarist and thus is subject to our argument.[12]

---

[9]See, for example, Griffin (1986), Nussbaum and Sen (1993), Scanlon (1998), and Sumner (1996) for a range of views.

[10]Of course, the ultimate normative import of a demonstration that a moral principle favors social decisions under which everyone is worse off will depend upon the appeal of the concept of well-being that one adopts. However, as we indicate in our introduction, this article concerns only the logical claim and does not directly address why it should be seen as problematic to view choices that hurt everyone as morally preferable.

[11]In fact, Rawls would determine who is the least well-off group of individuals by reference to what he calls "primary goods" rather than individuals' well-being. See Rawls (1971: 90-95; 1982). To that extent, his distributive theory is nonwelfarist.

[12]We note, however, that unlike nearly all other nonwelfarist principles, purely distributive nonwelfarist principles may be entirely moot in the symmetric case, the domain of our first demonstration (although related demonstrations are often possible). Our second demonstration, however, applies to such principles.

III. PARETO CONFLICT IN THE SYMMETRIC CASE

A. Demonstration

A basic, natural setting to consider is the symmetric case, that is, one in which every individual is similarly situated with respect to the social choices under consideration. (In particular, one could assume that everyone is identical in every relevant respect, having the same income, opportunities, and preferences.) In such symmetric settings, we now show that everyone will be made worse off whenever a nonwelfarist principle leads to a social decision different from that which would be made if the social goal were concerned exclusively with effects on individuals' well-being.[13]

We begin by offering a simple yet important observation: In symmetric settings, whenever a social choice raises aggregate welfare, it raises *everyone's* welfare; likewise, whenever a social choice reduces aggregate welfare, it reduces *everyone's* welfare.[14] To confirm this, consider a social choice that raises aggregate welfare. A fortiori, it must raise someone's welfare. But, by construction of the symmetric setting, everyone is affected identically. Hence, if the welfare of one individual rises, everyone's welfare must rise. (For a social choice that reduces aggregate welfare, the reasoning is, of course, the same.)

We now explain how our conclusion follows immediately from this observation. As just stated, any social choice that is opposed by welfarism in the symmetric case — that is, any choice that reduces aggregate welfare — necessarily is one that reduces everyone's welfare. Hence, whenever a nonwelfarist principle leads to a different social choice from that under welfarism — which is to say, whenever the principle favors a choice that reduces aggregate welfare — everyone will be made worse off.

To illustrate our argument, let us consider the principle of corrective justice, often traced to Aristotle (1980: V.4), under which one who wrongfully injures another must compensate his victim. This principle would favor, say, a legal regime under which victims may sue injurers for damages. Suppose that the injuries in question involve some sort of simple nuisance: Homeowners may dispose of leaves by burning them, causing smoke to interfere with their neighbors' enjoyment of their yards. Or parties may cause noise. Or renovations may cause incidental damage or disruption. Moreover, because we wish to examine precisely symmetric settings, we should suppose that each person is

---

[13]See also Kaplow and Shavell (1999).

[14]As we state in note 6, we are confining attention to welfarist approaches under which higher welfare is deemed better, not worse. In comparing a nonwelfarist approach to a welfarist one, there is no further need to choose among welfarist approaches because all of them (regardless of how they address the question of distribution) will favor the same choice in a symmetric setting: Either everyone is better off (if total welfare is higher), everyone is worse off (if total welfare is lower), or everyone is indifferent (if total welfare is the same); no distributive judgment is ever required.

equally often in the position of injurer and victim and that each would cause and suffer equal injury.

Comparing corrective justice and pure welfarism, there are two cases of interest. First, both evaluative approaches may favor legal liability — corrective justice by assumption and welfarism perhaps because the prospect of liability deters harmful activity. When nonwelfarism and welfarism align, there is nothing further to consider.

Second, and more interestingly, there may be a divergence. In this example, disagreement would arise when welfarism opposes liability, perhaps because liability does not deter harmful activity and but does entail administrative costs. Now, in this case, since each person is equally often an injurer and a victim, and since by assumption total welfare is lower under the rule favored by corrective justice, it must be that each individual's welfare is lower under corrective justice. This example thus illustrates our general claim about the conflict between nonwelfarist principles and the Pareto principle.

B. *Generality*

We offer three remarks regarding the broad applicability of the foregoing demonstration that, in the symmetric case, everyone is made worse off whenever a social choice is influenced by a nonwelfarist principle, in disagreement with a welfarist approach.

First, for virtually any nonwelfarist principle imaginable, one can construct symmetric cases within the domain of the principle, and thus the present demonstration will be applicable.[15] Thus, if instead of corrective justice we were to consider the principle that promises should be kept and focus on a situation in which one person may break a promise to another, we could construct a symmetric case simply by joining this situation with another that is identical except that the two parties' roles are reversed. It should be apparent that this method of construction is quite general.[16] Hence, virtually any nonwelfarist principle — which we of course assume will sometimes disagree with welfarism — will in some instances favor social choices under which everyone is made worse off.[17]

---

[15]As stated in note 12, an exception must be made for principles concerned purely with distribution, to which the present argument is inapplicable.

[16]Even if one wished to consider cases involving one party's death, one could still construct a symmetric case by supposing that this situation arose merely with a probability (less than fifty percent), with an equal probability that the situation with reversed roles arose.

[17]We also note that the unlikelihood of actually being in a symmetric setting is immaterial because our argument is entirely conceptual. As we elaborate in section V, we appeal here to logical consistency in moral argument (and, with regard to the present demonstration, to the special status of the symmetric case in many moral frameworks). Moreover, we observe that there is in reality rough symmetry with respect to the application of many moral principles. With regard to corrective justice, we are all prospective injurers and victims with regard to driving and many other realms of behavior. Likewise, the promise-keeping principle governs activity in which we all frequently engage, both as promisors and as promisees.

Second, our demonstration applies not only to purely nonwelfarist principles (for example, that promises should be honored even if the heavens would fall) but also to mixed views under which consequences for welfare also receive some weight. To confirm this we note that, as long as some weight is given to factors unrelated to welfare, there will be cases in which the mixed view will lead one to make a choice different from that under welfarism (and in those cases, everyone will be made worse off). The reason is that one can imagine cases in which considerations of welfare oppose the choice favored by the nonwelfarist principle, but by an amount that is less than the weight given to the nonwelfarist principle. Only if the nonwelfarist principle were given literally no weight could this sort of argument be avoided.[18]

Third, despite our alternative argument in section IV, and also our analysis in section V indicating that the foregoing demonstration should suffice, we observe that there is an important sense in which our result for the symmetric case has direct implications for asymmetric settings. The reason is that the characteristic of nonwelfarist principles that causes everyone to be made worse off in symmetric situations remains present in asymmetric situations, even though not everyone will ordinarily be made worse off. To reinforce this point, observe that the core difference between symmetric and asymmetric settings is that distributive effects arise in the latter. Yet, as we note in section II.C, distributive effects may be fully taken into account under welfarism; indeed, any view regarding the distribution of well-being is consistent with welfarism. Hence, merely introducing distributive effects cannot be a basis on which nonwelfarist principles could gain relative ground, so to speak. That is, the nonwelfarist principles under consideration must differ from welfarism in some manner *other than* how they address distribution; moreover, it must be such other difference that gives rise to the conflict with both welfarism and the Pareto principle in the symmetric case.[19] That other difference — the inherent feature of the nonwelfarist principle that produces the conflict in the symmetric case — will still be present in asymmetric cases.[20]

IV. PARETO CONFLICT IN GENERAL

Our second demonstration of the conflict between nonwelfarist principles and the Pareto

---

[18]To reinforce this point, it is helpful to keep in mind that the symmetric cases under consideration include literally every conceivable situation, modified (as we describe in the preceding paragraph) to create a symmetric setting. Hence, all manner of variation in individuals' preferences and opportunities, the nature of private and public institutions, and so forth is captured. To avoid conflicting with welfarism — and thus with the Pareto principle — in all such situations, the nonwelfarist principle would have to favor the same social choice as welfarism in every instance. But if this is the case, it seems evident that the nonwelfarist elements are being accorded no weight. (We set aside purely distributive nonwelfarist theories, see note 12, and also nonwelfarist principles used solely as tiebreakers, which we do not take to be of any real importance.)

[19]As we indicate in note 12, we set aside purely distributive principles in the present section.

[20]For a related argument, see the construction in note 23 in which an asymmetric case is translated into a comparison of choices in which distributive effects are eliminated.

principle is more general than the first, in that it does not make reference to symmetric settings. However, the demonstration is also more abstract, and we ask the reader's forbearance in studying it.[21]

Consider any nonwelfarist principle. (For concreteness, one may again think about the principle of corrective justice, under which wrongdoers are required to compensate their victims.) Suppose, moreover, that this principle is given some weight; that is, other things equal, it sometimes affects the social decision. Then, we can imagine two situations — call them *Fair* and *Unfair* — that have the following two characteristics: First, each individual is just as well off in *Fair* as he is in *Unfair*. Second, one situation — *Fair* — is definitely more fair than the other, *Unfair*, and hence it is deemed morally superior according to the nonwelfarist principle under consideration. (Although the stated relationship between the two regimes may appear most unlikely, keep in mind that this is merely a hypothetical example used for conceptual purposes; the only question is whether it can be imagined, and surely it can be, as we elaborate in the accompanying footnote.[22])

Next, consider a slightly modified unfair situation, which we will call *Unfair-II*, that is identical to *Unfair* except in one respect: There is a tiny savings (for example, in some sort of administrative cost) in *Unfair-II*, which is distributed uniformly per capita. Now, if fairness has any real weight, it must be true that *Fair* is deemed superior overall to *Unfair-II*: After all, *Fair* was definitely superior to *Unfair*, *Unfair-II* is every bit as unfair as *Unfair*, and the cost advantage of *Unfair-II* over *Unfair*

---

[21]The present argument is proved formally in Kaplow and Shavell (2001).

[22]To show that there can exist situations that are equal for each individual in welfare but that differ regarding some principle of fairness, it is helpful to consider that the *Unfair* situation may differ the *Fair* one in multiple ways. Some differences make *Unfair* less fair according to the nonwelfarist principle. Other differences can affect the distribution of income in a manner that just happens to offset any effect of the unfairness on individuals' well-being. This offset could be direct: An unfair regime that did not provide for corrective justice by allowing suits would seem to favor injurers over victims, but that same unfair regime might instead require injurers to pay higher fines to the state, with the revenue used to fund social insurance or some other program that benefits victims. Also, the offset could be unrelated: An unfair regime might also have a different tax or regulatory rule that had an offsetting distributive effect. As stated in the text, it is only necessary that we can imagine such a case. And, of course, we can imagine just about anything, as long as it is not internally contradictory.

An additional way to understand that the postulated construction is always possible with a nonwelfarist principle is to consider the opposite assumption, that there exists no possible case in which (a) all individuals are equally well off in two different situations and (b) the nonwelfarist principle deems one to be strictly superior to the other. Under this opposite assumption, it can be shown that the nonwelfarist principle is really a welfarist principle, which is to say that it is possible to determine any possible social choice by considering only individuals' levels of well-being. See note 8. First, suppose that in some case in which (a) holds that (b) does not hold, which is to say that the purportedly nonwelfarist principle views the choice as a matter of indifference. In that event, it would be sufficient to know the configuration of individuals' well-being in order to know what social choice is dictated by the principle — that is, the choice under the principle depends only on welfare . But such a principle is, by definition, welfarist. Second, suppose (impossibly, given the foregoing paragraph) that no pair of situations could have feature (a). This implies that each possible situation involves a different configuration of individuals' well-being. But in that case, any possible array of social choices could be stated knowing only the configuration of individuals' well-being (since each configuration is here assumed to be unique). Accordingly, any such principle of choice would be welfarist.

was stated to be tiny.

Finally, observe that everyone is worse off in *Fair* than in *Unfair-II*. (This is because everyone is equally well off in *Fair* and in *Unfair*, while everyone is worse off in *Unfair* than in *Unfair-II*.) Hence, we have demonstrated that there can exist situations in which the nonwelfarist principle favors a social choice, here *Fair* over *Unfair-II*, under which everyone is made worse off.[23]

Before closing we comment briefly on an implicit assumption in the preceding argument, namely, that in changing from situation *Unfair* to *Unfair-II*, the posited nonwelfarist principle continues to favor *Fair*. For concreteness, using our corrective justice example, one could suppose that, compared to *Fair*, both *Unfair* and *Unfair-II* (which are identical in fairness terms) involve a major sacrifice of corrective justice while the only difference between *Unfair* and *Unfair-II* is that the latter involves a trivial administrative cost savings relative to the former, perhaps a penny per person. One might nevertheless object to the generality of this example because of the possibility that a nonwelfarist principle may never have more than a little weight. But, however small that weight might be, we can always imagine an even smaller increment to welfare in moving from *Unfair* to *Unfair-II*. Perhaps the welfare gain is only the equivalent of a peanut, one that must be divided equally among all in the population.[24] The idea, we hope, is clear: As long as the nonwelfarist principle is given some real weight, however slight, it is possible to construct a case in which our demonstration applies and hence one in which the nonwelfarist principle will favor a social choice under which everyone is made worse

---

[23]Another way to reach this conclusion is with the following four-step construction (which is a heuristic sketch, not a proof). (1) Suppose that a nonwelfarist principle conflicts with welfarism in at least one given case. In particular, suppose that the nonwelfarist principle favors regime *N* and the welfarist principle favors regime *W*. By assumption, welfare is higher in *W*, and *N* is still superior under the nonwelfarist principle because it is more fair in some sense, and to an extent judged more important than the overall welfare advantage of regime *W*. (2) Construct *W′′* from regime *W* as follows: Maintain the same degree of (un)fairness according to the nonwelfarist approach and also the same degree of total welfare, but redistribute income such that the resulting distribution of well-being in *W′′* is the same as that in *N*. (If, for example, the redistribution reduces inequality and one's welfare assessment favors a more egalitarian outcome, the overall adjustment in generating *W′′* from *W* would reduce total income sufficiently to keep total welfare constant.) (3) Since *N* is better according to the nonwelfarist moral principle than *W′′* by the same amount that it was better than *W*, and since welfare is no higher in *W′′* than it was in *W*, it must be that *N* is deemed overall superior to *W′′* under the nonwelfarist principle. (4) However, *W′′* has higher total welfare than *N* and also the same distribution of welfare as *N* (that is how *W′′* is constructed). Hence, everyone must be worse off in *N* than in *W′′*, even though *N* is judged to be superior overall under the nonwelfarist principle under consideration.

[24]A further implicit assumption is that the change entailed in moving from *Unfair* to *Unfair-II*, which consists solely of distributing, say, a peanut to the population, does not fundamentally change what made regime *Unfair* inferior in the first instance to regime *Fair* as a matter of the given nonwelfarist moral principle. Since, however, we are referring to a nonwelfarist moral principle and, moreover, are supposing that the only modification involves solely an effect on welfare, this assumption seems plausible. Thus, nothing in the concept of corrective justice suggests that the principle can change fundamentally depending on whether there is one more peanut in the pool of social resources. (If the nonwelfarist moral principle, however, were a mix of a welfarist approach and, say, some deontological principle solely as a tiebreaker, this argument would be inapplicable, as we discuss in the following note.)

off.[25]

V. IMPLICATIONS

In sections III and IV, we demonstrate that any nonwelfarist principle will deem morally correct some social decisions that would make everyone worse off. Of course, few if any serious nonwelfarist principles always — or even usually — have this consequence; indeed, it is frequently observed by consequentialists and nonconsequentialists alike that there often will be alignment between deontological principles and welfarism with regard to the social choices required by each.

Nevertheless, the soundness of a theory must be assessed with regard to all of its implications. In this section, we elaborate on this familiar point and consider as well why the conflict that we establish constitutes an especially serious problem for many nonwelfarist theories. (Recall, however, that we do not argue here for the proposition that it is problematic for morality to require making everyone worse off.) In addition, we discuss the senses in which nonwelfarist principles are relevant under welfarism. This consideration will further clarify the meaning of welfarism and also help to reconcile our argument with widely-held moral intuitions, which often favor nonwelfarist principles.

A. *Logical Consistency*

We show that any principle that is not purely welfarist — that attaches any importance to nonwelfarist considerations — sometimes favors social choices that would make everyone worse off. Hence, as a matter of logical consistency, if one endorses the Pareto principle, one cannot give any weight to nonwelfarist principles. If a theory, in some part of its domain of application, contradicts a principle to which one subscribes, the theory has thereby been rejected.

This basic point is familiar in moral philosophy. For example, it is a hallmark of attacks on consequentialist views, utilitarianism in particular. Critics commonly present situations — sometimes bizarre — in which utilitarianism favors a result that our moral intuitions suggest is incorrect.[26] It is

---

[25]The reader may note that our present argument can be circumvented by supposing that a nonwelfarist principle never receives any weight except when there is a precise tie on welfarist grounds. (Perhaps a full cost-benefit analysis reveals that summing all the effects of a social choice on millions of people, taking distribution into account, results in a perfect tie, down to the last penny.) If this were the only role of nonwelfarist principles, however, they would hardly be worth discussing.

[26]In this regard, we observe that in important respects our demonstrations, though hypothetical, go more to the core of the principles in question than do some of the examples offered in criticism of utilitarianism. Notably, in our symmetric setting, we simply imagine an ordinary, typical example involving the nonwelfarist principle, when it is in conflict with welfare, and then suppose that we combine it with a second, identical case in which roles are reversed. Thus, even if one believes, for example, that a principle may be valid in ordinary but not extraordinary situations, our demonstration remains applicable. We also note that when the method of testing a result involves the application of moral intuition, the fact that there arise conflicts in bizarre hypothetical situations should be far

recognized that utilitarians may not reply simply by creating an ad hoc exception for such cases, leaving their theory otherwise intact.[27] Such a reply is inadequate because the situations presented, if they indeed show utilitarianism to be problematic in its stated domain (a highly contested question), reveal a fundamental problem with the theory and its underlying motivation.[28]

Another way to view the point about logical consistency is to recognize that there must be reasons to adopt nonwelfarist principles. Whatever the reasons may be, they must explicitly contemplate that effects on welfare are not always decisive (or, in pure nonwelfarist theories, never relevant). As we show, all such theories therefore endorse some social choices that make everyone worse off. Accordingly, they entail the principle that it can be morally desirable for society to act so as to make everyone worse off. This conclusion holds regardless of whether other (indeed most) applications of the nonwelfarist theory involve such a consequence.

---

less surprising or problematic than conflicts in core settings such as those that we consider.

[27]That is, to defend utilitarianism against such criticism, it is necessary either to show that the theory does not have the alleged implications or that, although it does, further analysis reveals that they are not morally unacceptable.

[28]Relatedly, we reject an argument that some have suggested to us, namely, that the conflict we adduce may be avoided by simply deeming the nonwelfarist principle to be inapplicable — more precisely, to reverse its conclusions — in those cases in which there would be a conflict with the Pareto principle. Such an approach is at least prima facie problematic for it calls into question the underlying rationale for the nonwelfarist theory. (Compare Sidgwick's (1907) discussion of the modification of moral principles to avoid conflicts with other moral principles.) This is particularly so given that the welfare consequences that give rise to the conflict are deemed irrelevant or insufficiently important by the nonwelfarist theories.

In addition, even if one modified the principles, as long as the modified version continued to be nonwelfarist, one would have to consider whether our two demonstrations still apply. Regarding the first, one could deem the principle inapplicable to all symmetric cases (creating the difficulty discussed in section V.B, below). Even then, the theory would have an odd character. Is it really the case that a moral principle might plausibly be deemed wholly irrelevant in a symmetric case but decisive — despite possibly large adverse welfare consequences — in another case that was ever so slightly asymmetric (perhaps one person was better off by one cent rather than worse off by one cent)?

This point also relates to our second demonstration, where we show that if a nonwelfarist principle is given any weight in any situation (however asymmetric), one could in principle find a closely related case in which a Pareto conflict results. And we note that since this argument is logically applicably to *any* nonwelfarist principle, it is thus applicable to any nonwelfarist principle that has been modified in any attempt to avoid Pareto conflicts. (There is no contradiction here with the idea that it is possible to modify a nonwelfarist principle to avoid all Pareto conflicts, because one can modify it enough that it becomes a purely welfarist principle.) Now, as our discussion at the end of section IV and in note 25 indicates, this analysis assumes that the nonwelfarist principle exhibits some minimal level of consistency and that it gives nontrivial weight to nonwelfarist considerations — not great weight, but more than just as a mere tiebreaker. But we suppose that serious proponents of nonwelfarist principles would grant these minimal assumptions, particularly when the sort of modification to situations used in our demonstration (such as giving everyone a fraction of a peanut) is independent of the basis for the nonwelfarist principles.

B. Importance of the Symmetric Case

In addition to the foregoing argument that appeals to the need for moral principles to be maintained in a logically consistent manner, we suggest that our demonstration in the symmetric case is particularly important. Our analysis has two steps. First, we explain that many moral theories — which are believed to provide a foundation for particular nonwelfarist principles — can be understood explicitly or implicitly as requiring that the validity of moral principles be tested in symmetric settings. This conclusion follows because the moral theories that we discuss in this regard are aimed at achieving impartiality in the assessment of moral rules and symmetric settings (often uniquely) embody impartiality. Second, it follows that those endorsing the theories under consideration face a challenge because, when in the symmetric case (where all moral principles must be tested), any nonwelfarist principle that is endorsed will always make everyone worse off when the principle disagrees with welfarism.

In advancing the present argument, we have in mind two particular types of moral theories. One is exemplified by the Golden Rule and Kant's categorical imperative. The Golden Rule commands an individual to treat another as he would like the other to treat himself. This command can be understood as telling us to construct a symmetric case in the manner that we discuss in section III: Although we begin with a given, asymmetric situation (referring to how one should treat another), we are commanded to add the mirror-image situation (referring to how the other should treat the one if the situation were reversed); taken together, we have constructed a symmetric case.[29]

Our interpretation of Kant's categorical imperative is similar. This formulation tests the morality of proposed principles by asking whether they would retain appeal if we generalized them to all situations, particularly including those in which we ourselves would be in the opposite position from our own, actual situation. Taken literally, it is hardly obvious that Kant's test requires us to consider symmetric settings, but following the suggestions of prior writers, we argue that when properly interpreted Kant's test does.[30] To understand why, consider, for example, a situation in which there is a strong person and many weak people. Initially, the strong person, in his interaction with a particular weak person, may favor the principle that "might makes right," because this would be to his advantage. Now suppose we insist that he generalize this principle to all situations. The strong person would continue to benefit, all the more so, because under the generalization he prevails over everyone and not

---

[29]Arguably, the statement in the text is incorrect because the Golden Rule can be interpreted literally as requiring us to derive the proper treatment by looking *only* at the reverse situation. But this view is no more neutral than the selfish perspective with which one begins, before considering the Golden Rule. For example, suppose the question is whether an individual should endure enormous pain or expense to help another trivially. If the individual considers only the reverse situation, in which the question is whether he would like the other to undergo great sacrifice to benefit himself, he would (selfishly) answer affirmatively. Then the implication would be that the correct moral rule is that everyone is required to make great sacrifices whenever they would produce even trivial benefits for others. Clearly, reversing the direction of bias (from excessive self-favoritism to excessive self-disregard) is not the purpose of the Golden Rule and is not what we take to be the intention of those who invoke it favorably.

[30]See, for example, Sidgwick (1907) and Hare (1997).

only over the one weak person now before him. Yet "might makes right" is just the sort of rule based on self-interest that the categorical imperative is understood to reject. In order to obtain this expected and appropriate outcome, it is necessary to reconstruct the generalized setting that one considers. Thus, one might suppose (contrary to fact) that each person equally often finds himself in situations in which he is strong and those in which he is weak. Now the principle "might makes right" indeed appears unattractive, for familiar reasons. Notice, however, that in order to align the imagined hypothetical scenario with our understanding of the categorical imperative, it is necessary to construct precisely a symmetric setting.

The second type of moral theory that involves the use of symmetric settings is one that explicitly puts individuals in a position of ignorance, such as that posed by Harsanyi (1953) and Rawls (1971). Here it is recognized that an impartial view is achieved by supposing that each individual is symmetrically situated.[31]

Moreover, even among those who do not advance a particular construct such as the categorical imperative or some version of a veil of ignorance, there is widespread acceptance of the view that impartiality is a necessary condition for any moral principle. And, as the foregoing discussion indicates, symmetric settings capture the notion of impartiality in a fundamental manner.

To complete our discussion, we combine the present argument that important moral frameworks require moral principles to be tested in symmetric settings with our previous conclusion in section III that nonwelfarist principles conflict with the Pareto principle in symmetric settings. Suppose that a theorist endorses one of the moral frameworks under consideration and also believes that some nonwelfarist principle is correct, in particular, in cases in which it conflicts with pure welfarism. To endorse such a moral framework means that a necessary condition for any moral principle to be acceptable is that it pass muster — be deemed correct — in the scenario contemplated by that framework. This scenario, in turn, should be understood as the symmetric setting. But how does the posited nonwelfarist principle fare in the symmetric setting in those cases in which the principle differs from welfarism? The answer is that following the principle makes everyone worse off in *every* such instance. Thus, to accept that the nonwelfarist principle passes the test is hold the view that, in all of the cases under consideration, the morally correct result involves making everyone worse off.

Accordingly, if one accepts both a nonwelfarist principle and the Golden Rule, the categorical imperative, the veil of ignorance, or other related constructs, one thereby accepts as a necessary, defining feature of one's nonwelfarist principles that it is morally preferable for everyone to be made

---

[31]See, for example, Rawls (1980: 529): "[T]he background setup of the original position . . . situates [the parties] symmetrically [so that persons are not] advantaged or disadvantaged by the contingencies of their social position, the distribution of natural abilities, or by luck and historical accident over the course of their lives." Other approaches to testing moral principles are similar to the construct of the original position in this respect. For example, Lewis (1946) suggests that impartiality can be captured by imaging that we occupy each role in society seriatim. If each individual is seen in this light, the situation is likewise symmetric.

worse off. In other words, one must not merely accept the possibility of conflict with the Pareto principle. One must comprehensively endorse it — that is, whenever and to whatever extent the endorsed moral principles differ from pure welfarism.

C. Implicit Importance of the Pareto Principle under Certain Moral Theories

It is, of course, entirely possible in logic to endorse nonwelfarist theories despite their conflict with the Pareto principle. Whether some or all individuals may be made worse off, whether slightly or substantially, may be deemed to be of secondary importance or entirely irrelevant. (And, as we note previously, we do not argue directly for the Pareto principle in this article.) Nevertheless, many nonwelfarist moral theories appear to be grounded in rationales that make at least an implicit appeal to some individuals' well-being. To that extent, the conflict with the Pareto principle raises an internal difficulty for such theories. After all, if one is truly motivated by a concern relating to some individuals' well-being, one has probably made a mistake if one is led by the concern to endorse as morally preferable a choice that makes everyone worse off.

This point is fairly clear with regard to nonwelfarist principles that seem motivated by concerns about protecting or assisting potential or actual victims. For example, the principle of corrective justice seems concerned about victims in demanding that they be compensated by their injurers. The sanctity of promises seems predicated, at least in part, on the impropriety of harming promisees (and also on damage to the institution of promise-keeping, which would tend to be to everyone's detriment). And so forth.[32] Yet our demonstration that all nonwelfarist principles will make everyone worse off in certain situations means that it cannot be an inherent feature of any nonwelfarist principle that some individuals are helped. To express our point in another way, many principles require us to be fair *to* certain individuals, but one must ask to whom one is being fair if every potential subject of concern is made worse off by a posited principle of fairness.

A similar tension also seems to underlie nonwelfarist principles that are predicated on ideals of

---

[32]As another sort of example, in defining welfare for normative purposes, some suggest that certain preferences be disregarded — such as negative other-regarding preferences — presumably (at least in part) out of a concern for those negatively regarded. Yet the logic of our argument suggests that adopting social choices that so disregard such preferences will sometimes make everyone worse off, including in particular the objects of such preferences. (We also note that, even setting aside our demonstrations regarding the Pareto principle, the idea that social choices should be made as if such preferences do not exist can be counterproductive: When the satisfaction of objectionable preferences is treated as nonexistent, then those who have such preferences are deemed to be worse off because some of their preference satisfaction is ignored; accordingly, under many moral theories, such individuals would be entitled to more resources. Moreover, if there is no way to control how these individuals expend their resources, they may well choose to expend some of their greater share precisely to satisfy further their objectionable preferences. These points supplement our claim in the text in suggesting that, if in fact one is motivated by a concern for some individuals, say, potential victims of some sort of behavior, then one must take a welfarist approach, for once welfare is ignored or downplayed, it is possible that the subjects of concern will be hurt rather than helped.)

personal freedom and autonomy or on libertarian conceptions of individual rights.[33] The reason for the difficulty is that consistent adherence to these (like other) nonwelfarist principles will sometimes favor social choices that would be rejected by every individual if in fact each was freely permitted to make his own choice about the matter. Given that individuals would unanimously reject the choice if given the opportunity, it would appear that insisting upon the choice does not promote but rather interferes with their freedom and liberty.[34]

D. Instrumental Relevance of Nonwelfarist Principles under Welfarism

Suppose that one does endorse the Pareto principle and, accordingly, rejects all nonwelfarist principles as proper guides for social decisionmaking. It nevertheless remains true that nonwelfarist principles are indirectly and importantly relevant under welfarism for a variety of reasons. We sketch these reasons briefly both to clarify the conceptual relationship between nonwelfarist principles and welfarism and also because these considerations help to reconcile a welfarist stance with the admitted appeal of many nonwelfarist principles to our moral intuitions.

First, welfare may sometimes be enhanced by instilling nonwelfarist principles in the form of common morality because such principles usefully guide individual behavior. Individuals have only limited ability to determine truly optimal actions, whereas moral principles can serve as useful guides. Moreover, individuals often will be inclined to act opportunistically, advancing their own interests at the expense of others, whereas individuals often will be led to act socially properly if they are rewarded for doing so by approbation and feelings of virtue and if they would be punished were they to act badly by disapprobation and feelings of guilt. These advantages of common morality help to reconcile welfarism with nonwelfarist principles and the moral intuitions that motivate them, and they are, of course, what

---

[33]Here, one might include such diverse theorists as Dworkin (1977), Nozick (1974), Rawls (1971), and Williams (1981).

[34]A well-known paper of Sen (1970) illustrates this problem. Sen suggests that adherence to the Pareto principle may be problematic because it conflicts with a certain notion of individual liberty, under which each individual's preference in some domain (say, which books to read) should be socially decisive with regard to his activity in that domain. A Pareto conflict arises in Sen's construction because he considers a case in which each of two individuals actually cares more about what the other person reads than about what he himself reads; thus, following Sen's principle of liberty, which requires that each read the book that he himself favors for himself, makes both individuals worse off than if each read the book that the other person would prefer him to read. Observe that in this case, Sen's principle implicitly entails the government's forbidding individuals from voluntarily waiving their rights in exchange for each others' waivers; that is, the notion of liberty that Sen advances (which he calls "liberalism") actually constitutes a limitation on individuals' freedom to be in a regime that would be to their mutual benefit. Thus, rather than constituting an argument against the Pareto principle, we believe that Sen's argument shows how certain notions of liberty can be internally inconsistent. (We further note that in one of Sen's (1992, pp. 144-16) subsequent defenses of upholding individual rights at the expense of welfare, he cites as motivations the need to limit governmental abuse of power and difficulties of enforcing contracts between such parties, arguments that are instrumental, welfarist justifications for rights of the sort that we note in the following section.)

underlies well-known two-level moral theories.[35]

Second, and related, certain nonwelfarist principles — notably, individual rights — might be usefully incorporated into rules, such as constitutions and other legal forms, in order to constrain the behavior of agents who cannot be trusted to use their discretion to maximize social welfare. Such principles are often easy to communicate, apply, and enforce precisely because they are relatively simple in character and thus do not require a full and often complex and contingent analysis of all the consequences of a given act. It may be conceded that rules that embody such principles will sometimes command actions that, all things considered, should properly be viewed as socially undesirable. Nevertheless, the expected consequences of imposing a rule may be more favorable than those in the absence of the rule.

Third, when attempting to determine what social choice would advance welfare, nonwelfarist principles might be employed for purposes of preliminary analysis as convenient proxies. Those assessing possible choices might have paths of inquiry illuminated by applying familiar nonwelfarist principles (which, as just noted, are often easier to apply than a full consequentialist analysis). Also, if a quick decision must be made and there is little preexisting analysis or evidence to draw upon, it sometimes will be sensible to rely on nonwelfarist principles to advance welfare. After all, in ordinary cases, many nonwelfarist principles often align closely with welfarism. This use of nonwelfarist principles, like the preceding two, is clearly instrumental: The principle is employed when — and because — it is expected to advance welfare.

Fourth, nonwelfarist principles will be relevant under welfarism to the extent that advancing such principles itself is part of individuals' welfare. Particularly when nonwelfarist principles are elements of common morality, individuals will tend to be upset when they observe violations of the principles, and they will tend to experience pleasure if the principles are obeyed. For example, principles of corrective and retributive justice are widely held and govern our social interactions; accordingly, many individuals are unhappy when those who cause harm are not held accountable, and they are pleased when violators receive their just deserts. Under many conceptions of welfarism, this set of effects on individuals' well-being, like any other, would be taken as a component of welfare. Hence, serving nonwelfarist principles in practice can, to an extent, directly contribute to individuals' well-being. Clearly, this too is an instrumental use of nonwelfarist principles; their importance in this regard is an empirical matter concerning individuals' actual preferences, not something determined by the philosophical soundness of underlying moral principles.

---

[35]Among those commonly associated with developing this approach are Hume (1739, 1751), Austin (1832), Mill (1861), and Sidgwick (1907) — although Hare (1981) traces the idea to Plato and Aristotle. Subsequent works include Brandt (1979, 1996), Hare (1981), Harrod (1936), Rawls (1955), and Sartorius (1972).

VI. CONCLUSION

We have shown that all nonwelfarist moral principles sometimes favor social choices that make everyone worse off. In symmetric cases, nonwelfarist principles always favor choices that reduce every individual's well-being when the favored choices differ from those prescribed by welfarism. This result is particularly important if one accepts the approach of many general moral theories under which, in essence, moral principles must tested as if in a symmetric setting. We also offered a more general demonstration of the conflict between nonwelfarist principles and the Pareto principle. We then explored how requirements of logical consistency make it impossible to endorse nonwelfarist principles without simultaneously endorsing that it may be morally correct for society to make choices that are to everyone's detriment. Finally, we noted a number of ways in which nonwelfarist principles may nevertheless be instrumentally important under welfarism, which helps to explain why our moral intuitions often seem to favor nonwelfarist principles.

We acknowledged at the outset that it seems implausible a priori that the Pareto principle could help us to choose among moral principles since most such principles are addressed to situations involving conflicting interests. When the question appears to involve whether to favor one type of party or another, it does not seem that any plausible principle would favor social choices that make everyone worse off. On reflection, however, such reasoning is mistaken. In general, one can construct symmetric cases out of asymmetric cases. Then any principle that helps one of necessity benefits all. Purely welfarist principles always benefit everyone. Nonwelfarist principles, in contrast, always make everyone worse off when they conflict with welfarism. Likewise, even without constructing symmetric cases, we show that, whenever a nonwelfarist principle is given weight, one can identify cases in which consistent application of the nonwelfarist principle favors a choice that reduces every individual's well-being.

This conflict does not, as a logical matter, prove nonwelfarist principles to be unsound. Furthermore, as we stated at the outset, we do not here attempt to defend the Pareto principle. Nevertheless, it seems apparent that many do endorse it. We suspect that its appeal lies in part in the clarity of the situations to which it applies: The Pareto principle favors a social choice only if everyone stands to gain; when some individuals gain and others lose, as is more typically the case, the principle is inapplicable. Whatever the reason for the appeal of the Pareto principle, we believe that the conflict that we adduce between this principle and nonwelfarist moral prescriptions does raise questions. In particular, it suggests that a cogent defense of any nonwelfarist principle should include arguments that explain why the principle is justified in light of its concomitant endorsement of the idea that it can be morally appropriate for society to make every one of its members worse off.

References

Aristotle. 1980. *The Nicomachean Ethics*. Rev. ed. Translated by David Ross and edited by J.L. Ackrill and J.O. Urmson. New York: Oxford University Press.

Austin, John. 1832. *The Province of Jurisprudence Determined*. Wilfrid E. Rumble, ed. Cambridge: Cambridge University Press (1995).

Brandt, Richard B. 1979. *A Theory of the Good and the Right*. Oxford: Oxford University Press.

Brandt, Richard B. 1996. *Facts, Values, and Morality.* Cambridge: Cambridge University Press.

Dworkin, Ronald. 1977. *Taking Rights Seriously*. Cambridge, Mass.: Harvard University Press.

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press.

Hare, R.M. 1981. *Moral Thinking: Its Level, Method, and Point*. Oxford: Oxford University Press.

Hare, R.M. 1997. *Sorting Out Ethics*. Oxford: Oxford University Press.

Harrod, R.F. 1936. Utilitarianism Revised. *Mind* 45: 137-56.

Harsanyi, John C. 1953. Cardinal Utility in Welfare Economics and in the Theory of Risk-Bearing. *Journal of Political Economy* 61: 434-35.

Hume, David. 1739. *Treatise of Human Nature*. Buffalo: Prometheus Books (1992).

Hume, David. 1751. *An Enquiry Concerning the Principles of Morals*. Tom L. Beauchamp, ed. Oxford: Oxford University Press (1998).

Kant, Immanuel. 1785. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor, Cambridge: Cambridge University Press (1997).

Kaplow, Louis, and Steven Shavell. 1999. The Conflict Between Notions of Fairness and the Pareto Principle. *American Law and Economics Review* 1: 63-77.

Kaplow, Louis, and Steven Shavell. 2001. Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle. *Journal of Political Economy* 109: 281-86.

Kaplow, Louis, and Steven Shavell. 2002. *Fairness versus Welfare*. Cambridge, Mass.: Harvard University Press.

Lewis, Clarence I. 1946. *An Analysis of Knowledge and Valuation*. La Salle, Ill.: Open Court Publishing Co.

Mill, John Stuart. 1861. *Utilitarianism*. Edited by Roger Crisp, Oxford: Oxford University Press (1998).

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Nussbaum, Martha, and Amartya Sen, eds. 1993. *The Quality of Life.* Oxford: Oxford University Press.

Rawls, John. 1955. Two Concepts of Rules. *Philosophical Review* 64: 3-32.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Rawls, John. 1980. Kantian Constructivism in Moral Theory. *Journal of Philosophy* 77: 515-72.

Rawls, John. 1982. Social Unity and Primary Goods. In *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams. Cambridge: Cambridge University Press.

Ross, W.D. 1930. *The Right and the Good*. Oxford: Oxford University Press.

Sartorius, Rolf. 1972. Individual Conduct and Social Norms: A Utilitarian Account. *Ethics* 82: 200-18.

Scanlon, Thomas M. 1998. The Status of Well-Being. In *Tanner Lectures on Human Values*. Vol. 19. Salt Lake City: University of Utah Press.

Sen, Amartya. 1970. The Impossibility of a Paretian Liberal. *Journal of Political Economy* 78: 152-57.

Sen, Amartya. 1992. Minimal Liberty. *Economica* 59: 139-59.

Sidgwick, Henry. 1907. *The Methods of Ethics*. Seventh edition. Indianapolis: Hackett Publishing Company (1981).

Sumner, L.W. 1996. *Welfare, Happiness, and Ethics*. Oxford: Oxford University Press.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers, 1973-1980.* Cambridge: Cambridge University Press.