The Problematics of the Pareto Principle

Daniel A. Farber[1]

The Pareto Principle asserts in one form that an outcome that is unanimously preferred by individuals should be chosen by society; or in another form that an outcome should be chosen if it is preferred by at least one individual and the remaining members of society are indifferent. It is little wonder that this principle, which has the ring of a self-evident truth, has been the "gold standard" for law and economics. The normative foundations of the Pareto Principle are more complex than they may appear. Despite its intuitive appeal, the Pareto Principle has limitations that are irrelevant in some spheres such as corporate law, but that may have serious import for fields such as constitutional law, family, and bioethics. Some of these limitations stems from conflicts between autonomy and welfare norms; others from bounded rationality; still others from ambiguities in the definition of "preference" or from subtle conflicts such as the Sen Paradox between the Pareto Principle and other seemingly self-evident norms. The Pareto Principle remains a useful rule of thumb, but it should not be followed blindly in all settings.

There can't be many moral standards with more intuitive appeal than the Pareto principle.[2] The basic intuition is simple: If one person prefers a certain outcome and it does not cost anyone else, society should give that person what he or she wants. This intuition seems readily supportable on a variety of grounds. One ground is that society is only the sum of its parts – if one person is better off and no one else is harmed, then society as a whole is better off. A second is that we should honor people's individual autonomy by respecting their preferences: "[a]ll participants would by definition consent to a transaction which left them better off, or as well off as before"; hence, "a moral analysis based on autonomy and consent would approve of

[1]Sho Sato Professor of Law, University of California at Berkeley. Thoughtful comments on an earlier draft were provided by Brian Bix and by participants in a workshop at Chicago-Kent law school.

[2]For definitions of the various Pareto standards, see Brian Bix, Jurisprudence: Theory and Context 181 (2d ed. 1999). Admittedly, the intuition in favor of Pareto may seem stronger to economists than others; so far as I know, there is no empirical evidence on this point.

transactions that were Pareto superior."[3] And a third is that society should care about the welfare of its citizens and seek to promote their well-being.

But although the appeal of the Pareto principle is indeed attractive, its normative status is actually quite complex. To begin with, the idea of a "preference" is by no means as clear-cut as it may seem. In a number of situations, it is not easy to know whether we should focus on the choices people actually make when they rank alternatives in advance (assuming that such rankings in fact exist) or instead on their rankings after the fact, which might be different. Or perhaps we should focus on the rankings they *would have had* if they had engaged in serious discussion of the issue with their fellow citizens, or if they were better at probability theory, or if they were more aware of what is actually good for their own welfare. And are we interested in their rankings of outcomes or their rankings of decisions, which may or may not be consistent with each other?

Moreover, the Pareto principle's normative foundations are also a bit shaky. It is often defended on the basis of its connection with social welfare. But for this connection to hold, the preferences in question must qualify as legitimate indications of individual welfare. As we will see, there are all kinds of reasons why we might not be willing to count the satisfaction of a given preference as a genuine improvement in individual well-being. For the social welfare argument to work, we must also be willing to embrace a purely individualistic concept of social welfare, which is defensible but at least subject to debate. Alternatively, Pareto can defended on the basis

[3]Id. As Bix points out, theorists such as Kant actually have a much narrower concept of autonomy than economists. Id. at 181 n.13.

of the liberal conception of autonomy.[4] Here again, however, the connection turns out to be subject to dispute. It has long been known that liberal autonomy and the Pareto principle may sometimes conflict; and it now appears that the conflict is potentially serious. As a general matter, Pareto stands for a form of paternalism rather than autonomy – we give people what they want (or possibly, what they *should* want) rather than what they have actually chosen.

This essay catalogues the limitations of Pareto. Although these limitations have been discussed somewhere in the literature (sometimes as limits on freedom of contract rather than Pareto), putting them all together highlights just how far from being self-evident and universal the Pareto principle really is. Part I begins with some background on the Pareto principle and its significance. In Part II, we consider the ways in which preferences may be diverge from individual or social welfare, and in Part III we examine situations where preferences are contingent and therefore do not offer a firm basis for social choice. Part IV considers flaws in preference formation due to the limits of human rationality. In Part V, we consider the relationship between the Pareto principle and social choice theory. Finally, Part VI asks the question, what is left of the Pareto principle?

After reading Parts II through V, the reader may well concluded that the Pareto principle is as full of holes as Swiss cheese. Should it be abandoned on that account? I believe that the answer is no (or at least, "not completely"), and that the Pareto principle can be a valuable guideline to social policy. But it should be considered to be a rule of thumb rather than a bedrock normative principle.

---

[4]See Anthony T. Kronman, *Wealth Maximization as a Normative Principle*, 9 J. Leg. Studies 227, 234 (1980)("The constraint imposed by the Pareto principle reflects a concern to protect the autonomy of individual persons. . . .")

I. Introducing the Pareto Principle

The Pareto standard arose out of nineteenth and early twentieth century debates about utilitarianism. When economists abandoned the idea of interpersonal comparisons of utility, they were left with a normative puzzle. Pareto provided the solution, as Bob Cooter explains:

> If utilities cannot be summed, can economists say anything about public policy? This is the problem to which Pareto found a solution that proved far more successful than he could have hoped.. A Paretian analysis, as proposed by its inventor and greatly refined by generations of economic theorists, first assumes that there is an initial distribution of resources, which is given outside the model. Once the initial distribution is described, the analysis proceeds to ask whether any reallocation of resources can make at least one person better off without making anyone else worse off. If the answer is "Yes," the reallocation is a *Pareto improvement.* If the answer is "No," the initial allocation is *Pareto efficient* (also called Pareto optimal"). Starting with an allocation that is inefficient, Pareto efficiency is achieved by reallocating resources until the opportunities for Pareto improvements are exhausted.[5]

We can rephrase the Pareto principle as a comparison of states of the world. If at least one person prefers State *A* to State *B*, and no one has the opposite preference, then society should adopt State *A* rather than State *B*.[6] (One could think of the motion to adopt State *A* as passing by

---

[5]Robert D. Cooter, *The Best Right Law: Value Foundations of the Economic Analysis of Law*, 64 Notre Dame L. Rev. 817, 820-821 (1989)(emphasis in original).

[6]The term "preference" itself is not free from ambiguity. See Lewis Kornhauser, *Preference, Well-Being and Morality in Social Decisions*, 32Journal of Legal Studies 303 (2003).

a vote of 1-0, with the rest of the world abstaining.) Although the Pareto principle only holds under restricted conditions, at least in those conditions it appears to provide an objective standard for ranking one state of the world as unambiguously better than another. (Note that the Pareto principle says nothing about what to do when people have conflicting preferences about states of the world; the fact that one state is Pareto efficient but not the other is irrelevant. Thus, while the Pareto principle does not endorse redistribution of wealth, it also does not condemn redistributions.) This standard seems to have a core of common sense: if Mary wants to see Movie #1 rather than Movie #2, and Bob either agrees with her or doesn't care, then the couple surely should go see Movie #1, which makes her happier and at worst makes no difference to him.

True, the Pareto principle is not everything. Pareto optimality is, according to Amartya Sen, "a very limited kind of success": "[a] state can be Pareto optimal with some people in extreme misery and others rolling in luxury, so long as the miserable cannot be made better off without cutting into the luxury of the rich."[7] Still, Sen says, though there many unappealing states of the world are Pareto optimal, "it has been thought reasonable to suppose that the very best state must be *at least* Pareto optimal."[8] So Pareto optimality seems to be at least a necessary condition of the best outcome, though far from being a sufficient condition.

The Pareto principle is especially significant because its logic turns out to be far more powerful than its modest formulation may suggest. Given some additional but fairly modest

---

[7]Amartya Sen, On Ethics & Economics 32 (1987). Sen goes on to say that "Pareto optimality can, like 'Caesar's spirit,' 'come hot from hell.'" Id.

[8]Id. at 35.

assumptions about continuity and about the form of utility functions, it turns out that the Pareto principle implies that social decisions can be made only by combining individual utility functions, eliminating any independent weight for concepts such as fairness. Thus, it is a very short step from Pareto to full-blown welfarism.

Although the details of the proof are clever and sophisticated, the basic idea is simple. The heart of the proof is follows. If some non-welfare factor really matters, then at least it must be able to break ties between two states of the world that have equal welfare. If we assume continuity, then the state of the world with the greater degree of fairness should still be favored when two states of the world are very close but not tied in terms of welfare. This means that we can end up favoring a state where welfare is lower, and it turns out that we can find such a state that is actually Pareto inferior.[9] We could avoid this result by dropping the assumption of continuity. But dropping this assumption essentially presumes that the non-welfare value is trumped by an infinitely small loss of welfare. So we can keep the non-welfare factor but only by giving it virtually no weight at all. Thus, modest and intuitively appealing as it may be, the Pareto principle potentially has far-reaching consequences.

The Pareto principle is generally considered to be the "gold standard" for policy analysis. When a change in legal rules produces a Pareto improvement, the argument is over – or so it is thought. While much dispute exists about Kaldor-Hicks efficiency[10] and about the relevance of distributional norms to law and economics,[11] the Pareto principle is often taken by practitioners

---

[9]See Louis Kaplow and Steven Shavell, *Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle*, 109 J. Pol. Econ. 281 (2001).

[10]This concept is defined in Michael J. Trebilcock, The Limits of Freedom of Contract 7 (1993).

[11]See id. at 833-834 (on redistribution).

of law and economics as being beyond controversy.[12] This is not surprising, for the Pareto principle has a strong intuitive appeal. To quote Cooter again:

> [Economic models of law] ask whether legal resources can be reallocated so as to make at least one person better off without making anyone worse off. *Thus the economic analysis of law attempts to maximize the value of legal resources to the people who enjoy them.* For every wasteful law, a more efficient one could be substituted and the savings could be distributed among the people affected by the change so that some are made better off without making anyone worse off. Waste is, consequently, an irrationality to expunge from law and policy.[13]

Who can possibly be in favor of wasting resources? Isn't the Pareto principle one of those rare moral precepts – a self-evident truth? Alas, its self-evidence is most apparent as a distance, but begins to disintegrate when we examine it closely.

II. Preferences versus Welfare

Recall the homely example used in the previous section. If Mary wants to see Movie #1 rather than Movie #2, and Bob either agrees with her or doesn't care, then the couple surely should go see Movie #1, which makes her happier and at worst makes no difference to him. And yet, on reflection, it is not so clear that the couple "should" see the first movie. Maybe Mary is wrong: the second movie will give her more pleasure or expand her horizons. Or maybe she only wants to go to the first movie because it will allow her to feel superior to Bob. That might not be

[12]See Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green, Microeconomic Theory 313 (1995)(Pareto efficiency as minimal criteria for defining social optimality); Mark Seidenfeld, Microeconomic Predicates of Law and Economics 54 (1996)((economists are "enamored" of Pareto principle).

[13]Trebilcock, supra note 10 at 822 (emphasis in original).

a good thing. Or maybe they should talk over the decision, as a result of which they might discover that it would be better for them *as a couple* (rather than as two individuals) to go to the other movie. In our movie example, these are fairly trivial concerns – after all, it was a trivial hypothetical – but in other settings, similar concerns may raise important moral questions.

Part of the appeal of the Pareto principle is found in a common paraphrase: society should take an action "if one person is better off, and no one else is hurt." But the principle itself is actually stated in terms of preferences and not in terms of welfare. There is a reason for this: preferences are presumably objective social facts, but determining a person's true welfare involves a normative judgment that most economists would prefer not to make. Yet the jump from preferences to welfare is problematic. Without such a jump, Pareto only stands for the notion that society should give people what they individually desire or at least what they think they desire. But as the example of Bob and Mary indicates, it can be a mistake to confuse what people individually desire with what makes them collectively better off.

A. Preferences and Individual Welfare

Imagine a one-person society: Robinson Crusoe on his island. For him, the Pareto principle seems tautological: what Crusoe prefers, "society" (which is just another name for Crusoe himself) also prefers. But does this give us any reason to believe that what Crusoe prefers will actually increase his welfare?   Is satisfying his preferences *necessarily* something we should consider a good thing for Crusoe.   Suppose a purely benevolent figure – call her Crusoe's Mom – is hiding on the island. Is it clear that Crusoe's Mom should try to arrange matters so his preferences will be satisfied as much as possible under the circumstances? It's not at all clear that she should do so, as we will see. She might try to satisfy his preferences in order to make him feel happier or to make his life better (which might of course not be the same thing). But the linkage between preferences and either happiness or welfare is contingent.

It might seem obvious that satisfying preferences makes people happier, but this is not always true. In reality, satisfying people's preference might or might not make them happier. For example, studies by psychologists show that "the very wealthy do not have a substantially more favorable perception of the quality of their lives than do the middle classes."[14] In contrast, education "produces a greater sense of well-being than wealth equal to the education's price."[15] In short, while most people have a preference for increased wealth, satisfying this preference may not in fact be a good way of making them happier. It is even less clear that great wealth would necessarily advance their "well-being" in any meaningful sense.[16]

"While being happy is a momentous achievement," according to Sen, "it is not the only achievement that matters to one's well-being."[17] Or, as Judge Posner puts it, "Happiness is important to most people, but it isn't everything. How many of us would be willing to take a pill that would put us into a blissfully happy dreamlike trance for the rest of our lives, even if we were absolutely convinced of the safety and efficacy of the pill and the trance?"[18]

---

[14]Herbert Hovenkamp, *The Limits of Preference-Based Legal Policy*, 89 Nw. U.L. Rev. 4, 37 (1994).

[15]Id. at 37-38.

[16]See Kornhauser, supra note 6.

[17]Sen, Ethics & Economics, supra note 7, at 60.

[18]Richard A. Posner *Norms and Values in the Economic Approach to Law* [in this volume].

The argument against equating the two, or otherwise defining well-being as a state of mind, is nicely summarized by Hausman and McPherson:

> Suppose there were an "experience machine" that could give people the highest quality experiences possible. These high-quality experiences might be intense sensations of pleasure or they might be experiences of climbing Everest or composing a symphony. Let them be whatever experiences mental-state theorists of well-being claim are ultimately and intrinsically good. The mental-state theorist would then have to say that all people would be better off permanently hooked up to a reliable experience machine rather than living their own lives and experiencing the decidedly mixed mental states that come with them. If one believes that those who are hooked up to the experience machine are missing out on some of the intrinsically good things in life . . . then one cannot accept a mental-state view of well-being.[19]

The distinction between happiness and well-being is a classic issue regarding utilitarianism. Completely decisive philosophical arguments do not exist but the argument against defining utility as a mental state seems to come close. Yet, if we reject happiness as the ultimate social goal, what do we use instead? Sen and others have argued for a richer conception of human flourishing. In Sen's version, for example, society should aim at extending human capabilities.[20] Martha Nussbaum also argues on behalf of the capabilities approach, using Aristotelian philosophy as her starting point. She concludes that aggressive social action is

---

[19]See Daniel M. Hausman and Michael S. McPherson, Economic Analysis and Moral Philosophy 74-75 (1996).

[20]Amartya Sen, *Capability and Well-Being*, in Martha C. Nussbqaum and Amartya Sen, The Quality of Life 30 (1993).

needed to improve the capabilities of women in many countries around the world.[21] Whatever may be the merits of these or other efforts to define individual welfare objectively, we certainly cannot assume that what people want on any given occasion will actually promote their objective welfare. Thus, if we want to maintain a defense of the Pareto principle in terms of people being "better off," we must define the relevant preferences counterfactually, as the preferences people *would have had* if they understood their true welfare.

B. Preferences and Social Welfare

Apart from questioning the linkage between preferences and individual welfare, we might also question whether individual preferences invariably translate into some measure of social well-being. There are three problems here: (a) satisfying some kinds of antisocial preferences might not contribute to a better society, (b) we may not be able to aggregate individual preferences in a straightforward way, and (c) we may think that society's decision making process should give people the opportunity to deliberate collectively over their preferences rather than simply taken those as given. We discuss these problems in turn.

As to the first problem, it has often been suggested that some preferences, such as those based on racism or sadism, do not deserve to be counted in making societal decisions.[22] For example, Hovenkamp argues that some preferences must be considered irrelevant because they violate the premises of a constitutional democracy:

---

[21]See Martha C. Nussbaum, Women and Human Development: The Capabilities Approach (2000). For a critique of her views, see John Lewis, *Giving Way: Martha Nussbaum and the Morality of Privation*, 8 U. Chi. L. Sch. Roundtable 215 (2001).

[22]See Hausman & McPherson, supra note 19, at 78-79.

> A constitutional democracy with a meaningful bill of rights places limits on the range of preferences that the state will recognize. For example, even if every citizen of Smallville is an evangelical Protestant who favors Bible reading the public school, such Bible reading is forbidden by the federal constitution. . . .
>
> By its nature, a bill of rights removes some kinds of preferences from society's consideration, while it raises others to a privileged level.[23]

The idea of laundering preferences is also endorsed by Adler, one of the leading law school advocates of welfarism.[24]

In contrast to Hovenkamp, Kaplow and Shavell argue that all preferences are equally entitled to respect. We should not, therefore, discard preferences such as racism or sadism in calculating social welfare:

> To trump preferences is, in essence, to redefine individual's well-being in a manner that substitutes some other preferences – ones that are cleansed, so to speak – for individuals' actual preferences . . . But such an approach is troubling from the perspective of welfare economics because the moral force and appeal of welfare economics lies in promoting the actual well-being of people, not in advancing some hypothetical notion of satisfaction that is distinct from that of the individuals who are the objects of our concern. Furthermore, employing a cleansed version of preferences, rather than actual preferences

---

[23]Hovemkamp, supra note 14, at 63-64.

[24]See Matthew D. Adler, *Beyond Welfarism and procedure: A Welfarist Theory of Regulation*, 28 FSU L. Rev. 241, 262-267 (2000).

> may lead one to favor policies that make everyone worse off, just as when a notion of fairness is pursued at the expense of individual's well-being.[25]

Thus, they say, "[t]he idea of an analyst substituting his or her own conception of what individuals should value for the actual views of the individuals themselves conflicts with individual's basic autonomy and freedom."[26]

The trouble with the Kaplow and Shavell argument on this point, as we will see later, is that the Pareto principle is only weakly linked with autonomy. One might instead argue on the basis of liberal tolerance that society should count all preferences equally, even intolerant ones. But this requires a very strong form of tolerance. It is one thing to argue against social intervention to change or frustrate anti-social tastes. It is another to argue that society has a collective interest in actually advancing those tastes (at least when doing so does not harm other individuals) -- for example, that if people like telling racist jokes and society considers such jokes harmless, a liberal society should actively try to promote opportunities for racists to tell these jokes to each other. (And if we assume continuity, that means that society should continue to promote those jokes even when they do some degree of harm to others.) Such an argument would not be impossible to make but it is clear that we are far from the sort of self-evidence that the Pareto principle originally seemed to promise.

Another problem with using the Pareto principle for making societal decisions also deserves mention. There is a technical difficulty involved in applying the Pareto principle under conditions of uncertainty. It is quite possible that people are unanimous in their preferences over

---

[25]Louis Kaplow and Steven Shavell, Fairness versus Welfare 419-420 (2002).

[26]Id. at 421-422.

choices, but have varying views about the outcomes and probabilities involved. Simply aggregating the preferences may be misleading.[27] For example, consider the question of whether to adopt a policy requiring all welfare recipients to work. Liberals might support the policy because they believe that it will prevent cuts in benefits levels, which they oppose. Conservatives might support the policy because they believe it will lead to benefit cuts, which they favor. So there is unanimous support for the proposal. But this does not mean that adoption of the proposal will lead to a state of the world preferred by everyone over the status quo. In fact, it is logically impossible for both liberals and conservatives to be satisfied, since either benefits levels will go down or they will not.

Presumably, the solution to this problem is to apply the Pareto principle only to preferences about outcomes, as opposed to preferences about decisions. The decision-maker then applies her own best estimate of probabilities to determine whether a particular decision would be Pareto preferred if the participants shared his best estimate of the outcomes. Alternatively, the decision-maker could pool the probability estimates of the individuals involved. But separating the probabilities and the preferences creates difficulties of its own. Regardless of whether the decision-maker uses pooling or an independent assessment of probabilities, the probability figure she derives will not generally correspond to those of the affected individuals. It is even possible that she would find a policy to be Pareto optimal although it is universally opposed, because she determines that using the correct probabilities, everyone would prefer the expected outcome of the proposed action over the status quo. To the

[27]See Hausman & McPherson, supra note 19, at 88.

extent that attachment to the Pareto principle rests on a belief in individual autonomy, the decision-maker is on questionable ground here.

This objection to rejecting individuals' probability estimates disappears if our basis for adopting the Pareto principle is individual welfare rather than autonomy. But the welfare analysis becomes much more difficult than the Pareto principle seems to promise, for we can no longer simply base the societal decision on actual individual rankings. Applying his own estimate of probability (or a pooled probability) to the outcome preferences of the individuals requires that the decision-maker have very good information about those preferences. Rather than simply knowing that individuals choose action A or action B, the decision maker must be able to attach cardinal utilities to the two outcomes in order to calculate expected utility. As a practical matter, this may make application of the Pareto principle quite difficult.

A final difficulty in moving from individual preferences to social decisions is that we might not want to take those preferences as pre-political "givens." Communitarians believe that "communities, through their collective choices, may decide to adopt some new conception of the common good, and preferences are likely to adapt accordingly as individuals participate in this process of political reconceptualization of their community values."[28] Going even further, some comunitarians argue that private preferences are essentially irrelevant – what matters are not the preferences a person would express in market transactions but rather the values she would

---

[28]Trebilcock, supra note 10 , at 155-156.

express in her role as a citizen.[29] The distinction is said to be that choices involve moral commitments and beliefs rather than merely wants and desires.[30]

III. Contingent Preferences

Preferences may depend on circumstances. "That some nontrivial part of people's preferences is a product of their environment, rather than logically prior to all experience, seems so obvious that it is not worth debating."[31] Thus, some people "may want things precisely because they cannot have them ('the grass is always greener on the other side of the fence'), while other people spurn what is beyond their reach, like the fox who judged the unobtainable grapes to be sour."[32] No less an advocate of economic analysis than Judge Posner tells us that "people's preference often do not coincide with what a neutral observer would think good for them or what they would think good for themselves if they lived in a different kind of society or just if they knew more about alternative ways of living."[33]

A. Before and After the Transaction

[29]See Jane B. Baron and Jeffrey L. Dunoff, *Against Market Rationality: Moral Critiques of Economic Analysis in Legal Theory*, 17 Cardozo L. Rev. 431, 440 (1996).

[30]Id. at 443.

[31]Hovenkamp, supra note 14, at 51.

[32]Hausman & McPherson, supra note 19, at 79.

[33]Posner, supra note , at 12. Posner views preference stability as a methodological commitment, adopted to make prediction more feasible, but this does not eliminate the normative problems raised by the Pareto principle when in fact preferences are contingent.

Before the fact, the parties to a contract make think it will make them better off, but one or both of them may be wrong. As Trebilcock points out, from an *ex ante* perspective, this later regret is irrelevant, assuming the earlier decision was fully informed, rational, and voluntary. But from an *ex post* perspective, performance of the contract is no longer Pareto optimal. Trebilock calls this the "Paretian dilemma."[34] Economists strongly favor the *ex ante* perspective, which in effect resolves the dilemma in favor of the earlier preferences of the parties.[35] Others would favor the use of *ex post* preferences. Indeed, they say, "the very point" of laws may be to "cultivate and elevate preferences."[36]

Although the *ex ante* perspective is not untenable, the normative case for it is far from straightforward.[37] Even where preferences have not changed, the *ex ante* perspective is not so obviously the correct moral vantage point.

For example, suppose that two lawyers are trying to settle a case. They both prefer the settlement to the risk of waiting for the jury to come back with a verdict. Thus, ex ante, enforcing the settlement seems Pareto optimal, since it is preferred by both to the alternative of non-settlement. Unbeknownst to either of them, the jury has already come back to the courtroom with a verdict. When she learns this, the lawyer for the winning party prefers to nullify the settlement. So *ex post*, the settlement is not Pareto optimal. It is not clear that her original

---

[34]Trebilcock, supa note 14, at 244.

[35]See Cooter, supra note 5, at 824; Frank Easterbrook, *Forward: The Court and the Economic System*, 98 Harv. L. Rev. 19-33 (1984). Gordon Tullock, *Two Kinds of Legal Efficiency*, 8 Hofstra L. Rev. 659, 863 (1980).

[36]Baron and Dunof, supra note 29,at 445.

[37]For more extensive discussion of my views on this point, see Daniel A. Farber, *Economic Efficiency and the Ex Ante Perspective*, in Jody S. Kraus and Steven D. Walt, The Jurisprudential foundations of Corporate and Commercial Law 54 (2000).

consent to the settlement should be morally binding, or that the *ex ante* perspective on this transaction is more valid than the *ex post* perspective. It is true that she consented to the transaction, but only under a mistake of fact. (Of course, we might say that morally she assumed the risk, but this seems rather circular – it assumes that she should have known she would be morally bound even if the jury had already returned – and in addition there is no reason to think that either party was consciously aware of this particular risk.) We may enforce such a contract, but that is on the grounds of social utility rather than any *a priori* argument for applying the Pareto principle *ex ante*.

To sharpen the issue, consider the following hypothetical. Imagine a society of sadists. The Torture Show, a television program, kidnaps randomly selected individuals and tortures them to death on camera. (In some ways, this is only an exaggeration of the basic idea of some actual "reality TV" shows, where individuals are humiliated for the entertainment of others!) Assume that the odds of being a victim are so small that *ex ante*, everyone would prefer to have the show than not to have it. *Ex post*, however, the victims' preferences regarding continuation of the program are reversed – after all, they are sadists rather than being insanely masochistic. (But note that the victim's underlying goals have not changed -- he is still a sadist and still happy to watch *other* people being tortured.) Does the Pareto principle require that the show continue? One argument against the show is that sadistic preferences do not count; thus, even *ex ante* we give no weight to the pleasure anyone might derive from watching and only count the expected cost of being a victim. We will consider that argument later, but for the moment let us put it aside. Is it clear that the *ex ante* perspective is appropriate here?

The essential claim behind the *ex ante* perspective is that the victims would have accepted the gamble in advance so that it is fair to hold them to it now. But there are two possible arguments to the contrary, neither of which can be dismissed out of hand. The first counter-argument is that, although the victims would have consented to taking the risk of becoming victims, they did not actually do so. Even in the unlikely event that we have perfect information about their preferences, their hypothetical consent may not carry the same moral force as actual consent. To say that a person can be subjected to pain and death because he actually agreed to it is one thing; to say that he could be so treated because he *would* have agreed seems much weaker. The second counter-argument is that even actual consent might not be enough. Perhaps it is not morally permissible to torture people to death for fun even if they have consented. This second argument takes the idea of inalienable rights seriously -- people have a non-alienable right to be free from this kind of conduct, which they retain even if they attempt to bargain it away.

If we endorse the Pareto principle, adopt the *ex ante* perspective, and refuse to launder away tastes like sadism, then we are committed to endorsing The Torture Show. Part of the power of this example is that the preferences involved are so distasteful, suggesting preference laundering as a solution to the problem rather than abandoning the *ex ante* perspective. But there are also scenarios with less distasteful preferences where the same problem arises.

For instance, assume that the death penalty is a highly effective way to deter crime (a controversial assumption). It has always been in effect in the jurisdiction. For any citizen who is not actually being executed at any given time, the crime control advantages outweigh the risk of being executed in the future. John Doe, an innocent person, is now on death row. Must we say that Doe has no reason to complain of being executed, because in the past he would have endorsed the death penalty if anyone had thought to ask him? It seems difficult to give this much weight to a counterfactual act of consent. Even if he had consented, consent to a risk might not automatically translate into consent to a harm. If capital punishment is adopted for reasons of social policy, some risk of executing an innocent man is inevitable, but this seemingly makes such an execution morally excusable, not morally desirable. (One might be tempted to say that Doe has benefitted from the existence of the death penalty in the past, but this might not be factually true – quite likely he personally would not have been a murder victim even without the death penalty.) Are we committed to calling the death penalty Pareto optimal by disregarding an innocent person's current objection to being executed? And are we sure that even actual consent to be wrongfully executed would be enough to justify the execution?

If we are not prepared to endorse the *ex ante* perspective universally, the Pareto principle becomes ambiguous. Do we mean pre-transaction preferences or post-transaction ones? We might rescue the principle, perhaps, by applying it only when no one would object to a transaction *either* ex ante *or* ex post. But the number of legal problems involving such situations must be vanishingly small. If everyone likes an idea in advance and like it after the fact, what would they be doing in court?

In the examples we have considered, individuals' basic values and tastes have remained unchanged; it is only the circumstances that have changed. More profound problems are introduced when the transaction actually causes changes in values and tastes, making it difficult to determine which are the person's "true" values. When a person's goals have changed substantially, it is not at all clear that his or her earlier goals should be privileged over later ones. As Kronman says:

> From the standpoint of his present values, which he cannot shake off or suspend, his past actions may seem pointless or evil; in this respect, he is likely to regard his earlier decisions as a foreign element whose continuing influence appears senseless from the standpoint of his present goals.[38]

For example, addictions and habits reflect preferences which are themselves generated by past behavior. In this situation, Kaplow and Shavell argue that we should accept the *ex post* preferences of the individual as more accurate gauges of welfare. Thus, "if it is discovered that individuals are able to adapt to certain physical disabilities more or less readily than is commonly supposed, the valuations employed in measuring tort damages or in performing cost-benefit analysis (for example, of highway safety improvements) should reflect actual harm rather than the victim's uninformed ex ante estimates."[39] This example seems dubious. If a person was afraid of being maimed but later found that the pain killers made him happier than ever, we would still be reluctant to call his being maimed a Pareto improvement.

Another example, which may be more favorable to the Kaplow and Shavell view, may be presented by the issue of surrogacy contracts. When a woman enters into a contract to bear a child for another couple, it may seem like a very favorable arrangement. But after giving birth, she might unexpectedly find herself with a much different attitude toward giving up the child. She might even feel that, having gone through the experience of childbirth, she is now a different person than she was when she entered into the contract. A rule enforcing the surrogacy contract is Pareto optimal based on *ex ante* preferences but not on the basis of *ex post* preferences.

---

[38] See Anthony T. Kronman, *Paternalism and the Law of Contracts*, 92 Yale L.J. 763, 782 (1987).

[39] Kaplow and Shavell, supra note 25, at 20.

In such situations, the Pareto principle becomes indeterminate unless we simply resolve the situation by fiat, as Kaplow and Shavell themselves do. On paternalistic grounds, perhaps their solution might have something to say for it, but it is difficult to support on autonomy or consent-based grounds – essentially it involves exposing people to risks that they would not voluntarily undertake because we think we know better than they do how they will feel afterwards. This is a strategy parents frequently adopt with small children (with mixed success). It is less clear that such a strategy should be adopted by society toward its members. In any event, this interpretation of the Pareto principle moves us far away from autonomy values toward paternalistic supervision of individual welfare.

B. Before and After the Legal Rule

A particular problem arises when the legal rules themselves affect preferences.[40] Then the policymaker must decide beforehand which set of preferences to favor, a decision that obviously cannot be made by looking at those preferences themselves.[41]

For instance, cognitive psychologists have established a recurring pattern in human behavior: people will demand more to sell something than they would be willing to pay to buy it. For example, in one experiment, students who were randomly given a candy bar instead of a coffee mug were unwilling to trade for the mug, but those randomly receiving the mug were equally unwilling to trade for the candy bar. This "endowment" effect is strongest when people

---

[40]For discussion of this situation, see Cass R. Sunstein, *Legal Interference with Private Preferences*, 53 U. Chi. L. Rev. 1129 (1986).

[41]See Hausman & McPherson, supra note 19, at 76-77. For example, changes in contract remedies may induce changes in preferences for "fairness", according to Oren Bar-Gill and Chaim Fershtman, *Law and Preferences*, 20 J. L. Econ. & Org. 331 (2004).

feel some moral entitlement to what they possess, even if the grounds for such a feeling are slim.[42]

In this situation, the Pareto standard is completely indeterminate. If people happen to have coffee mugs initially, they universally prefer the mugs to the candy bars. Hence, it appears that possessing mugs is Pareto superior to possessing candy bars. If they happen to start off with candy bars, however, we find the opposite results. In particularly, if we are in the position of assigning legal entitlements to mugs versus candy bars, we cannot apply the Pareto principle because the preferences depend on how we assign the entitlements.

Again, this is less of a difficulty if we take a paternalistic view of Pareto than an autonomy-based one. From the paternalistic view, we need merely decide which situation makes people happier or better off – how they rank alternatives themselves or which they would choose has no relevance except as evidence of what is really good for them. Of course, the paternalistic Paretian is left with the considerable problem of determining on people's behalf what is really good for them.

Another instance of shifting preferences arises when adoption of a legal rule causes a conceptual shift in how people regard the relevant transaction. One concern is that legalizing certain market transactions will have harmful effects on how people view themselves and each other. The Pareto principle is strongly linked with the idea of market transactions. Assuming fully informed, rational, and uncoerced decisions, an agreement to exchange two commodities must be a Pareto improvement – no one would agree to such an exchange unless it satisfied their

[42]For an extensive discussion of this experiment and other research findings, see Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler, *Experimental Tests of the Endowment Effect and the Coase Theorem*, in Cass R. Sunstein (ed.), Behavioral Law and Economics 211 (2000).

preferences at least as well as the status quo. But society has always had doubts about certain exchanges, such as the sale of sexual services, infants for adoption, and body organs. One objection is that we should not transform sexual activities, children, or body parts into commodities. Instead, we should view these things as integral to our definition of our selves, and thus incapable of being property.[43] A specific fear is that "permitting transactions such as these, along with the market rhetoric and manifestations that accompany them, may change and pervert the terms of discourse in which members of the community engage with one another."[44] For example, "[o]ne concern with respect to legal sanctioning of *inter vivos* organ sales is that people will begin to view their organs entirely as commodities."[45] Similarly, critics of surrogacy contracts fear that such contracts will reinforce oppressive gender roles.[46]

Rather than being worried that a new legal rule will change preferences, we might welcome this result if we believe that current preferences are distorted. One long-standing concern about basing social policy on existing preferences is that those preferences may be warped by existing oppressive institutions. As Sen says:

> A person who has had a life of misfortune, with very little opportunities, and rather little hope, may be more easily reconciled to deprivations than others reared in more fortunate and affluent circumstances. . . . The hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie may all

---

[43]See, e.g., Margaret J. Radin, *Market Inalienability*, 100 Harv. L. Rev. 1849 (1987).

[44]Trebilcock, supra note 10 , at 26.

[45]Id. at 35.

[46]Id. at 50.

take pleasures in small mercies, and manage to suppress intense suffering for the necessity of continuing survival, but it would be ethically deeply mistaken to attach a correspondingly small value to the loss of their well-being because of this survival strategy.[47]

Worse, individuals may adjust to oppression and even come to prefer their oppressed circumstances over the potential for freedom. For instance, women may adjust to sexist expectations and develop preferences for limited career options and dependence on men.[48] But for these women, "liberties, high wages, and protection from domestic violence may make them better off than giving them what they prefer."[49] If the legal rules are changed, perhaps they will learn to value their new freedoms and benefits.

How much to worry about preference shifts is a serious question. We have relatively little empirical evidence about how legal rules shape popular preferences. Sometimes, significant effects seem apparent. For example, non-smokers frequently say they have become much more aware of the unpleasantness of tobacco smoke since anti-smoking ordinances created so many smoke-free public spaces. Even those who might have been indifferent originally to the passage of the ordinances may strongly oppose repeal. In other arenas, effects may be more mixed. Sexual harassment laws have no doubt made some men change their own views of appropriate conduct in the workplace. But others no doubt view these laws only as an annoying impediment to their activities. As a general matter, we know relatively little about how much

[47]Sen, Ethics and Economics, supra note 9, at 46.

[48]See Hausman and McPherson, supra note 19, at 79.

[49]Id.

legal rules are internalized over time. The more widespread this phenomenon is, the less useful preferences at any given time are as guides to legal policy, and thus the smaller the scope for the Pareto principle.

IV. Limited Rationality

Even if actual preferences are left unaffected by transactions or rule changes, they may not provide a useful standard for societal decisions. It is no news, except perhaps to the most sheltered economic theorists, that "[p]eople's preferences are sometimes irrational in the sense that they are not mathematically consistent with other preferences or perhaps in the weaker sense that they simply do not seem intelligent in the light of the information that people have at the time they make a choice."[50] Cognitive psychologists have documented an array of errors that human beings are prone to.[51] These errors make it harder to connect the proposition "individuals prefer *A* to *B*" to the proposition "society should choose *A* over *B*."

A. Cognitive Limitations

Sometimes people's preferences are disarranged because of their inability to process information. We cannot necessarily say as a result that any given preference is wrong. Rather, we must say that preferences are out of alignment with each other – for example, that preferences for specific outcomes when their achievement is a certainty are out of whack with preferences for various probabilistic packages of outcomes. Given such inconsistencies, we cannot take

---

[50]Hovenkamp, supra note 14, at 46.

[51]A good survey can be found in Russell B. Korobkin, and Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 Cal. L. Rev. 1051 (2000). Some of the implications of these results for autonomy norms are explored in Cass R. Sunstein and Richard H. Thaler, *Libertarian Paternalism is Not an Oxymoron*, 70 U. Chi. L. Rev. 1159 (2003), and Colin Camirer, et al., *Regulation for Conservatives: Behavioral Economics and the Case for "Assymetric Paternalism,"* 151 U. Penn. L. Rev. 1211 (2003).

preferences as given but must make adjustments. This necessarily involves a certain degree of paternalism, further undermining the autonomy argument for the Pareto principle.

Thanks to the efforts of cognitive psychologists, we now know a good deal about human intellectual limitations. For example, we know that humans make systematic errors in estimating risks. They tend to be overly optimistic, thinking they are less subject to risks than the average person. On the other hand, they overestimate risks based on the amount attention given to the risks by the media. When an event has happened, they also overestimate the prior probability of its occurrence. Thus, people's decisions under uncertainty may reflect mis-perceptions of probabilities rather than true preferences about outcomes.[52]

Preferences may also be disarranged because preferences over specific attributes do not match preferences over complicated packages of attributes. When faced with complex tasks, people take shortcuts. For example, rather than trying to find the optimal outcome, they "satisfice" by settling for the first acceptable option rather than continue to search.[53] Even when given correct probability information, people mis-process the information, ignoring information about base rates and focusing instead on the representativeness of an outcome.[54]

These cognitive deficits clearly make practical application of the Pareto principle more difficult. We can no longer take completely for granted that voluntary, fully informed exchanges meet the Pareto principle. What is less clear is whether irrationality causes any normative difficulties for the Paretian. The effect of these cognitive difficulties is to take the analyst farther

---

[52]See Sunstein, *Introduction*, in Sunstein, Behavioral Law and Economicss, supra note 42, at 4-5.

[53]See Korobkin and Ulen, supra note 51, at 1078.

[54]Id. at 1086-1090.

way from what people actually prefer regarding certain choices, and toward what people would prefer if their thinking met the economist's conception of rationality. Whether this is normatively troublesome depends on our degree of toleration for paternalism. In any event, we are once again pushed away from autonomy and preferences toward paternalism and objective welfare as the basis for applying and justifying the Pareto principle.

B. Incomplete or Incoherent Preferences

Reliance on preferences can also be problematic when people have incomplete or incoherent preferences. This leaves the Pareto principle impotent to the extent we are really serious in phrasing it in terms of "preferences." To the extent that "preference" is really just shorthand for "what people would really prefer if they knew what was truly good for them," we can continue to endorse the Pareto principle, though only as a minimalist form of utilitarianism.

We can begin by considering situations in which people have seemingly inconsistent preferences. For example, they frequently use a different discount rates for different time periods, resulting in preference reversals due to the mere passage of time.[55] This leads to situation where it is rational to make a plan, which it will then be rational to break despite the absence of any relevant change in preferences or circumstances. Similarly, people often demand more for accepting a delay than they would be willing to pay to accelerate the same event.[56] Such situations, as well as those involving inadequate "will power," can be described by ascribing

---

[55]See id. at 1120.

[56]See id. at 1122.

different "selves" to the same person at different moments of time – in effect, the self who decides on a diet is a different person than the self who later sneaks a candy bar.

Other inconsistencies seem to show that preferences in some sense do not exist until people have reason to construct them, and that the results of the construction depend on the context. For example, the choices people make in *pricing* risky alternatives are not always consistent with their *rankings* of the alternatives.[57] The general conclusion is that preferences are constructed not found:

> The discussion of the meaning of preference and the status of value may be illuminated by the well-known exchange among three baseball umpires. "I call them as I see them," said the first. "I call them as they are," claimed the second. The third disagreed, "They ain't nothing till I call them." Analogously, we can describe three different views regarding the nature of values. First, values exist – like body temperature – and people perceive report them as best they can, possibly with bias (I call them as I see them). Second, people know their values and preferences directly – as they know the multiplication table (I call them as they are.) Third, values or preferences are common constructed in the process of elicitation (They ain't nothing till I call them). The research [on preference reversal] . . . is most compatible with the third view of preference as a constructive, context-dependent process.[58]

A somewhat similar problem of preference construction may exist when the choice is between sharply disparate values, such as money versus human life, or career versus family.

---

[57]See Richard H. Thaler, The Winner's Curse: Paradoxes and Anomalies of Economic Life 79-91 (1992).

[58]Id. at 90-91.

Economists assume that preferences always exist between any two outcomes, so that a person can readily make a choice. But a number of philosophers insist on the incommensurability of values and our inability to give them all a common metric.[59]

If we support the Pareto principle because we truly care about fulfilling people's preferences, this gives us a real problem: people may not have any ascertainable preferences at a given time. If instead, we are using Pareto as a stand-in for utilitarianism, we face some very difficult measurement problems. Essentially, in the instances where people find themselves stymied in making decisions for their own lives because of the difficulty of comparing radically different values, we must be prepared to make the comparison for them if we are to apply something like the Pareto principle.

We must also be willing to decide the "correct" way to frame specific choices, because what people prefer is often a function of how choices are presented. Framing effects are one of the best established findings in cognitive psychological. Choices depend not only on what options are present but on *how* those options are presented. For example, people will see a penalty for using credit cards differently from a bonus for using cash, and will respond differently to the two.[60] Yet, the economic substance of the two is exactly identical. Frames can lead to inconsistent attitudes toward risk aversion, and can be affected by fairly small changes in presentation.[61]

---

[59]See Baron and Dunoff, supra note 29, at 486.

[60]See Edward J. McCaffery, Daniel J. Kahneman, and Matthew L. Spitzer, *Framing the Jury: Cognitive Perspective on Pain and Suffering Awards*, in Sunstein, Behavioral Law and Economics, supra note , at 262.

[61]Korobkin and Ulen, supra note 51, at 1104-1107.

A similar problem is presented by menu effects. If preferences are well-behaved, whether *A* is preferred to *B* should not depend on how either compares with *C*. For example, suppose that I prefer chocolate to vanilla ice cream, and both of those to strawberry. Then I should always choose chocolate, regardless of which of the other two is on the menu. Yet in many situations, individual choices turns out to be context dependent. For example, juries presented with certain facts might choose a verdict of murder over a verdict of voluntary manslaughter if those are the only two choices. But if they are also offered the choice of negligent homicide, they switch to voluntary manslaughter. This is exactly like always choosing chocolate over vanilla, unless strawberry is also on the menu, when you reject chocolate for your former second-choice of vanilla.[62]

How are we to specify the "correct" setting for a decision, the one that will reveal what people "really" prefer? It is difficult to see any alternative except picking the frame that is most likely to lead people to the right choice, the choice that really furthers their welfare or happiness. So the framing problem reinforces our growing suspicion that the Pareto principle is really just a minimalist version of utilitarianism.

V. Preferences, Rights, and Social Choice

Even if preferences are perfectly connected with individual welfare and perfectly stable, some questions remain about the Pareto principle. As we will see, accepting Pareto may force us to abandon other desirable aspects of a decision-making process.

A. Autonomy and Pareto

[62]See Mark Kelman, Yurval Rottenstreich, and Amos Tversky, *Context Dependence in Legal Decision Making*, in Sunstein, Behavior Law and Economics, supra note 42, at 61 (reporting results of an analogous jury experiment).

Amartya Sen proved that the following three conditions cannot be simultaneously satisfied by any social choice function:

(1) The social choice method must apply to any possible array of individual preferences.

(2) The Pareto principle applies.

(3) Each individual gets the final say about the choice between at least one set of alternatives (for example, whether that individual reads or does not read a specific book).

Thus, Sen observed, the Pareto principle seems logically inconsistent with a minimal conception of liberalism in which individuals have the final say over even a single decision affecting themselves alone.[63] Sen provided a seemingly contrived example of this paradox, in which two friends must decide whether one of them will read *Lady Chatterly's Lover*. Because of the nature of the example, it is not difficult to dismiss the paradox as a very clever but practically insignificant hypothetical.[64] Still, as Sen says, the logical implications are disturbing:

> [I]t turns out that a principle reflecting liberal values even in a very mild form cannot possibly be combined with the weak Pareto principle, given an unrestricted domain. If we do believe in these other conditions, then the society cannot permit even minimal liberalism. Society cannot then let more than one individual be free to read what they like, sleep they way they prefer, dress as they care to, etc., *irrespective* of the preferences of others in the community.[65]

---

[63]Amartya Sen, *The Impossibility of a Paretian Liberal*, in Choice, Welfare, and Measurement 285 (1982).

[64]Id. at 288.

[65]Id. at 290.

Sen concludes from such considerations that Pareto is "unacceptable as a universal rule."[66]

Quite apart from Sen's technical result, one must question the degree of the connection between the Pareto principle and the idea of personal autonomy. The Pareto principle does not merely mean that people are entitled to make choices. It also says that it is perfectly legitimate for society to make choices for them, if society believes that the affected individuals will like the results and no one else will be harmed. The ultimate grounding of the Pareto principle is not autonomy but paternalism. As a practical matter, the best paternalistic strategy may be to let people make their own decisions because they have better incentives or information about their own preferences. But this is a purely contingent result; the Pareto principle does not preclude the possibility that no individual would ever be allowed to make a decision about his own life become some super-computer was better than they were at figuring out how to maximize their preferences. It is not surprising, therefore, that Sen was actually able to prove a logical inconsistency between Pareto and libertarianism.

In a recent book, Kaplow and Shavell extend Sen's result to argue that Pareto is inconsistent with any non-welfarist moral values such as fairness. They contend that "individuals will be made worse off overall whenever consideration of fairness leads to the choice of a regime different from that which would be adopted under welfare economics because, by definition, the two approaches conflict when a regime with greater overall well-being is rejected on grounds of fairness."[67] In particular, when people are symmetrically situated *ex ante* -- for example, equally

---

[66]See Amartya Sen, *Liberty, Unanimity, and Rights*, in Sen, Choice, Welfare, and Measurement, supra note 63, at 313.

[67]Kaplow and Shavell, Fairness versus Welfare, supra note 25, at 52.

likely to cause accidents or be accident victims -- Kaplow and Shavell say that it is "*always* the case that everyone will be worse off when a nation of fairness leads to the choice of a different legal rule from that chosen under welfare economics."[68] Their explanation is simply that:

> Because everyone is identically situated, whenever welfare economics leads to the choice of one rule over another, it must be that everyone is better off under the preferred rule. Hence, whenever a notion of fairness leads one to choose a different rule from that favored under welfare economics, everyone is necessarily worse off as a result.[69]

These are controversial arguments, but to the extent that they are accepted, they can as easily be used to raise questions about the Pareto principle as about fairness or other moral norms. One intuition is that a situation is better if everyone prefers it. Another intuition is that a situation is worse if it violates human rights. If they are inconsistent, which should prevail? To quote the title of one of their articles, Kaplow and Shavell prefer to say that *Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle*.[70] But they could just as well have titled the article, *The Pareto Principle is Inconsistent with Any Moral Theory Based on Fairness or Human Rights*. The two titles in fact mean exactly the same thing.

B. Pareto and Social Choice

If we are designing an institution for making societal decisions (i.e., a government), should we insist that the institution's decisions be consistent with the Pareto principle? This may

[68]Id. at 52 (emphasis in original)..

[69]Id.

[70]Kaplow and Shavell, supra note 9.

seem desirable, but it could in practice conflict with other desirable characteristics of social decision making.

As a matter of common sense, there seem to be a number of obvious attributes that any system for making social decisions should have. One of them is the Pareto principle. Another is that choices should be transitive: if society prefers *A* to *B* and *B* to *C*, then it should also prefer *A* to *C*. Still another is that the decision should not be affected by the presence of irrelevant alternatives: eliminating the least favored alternative from the agenda should not affect the ultimate choice between the other, more favored alternatives. (Recall the effect of the strawberry ice cream on the choice between chocolate and vanilla.) Alas, Arrow's Theorem is only one of a host of proofs that no mechanisms of social choice can satisfy some such set of obvious basic requirements.[71] Thus, we must sacrifice some desirable aspect of decision-making or another. The only question is which one.

In some situations, it may well be the Pareto principle that should be sacrificed. For example, majority voting systems are potentially subject to chaotic cycling, which could make the legal regime dangerously unstable. In order to attain stability, it may be useful to introduce some friction into the system, so that laws can only be passed if they make fairly large changes that possess wide popular support. But this friction may block changes that would help some small group of people without harming anyone else. Such a sacrifice of the Pareto principle may be worthwhile in systemic terms.

[71]For a discussion of Arrow's Theorem, see Daniel A. Farber and Philip P. Frickey, Law and Public Choice: A Critical Introduction 38-40 (1991). A summary of the leading impossibility results can be found in Prasanta K. Pattqanaik, *Some Paradoxes of Preference Aggregation*, in Dennis C. Mueller, Perspectives on Public Choice: A Handbook 201 (1997).

Similarly, we might sacrifice the Pareto principle to reduce transaction costs. Eliminating waste sounds like a good idea, but like all good ideas, it may have diminishing returns. It may simply not be worth the trouble of searching out every possible Pareto improvement; we might do better to settle for an outcome that seems reasonably close to Pareto optimality. We might also bend the Pareto principle to ensure that outcomes have other desirable qualities. For example, in designing mechanisms for dividing goods among individuals, it may be impossible to find one where the results are guaranteed to be Pareto efficient and also leave each individual equally satisfied.[72] We – or the individuals involved – might care more about equity than efficiency.

The Pareto principle may also preclude recourse to other (non-preference based) values in ways that are counter-intuitive. The logic here resembles Kaplow and Shavell's argument for excluding fairness, but applies to a broader range of values. Suppose, for example, that we believe that animals have rights but that those rights are highly subordinate to human welfare. Thus, animal rights become significant in making decisions as tie-breakers, when human welfare is unaffected. For example, one might consider the impact of a decision on animal rights when the affected humans have no preferences about the outcome one way or the other. This seems like a very modest and reasonable position, even for someone who does not put much stock in the idea of animal rights.

And yet, this position is incompatible with the Pareto principle, at least if we assume that the social choice function is continuous. Consider two outcomes, *A* and *B*, and assume that all

[72]See Steven J. Brams and Alan D. Taylor, Fair Division: From Cake-Cutting to Dispute Resolution 236 (1996).

the affected humans are indifferent between them but that *A* is much better for animals. Then society chooses outcome *A*. Now suppose instead that one individual has a minute preference for outcome *B*. If the social choice function is continuous, then this tiny preference shift should not change the result. So society still picks outcome *A*. But this violates the Pareto principle, for now every affected person is either indifferent or prefers outcome *B*. Precisely the same argument applies to environmental values such as the preservation of endangered species,[73] or to the preservation of great works of art. Thus, the Pareto principles means that if everyone else is indifferent and one person has a faint distaste for Rembrandt, it is better for society to destroy all of the Rembrandts in the world. This is far from being an intuitively compelling conclusion. One might reasonably think that the world would be poorer if the last redwood or the last Rembrandt were destroyed, even if everyone alive were too insensitive to care.

Thus, in at least some circumstances, we might be willing to allow modest incursions on the Pareto principle, thereby tolerating a certain amount of economic inefficiency, because we would like the decision-making process to have other desirable attributes. An ideal system for making decisions would incorporate all of the features we desire, but unfortunately no such system is possible. The Pareto principle does not seem uniquely entitled to a sacrosanct status if such tradeoffs must be made.

VI. Conclusion

In the course of working on this essay, I found myself driven to three conclusions.

[73]See Daniel Farber, *From Plastic Trees to Arrow's Theorem*, 1986 U. Ill. L. Rev. 337, 346 n.22 (arguing that the Pareto principle should be softened when individual preferences are only slightly in favor of the outcome which harms the environment).

The first conclusion is that the Pareto principle has no deep connection with individual autonomy or freedom. Sen's paradox is not just a clever parlor trick, but a symptom of a profound gap between the Pareto principle and libertarianism. The libertarian believes that people should get what they choose; the Paretian believes that people should get what they *would* have chosen. Moreover, in a host of different circumstances, the Pareto principle either breaks down because of the absence or inconsistency of the relevant preferences, or else it must be rephrased in terms of individual welfare rather than individual preferences. Deep down, the Pareto principle rests on paternalism not autonomy.

Second, partly for this reason, I have also concluded that the Pareto principle is a close cousin of utilitarianism.[74] The Paretian limits himself to the easiest cases for applying utilitarianism, those where the only utility changes are positive. This eliminates the problem of measuring utility, although we may still be perplexed about whether to consider the satisfaction of certain tastes as genuinely improving individual well-being. But although it avoids some of the practical problems of full-blown utilitarianism, it raises many of the same ethical issues. Indeed, as we have seen, it takes only the addition of a continuity assumption to change the Paretian into a utilitarian (or a close cousin). As with utilitarianism, we can hope that the Pareto standard will ordinarily lead us to results favoring human liberty, equality, and other values. But there is no guarantee that this will always be so, and in case of conflict, we must be prepared to sacrifice human dignity and the like for greater social utility. Whether this seems problematic

[74]For this reason, Kaplow and Shavell's conclusions should be no surprise. If only welfarism can accommodate the Pareto principle, that is because the Pareto principle itself is welfarist to the core. See Brett McDonnell, The Economists' New Arguments 4 (forthcoming).

depends partly on how often we expect such conflicts to arise, and partly on whether we are willing to abandon our other moral intuitions.

Third, although I am no longer willing to accept the Pareto principle as a self-evident truth,

I continue to think that it is often a useful standard. But several conditions have to be met for us to be willing to apply the Pareto principle with any confidence:

1. To the extent that preferences are distorted by cognitive deficiencies or other factors, we have to be sufficiently sure of our ground to be comfortable with making paternalistic corrections.
2. Preference satisfaction cannot conflict with another goal (if any) we hold dearly, such as autonomy, individual rights, or cultural and environmental values.
3. Preferences must be relatively stable, unaffected either by the transaction or the legal environment.
4. Making Pareto optimality an absolute requirement must not require undue sacrifice of other desirable aspects of decision making such as stability or efficiency.

Most people will find that these conditions are satisfied at least some of the time, many will find them satisfied often, and a few will find them satisfied always. Thus, the Pareto principle is likely to remain a useful tool. But we should not imagine that Professor Pareto showed us the way to normative bedrock.