

PATERNALISM, UTILITY, AND FAIRNESS

Author(s): Richard J. ARNESON

Source: *Revue Internationale de Philosophie*, Vol. 43, No. 170 (3), L'Éthique égalitariste / Egalitarian Ethics (1989), pp. 409-437

Published by: Revue Internationale de Philosophie

Stable URL: <http://www.jstor.org/stable/23946758>

Accessed: 20-06-2018 23:36 UTC

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.org/stable/23946758?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>



Revue Internationale de Philosophie is collaborating with JSTOR to digitize, preserve and extend access to *Revue Internationale de Philosophie*

PATERNALISM, UTILITY, AND FAIRNESS

Richard J. ARNESON

Liberals of various philosophical persuasions, perhaps including John Locke and J. S. Mill, have espoused a familiar principle that I will call the *liberty principle*, which holds that people should be left free to do whatever they choose unless their conduct threatens harm (in specified ways) to nonconsenting others. One implication of the liberty principle is *antipaternalism*: restriction of a person's liberty to carry out a voluntarily chosen course of conduct should never be imposed for the purpose of benefitting either that person herself or others who voluntarily consent to be affected by that conduct⁽¹⁾. Another implication is *no enforced charity*: if a person's conduct threatens no harm to nonconsenting others, restricting his liberty to carry out that conduct in order to force him to act for the benefit of others is never justified⁽²⁾. In *On Liberty* Mill attempted to develop a utilitarian argument for the liberty principle but critics have

(1) This definition needs further refinement, but will serve my purposes. As usually understood, antipaternalism forbids restricting a person's liberty for her own good against her present will, unless her present will is bound by voluntary prior commitment. On the definition of paternalism, see Gerald DWORAKIN, "Paternalism" and "Paternalism: Some Second Thoughts", both in *Paternalism*, ed. Rolf Sartorius (Minneapolis, University of Minnesota Press, 1983), pp. 19-22 and 105-107; also my "Mill versus Paternalism", *Ethics* 90 (July, 1980): 470-489; see esp. pp. 471-472. For a good criticism of my suggestion, along with many other instructive observations, see Joel FEINBERG, *The Moral Limits of the Criminal Law*, vol. 3, *Harm to Self* (Oxford and New York, Oxford University Press, 1986), pp. 3-23.

(2) On the question of whether Mill intended the liberty principle to forbid enforced charity, see D. G. BROWN, "Mill on Liberty and Morality", *Philosophical Review* 81 (1972): 133-158; David LYONS, "Liberty and Harm to Others", *Canadian Journal of Philosophy*, supp. vol. 5 (1979), pp. 1-19; and Fred BERGER, *Happiness, Justice and Freedom: The Moral and Political Philosophy of John Stuart Mill* (Berkeley and Los Angeles, The University of California Press, 1984), pp. 253-258.

insisted that either the argument, if it is genuinely utilitarian, doesn't work, or if it does work, then it isn't genuinely utilitarian (3).

In this paper I accept for the sake of argument the truth of all the empirically doubtful premises that Mill assembles to make his utilitarian case for the liberty principle, and I assume further that the logic of the arguments that Mill constructs with these premises is impeccable. My worry is that Mill's arguments even if successful in their own terms may be objectionable on grounds of fairness. Mill's arguments for the liberty principle appealing to the values of individuality and autonomy have an ideological thrust insofar as they ignore altogether the predictable distributive effects on more and less able agents of the operation of his proposed rules of "no paternalism" and "no enforced charity". Once noted, the point is obvious with respect to the "no enforced charity" component of the liberty principle, so my argument concentrates on Mill's grounds for the rejection of paternalism.

Joel Feinberg has recently produced a nonutilitarian defense of the antipaternalism component of the liberty principle that is noteworthy for its sophisticated account of the concept of voluntary choice and for its spirited defense of a principle of personal sovereignty seen as underlying an antipaternalist policy (4). I show that Feinberg's account of voluntary choice provokes its own fairness objection against antipaternalism and that his appeal to personal sovereignty is just as problematic as Mill's utilitarianism.

I. MILL'S ARGUMENT

The unfeasibility of perfect screening

Mill writes, "But the strongest of all the arguments against the interference of the public with purely personal conduct, is that when it does

(3) See Henry David AIKEN, "Utilitarianism and Liberty : John Stuart Mill's Defense of Freedom", in his *Reason and Conduct : New Bearings in Moral Philosophy* (New York, Alfred A. Knopf, 1962), pp. 292-314. See also the references cited in footnote 17 of chapter 1 of John GRAY's *Mill on Liberty : A Defence* (London, Routledge and Kegan Paul, 1983), pp. 131-132. C. L. TEN tries to show "that Mill's case for liberty is not wholly reconcilable with any consistent version of utilitarianism" and defends the nonutilitarian libertarianism he imputes to Mill in *Mill on Liberty* (Oxford, Oxford University Press, 1980). See especially the summary of his argument on p. 9, from which the quotation above is taken. See also FEINBERG, *Harm to Self*, pp. 57-62.

(4) FEINBERG, *Harm to Self*. (Further page references to this work are in parentheses in the text).

interfere, the odds are that it interferes wrongly, and in the wrong place" (5). This comment signals the level of abstraction at which Mill's argument is pitched. We are considering a rule regarding the treatment of paternalistic proposals that is to guide a policymaking public authority that must make decisions with limited and imperfect information about the impact the policies it is considering would have on the welfare of individual citizens. Moreover, the cost of tailoring a policy of interference to the individual case would be prohibitive even if perfect information were available. Any feasible policy must make broad rough classifications. We seek what is in effect a constitutional rule to guide policymaking in the nonideal world.

Consider a stylized description of a paternalistic decision. The public authority is deliberating about whether to prohibit a type of conduct that all agree harms no one except the agents themselves and others who freely consent to bear its consequences. Mill claims that for all proposed paternalistic restrictions, the loss to those hurt by the restriction outweighs the benefit to those helped by it. Regarding this claim, critics chide Mill for lapsing into uncharacteristically dogmatic judgment on empirical issues about which a utilitarian should be openminded. Let us give Mill the benefit of the doubt by assuming at least provisionally that his claim turns out true. What then?

In practice, paternalistic restrictions will constrain the liberty of some who would be better off in the absence of the restriction and of some who would be better off if the restriction is imposed. To some extent sheer luck determines who gains and who loses from paternalism, but we can abstract from this factor by concentrating on the expected gain or loss in welfare that an individual faces from a given proposed paternalistic restriction. Such restrictions override the judgment of the individual concerning what is best for herself. The better one's own judgment in these matters, the greater the chance that the paternalistic restriction will have bad consequences in its application to one's own case. The "traditions and customs of other people", Mill writes, should not be a law to the individual, because "their experience may be too narrow; or they may not have interpreted it rightly", or "their interpretation of experience may be correct, but unsuitable to him", or again, to conform to custom "merely

(5) John Stuart MILL, *On Liberty*, in *Collected Works*, vol. 5, ed. J. M. Robson (Toronto and Buffalo, University of Toronto Press, 1977), p. 283. (Further page references to this work are in parentheses in the text.)

"as custom" does not develop the individual's powers of practical judgment (p. 262). Individuals obviously differ markedly from one another in their abilities reasonably to decide whether they will be better off following or flouting commonsense maxims of prudence. The deliberative abilities required for practical judgment involve choice of goals as well as choice of means. Individuals also differ widely in their capacities to make good any of their decision-making defects, when plans go awry, by their perseverance and grit, shrewd and creative improvisation, and so on. Overall we know that if we were to rank individuals according to their self-regarding virtues, their composite scores would range all the way from "wise" to "foolish" and from "competent" to "incompetent".

Mill surmises that adhering to a policy of no paternalism will stimulate the development of people's skills at rational deliberation regarding their ends and of the character traits that will best enable them to achieve their ends. He may be right about this. Still, people differ widely in their native capacities for deliberation about plans and skillful execution of them. So the combination of the good effect of stimulating one's capacities and the bad effect of letting one's present incapacities express themselves in action will affect people's prospects of welfare in different ways, depending on their achieved and potential levels of self-regarding virtue and other particulars of their situation.

The Fairness Objection

To reiterate, Mill's claim is that every feasible paternalistic social rule would lower welfare on balance. This utilitarian case for strict antipaternalism could be correct yet have disturbing implications for the distribution of welfare. A ban on paternalism in effect gives to the haves and takes from the have-nots. Left unrestrained in self-regarding matters, more able agents are more likely to do better for themselves choosing among an unrestricted range of options, whereas less able agents are more likely to opt for a bad option that paternalism would have removed from the choice set. In Mill's utilitarian calculation, it turns out that under a no-paternalism rule the losses, if any, suffered by the less able (whom we suppose are on average worse off to begin with) are outweighed by gains enjoyed by the more able (whom we suppose are on average better off to begin with) (6).

(6) Two clarifications are needed here. First, my argument does not require the assumption that any paternalistic rule that is justified as promoting the utility of its intended

No doubt in Mill's most optimistic scenario, everybody gains under no paternalism. But to give much credence to this possibility requires us to take the more fanciful flights of Mill's rhetoric very seriously indeed. One of the more puzzling features of Mill's argument is his fixation of attention on the most wonderful possibilities of human development that might unfold if society scrupulously refrained from coercive interference with liberty in self-regarding matters. The serious worries about antipaternalism become apparent only in exploring worst-case scenarios and the likelihood they will befall at least some unfortunate persons. Leaving aside the completely unsupported conjecture that everybody gains by strict antipaternalism, we may worry that for all but the most hardboiled utilitarians, a policy that maximizes utility by making the worse off even worse off than they otherwise would be is unacceptable.

Here we touch one strand of the many and various justice-based objections to utilitarianism. Utilitarianism is distribution-insensitive : the distribution of utility across persons makes no difference whatsoever to utilitarian judgment of actions and policies except insofar as implementing one or another distribution of utility causally affects the total amount of utility in the long run. The insensitivity of utilitarianism to distributional concerns encompasses several distinct criticisms, one being that utilitarianism goes wrong in regarding only aggregate totals or averages of welfare while ignoring altogether the value of equal distribution of welfare among persons. The principle of equality of welfare holds that other things equal it is morally bad that social arrangements leave some people worse off than others. As stated, the principle is vague in not specifying how to weight

beneficiaries (agents who would otherwise harm themselves) must benefit the less able to a greater extent than it benefits the more able. There might be a self-regarding vice to which more able agents are peculiarly susceptible and the bad effects of this vice might be correctible by paternalism. All that my argument requires is that on the whole and on the average, paternalistic rules will be more to the advantage of less able agents. Second, my argument does not require the assumption either that the less able agents are always among the worse off or that paternalism (when justified as above) is never more to the advantage of better off agents. Both these assumptions are evidently false. A person who is very inept at managing his life may nevertheless be blessed by external advantages, such as wealth and wise friends, which insulate him from the bad effects toward which his self-regarding faults tend. And there might be self-regarding vices (such as consumption of very expensive but dangerously harmful recreational drugs) to which people who have better than average welfare prospects are peculiarly liable. In this regard my argument requires only the assumption that on the whole and on the average, being among the less able identifies an individual as facing below-average welfare prospects.

the values of more welfare against more equality of welfare and against other independant moral concerns when there is conflict among them. Maximin, Nicholas Rescher's effective average, and Paul Weirich's weighted utilitarianism are rules that offer alternative ways of resolving one aspect of this conflict, by proposing alternative rules of trade-off between the norms of "Maximize welfare !" and "See to it that welfare is equally distributed !" (7). (Weirich's suggested rule holds that the moral value of a gain of welfare varies directly with its size and inversely with the final relative welfare position of the person who gains. For a marginal gain of welfare, the rule holds that the value of the gain is inversely proportional to the initial relative welfare position of the person who gains. In other words, the worse off a person is in terms of welfare, the better it is, from a moral standpoint, to secure a gain of welfare for that person). On any of these distribution-sensitive alternatives to utilitarianism, which value equality of welfare for its own sake (8), it could be the case that even if a strict prohibition on paternalism should turn out to be welfare-maximizing, considerations of distributive fairness would recommend the enactment of some paternalist rules. Paternalism might then prove morally acceptable even if arguendo we concede to Mill all of the empirically controversial assumptions that he uses to build his case for an absolute no-paternalism rule. Mill's brand of utilitarian libertarianism, even if genuinely utilitarian, might be vulnerable to objections of fairness. Indeed, given the specific character of the arguments that Mill makes, a stronger conclusion is warranted : Mill's arguments for strict antipaternalism, to the extent they are accepted, simultaneously show that this antipaternalism is justified on straight utilitarian grounds and that it is unjustifiable on weighted utilitarian grounds that balance utility against equality.

(7) A maximin welfare policy holds that institutions should be arranged so as to maximize the welfare of the worst off. On maximin and on its leximin extension see Amartya SEN, "Equality of What?", reprinted in his *Choice, Welfare and Measurement* (Oxford, Basil Blackwell, 1982), pp. 353-369. On weighted utilitarianism, see Paul WEIRICH, "Utility Tempered with Equality", *Nous* 17 (1983), 423-439. On effective average, see Nicholas RESCHER, *Distributive Justice: A Constructive Critique of the Utilitarian Theory of Distribution* (Indianapolis, Bobbs-Merrill Co., 1966), pp. 31-38.

(8) The statement in the text is not quite right. The three views mentioned in the previous footnote all give greater weight to increasing the welfare of the worse off than to increasing the welfare of the better off. But if the position of those below average in welfare cannot be improved, a principle of equality of welfare would be averse to increasing the welfare of the better off, but weighted utilitarianism and the leximin version of maximin would favor such improvement even though it increases inequality.

Succeeding subsections will defend this preliminary conclusion against further arguments that are either raised by Mill or that might plausibly be urged on his behalf.

Effects on Third Parties : Mill on Progress

In the foregoing discussion the implicit assumption has been that in considering whether a paternalistic proposal is acceptable we are entitled to reckon only gains and losses to the persons whom we propose to restrict and to assume that the effects of restriction on third parties (parties not consenting to the transaction) will be nonsignificant. Mill emphatically does not make this assumption, so we need to see what happens to the argument when it is dropped. Mill observes that where something other than the individual's own character is the rule of conduct, "there is wanting one of the principal ingredients of human happiness, and quite the chief ingredient of individual and social progress" (p. 261). One of the sources of social progress that Mill has in mind is that in a society that eschews paternalism individuals will engage in a wide variety of "experiments of living" and will learn from the successes and failures of the other people's experiments. Nonpaternalism is the antidote to Chinese stationarity (pp. 273-274). In the short run people can adapt their life plans in order to take into account lessons learned from the experiences of others, and in the long run individual judgment selects the best from each generation's crop of experiments in living, so over time there is progress in individual culture. Whether my life plan succeeds or fails, my open engagement in it benefits others either through the cautionary lessons it imparts or as a model worthy of emulation. Refraining from paternalistic interference with my life has spillover benefits for those who would learn from it. As Mill puts it, "the worth of different modes of life should be proved practically, when any one thinks fit to try them" (p. 261).

Mill lays heavy stress on one possible case of progress by way of free experimentation : the less able agents benefit by observing the creative innovations exhibited by the realized life plans of the most able. He writes : "The initiation of all wise or noble things, comes and must come from individuals ; generally at first from some one individual. The honour and glory of the average man is that he is capable of following that initiative ; that he can respond internally to wise and noble things, and be led to them with his eyes open" (p. 269)

More generally, all people in society benefit in uncontroversial ways over the long run by the increases in material progress that are stimulated by innovations produced by persons of genius that would not have come about if the geniuses had not been free to go their own way unhampered by paternalistic restrictions. Since the geniuses who will contribute to material progress cannot be identified in advance of their contributions, in order to give free space to them one must give free space to all persons. Rising material progress is a predictable byproduct of refraining from paternalistic interference.

It should be evident that the fairness worries about Mill's antipaternalism are not allayed by this optimistic invocation of social progress. As before, I will not challenge the factual assertions that Mill makes, but simply check to see if they carry the argument as he imagines. The appeal to these spillover benefits of free experiment does not of course rule out a straight utilitarian justification of some paternalistic rules, because it might be the case that almost all the benefits of free experimentation are generated by permitting individuals broad but not unlimited freedom to carry out their life plans without suffering paternalistic interference. Mill himself mentions the possibility that a paternalist policy could be limited to blocking individuals from following a very small number of life plans universally condemned by experience (p. 281). For example, a government might generally permit the consumption of dangerous recreational drugs while banning the use of a few drugs that are reliably deemed to be extremely harmful and to offer extremely unfavorable short-term pleasure to long-term deterioration ratios. Sidestepping this issue of the bearing of spillover benefits on the adequacy of Mill's utilitarian argument, I want to determine whether these benefits might mitigate the fairness objection to Mill's strict antipaternalism. Any mitigating appearance is due to myopic concentration on just one of the several cases that need to be considered.

We need to consider socially valuable learning in the short run (affecting the present generation) and in the long run (affecting future generations) from successful and unsuccessful experiments in living carried out by more and less able agents. The learners in question who benefit from the experiments of others can of course be either more or less able. Moreover, it is obvious that mislearning also occurs, as when publicity about the stimulus to creative artists provided by selective drug usage prompts others to imitate their example, to their detriment.

A complication about the definition of "paternalism" arises here. If in the light of cases like that just mentioned we restrict the liberty of

experimenters in order to protect the welfare of those who are unduly likely to draw incorrect lessons for their own lives from the experiments, is such restriction paternalistic? The answer is "Yes" insofar as the restriction counts as a restriction of the liberty of the potential mislearners; the restriction would then be an instance of what Gerald Dworkin calls "impure" paternalism, in which the class of persons whose liberty is restricted is not identical with the class of persons for the sake of whom the restriction is imposed⁽⁹⁾.

However this definitional point is decided, the possibility of mislearning alerts us to a more general phenomenon, namely that more and less able agents have significantly different capacities to gather and correctly to interpret information generated by other people's experiments in living. Also, agents vary in their capacities to integrate the insights so gained into their own self-regarding decisionmaking. The benefits and costs of learning from free experimentation in the ways that Mill extolls will in many cases fall unevenly on more and less able individuals. Once again it appears likely on Mill's own assumptions that a strict ban on paternalistic restriction of liberty will in some of its applications benefit the better off at the expense of the worse off, and so be unfair even if utility-maximizing.

Appeals to the benefits to society from unsuccessful experiments in living and to the benefits to future generations to be gleaned from the experience of the free experiments of the present generation also raise worrisome fairness issues. Take the second appeal first. Mill views with Victorian optimism the long-run welfare prospects of a regime of strict antipaternalism. Imagine he is right about this. Strict antipaternalism might then be justified by utilitarian calculation even if this policy guide is disadvantageous for the members of the present generation, provided that their losses are offset by gains to future members of society. But this use of the present generation as cannon fodder for the future is morally problematic, particularly so on the assumption that with social progress the average welfare level will rise with each succeeding generation. Again Mill is proposing a taking from the worse off to advantage the already better off. At some trade-off ratio this might be acceptable policy, but straight utilitarianism gives no weight at all to equality of welfare and thus countenances sacrifices of present persons for the sake of future persons that many will regard as unfair.

(9) Gerald DWORKIN, "Paternalism", p. 22.

A similar point holds for beneficial social learning from the egregious imprudence of the less able agents. The “unblinking accountant’s eye” (¹⁰) of the utilitarian registers equally the loss suffered by the less able agent who stumbles in the gutter in a laissez-faire regime and the subsequent gain accruing to more able agents who shrewdly learn from this mishap how to avoid similar stumbles. But according to welfare egalitarianism adherence to antipaternalism in such cases amounts to unfairly using one person as a mere means to the benefit of others.

Mill’s discussion of spillover benefits is polemically slanted insofar as it highlights harmonious, rosy possibilities and ignores equally likely but more troublesome cases which pose acute conflicts of distribution.

Strong and Weak Paternalism

A plausible objection to my argument to this point is that Mill is at most (depending on what one makes of his discussion of voluntary slavery contracts (¹¹)) unequivocally opposed only to strong paternalism, that is, paternalistic restriction that involves an overriding of the restricted person’s own judgment of the ultimate goals he aims to fulfill by his action. The claim then is that Mill is not opposed to weak paternalism, that is, to restriction involving an overriding only of the restricted person’s own judgment of the means that are best suited to advance his chosen goals. But the worry that Mill’s antipaternalism might be unfair to less able agents dissolves — or so the argument goes — once it is understood that the scope of his antipaternalism is limited to the strong variety.

I have two responses by way of rebuttal : (1) the textual basis for interpreting Mill as a weak paternalist is shaky, and (2) in any event there are cases where the welfare egalitarian argument for strong paternalism would be persuasive against Mill, if Mill were indeed a clear-cut opponent of strong paternalism.

Mill discusses acting on a mistaken belief about the means needed to achieve one’s goals when he sketches the example of forcibly detaining someone who is about to venture unawares onto an unsafe bridge. The ground of interference is the reasonable presumption that the person is

(10) Bernard WILLIAMS, “A Critique of Utilitarianism”, in *Utilitarianism For and Against*, J. J. C. Smart and Bernard Williams (Cambridge, Cambridge University Press, 1973), p. 113.

(11) On this point, see BERGER, pp. 267-268, and FEINBERG, pp. 71-79.

unaware of the rickety condition of the bridge. Once the individual is apprised of the true condition of the bridge, "when there is not a certainty, but only a danger of mischief, no one but the person himself can judge of the sufficiency of the motive which may prompt him to incur the risk" (p. 294). This passage is puzzling. If we trust the person's prudential judgment, why not presume he can correctly decide whether the gain to himself from venturing on the bridge is worth the certain loss to himself from its collapse? (He may be bent on suicide and not inclined to explain this intention to would-be do-gooders like us). If we do not trust the person's judgment, why presume that even with correct information the person can reasonably factor the risky or uncertain prospect of venturing on the bridge into his prudential decisionmaking (¹²)? If our ignorance of the person's goals precludes anything beyond temporary detention for the purpose of conveying relevant information when there is a chance the bridge may collapse, why doesn't this same ignorance preclude sustained interference when the collapse of the bridge is a certainty? The principle suggested by Mill's treatment of the example is not tolerance of weak paternalism and only weak paternalism, rather something like "Paternalistic restriction of a person's liberty is always wrong unless it is done either (a) to prevent the person's certain death or (b) as a temporary expedient in order to supply information that appears to be importantly relevant to the person's choice and of which he appears to be ignorant" (¹³). Except perhaps with regard to immediately life-threatening situations, Mill does not seem ready to countenance paternalistic restriction that is needed to counteract the decisionmaking disabilities of less able agents. As Mill elsewhere remarks, "If a person possesses any tolerable amount of common sense and experience, his own mode of laying out his existence is the best, not because it is the best in itself, but because it is his own mode" (p. 270). One's own mode of laying out one's life includes choice of means as well as choice of ultimate goals, and the precondition of one's own mode being best is not that it is particularly

(12) Robert GOODIN, "Anticipating Evaluations : Saving People From Their Former Selves", chapter 3 in his *Political Theory and Public Policy* (Chicago, University of Chicago Press, 1982), pp. 39-47, discusses common defects in people's incorporation of risky and uncertain prospects into their decisionmaking.

(13) See the discussions of Mill's bridge-crossing example in GRAY, pp. 91-92, FEINBERG, pp. 124-127, and C. L. TEN, pp. 109-117. These authors believe the drift of this passage is toward soft antipaternalism, but for convenience I ignore this possible reading of Mill and postpone discussion of soft antipaternalism to section II.

well-chosen, much less that it could not be improved upon by intelligent interference, but rather that one has a "tolerable" level of decisionmaking talent — which in view of the unsafe bridge passage I would interpret minimally as the requirement that the agent be neither feeble-minded nor insane.

The welfare egalitarian fairness argument justifies some strong paternalism. In extreme cases the decisionmaking processes of less able agents that shape their fundamental personal values and goals may be defective to the point that coercive interference with their self-regarding choices may be justified as part of a process whereby individuals are encouraged to rethink their goals with fuller deliberative rationality. In such extreme cases strong paternalism can foster individuality, the value that Mill considers a prime ingredient and determiner of an individual's happiness. The problem for the weak paternalist position is that it concedes too much ground to the opposition. Suppose that factual ignorance, errors in deductive reasoning, failure of uptake from statistical data, framing errors, the basing of choice on nonevidential cognitive dissonance reinforcement, and the like (¹⁴) can sufficiently distort a person's judgment as to how best to achieve his ends to warrant paternalistic interference — as the weak paternalist admits. Why then don't the same irrational proclivities, operating through the processes that determine people's basic preferences, sometimes warrant strong paternalistic interference — even though the weak paternalist must on principle deny this possibility? Moreover, the same distributive fairness worries that bolster the argument for acceptance of weak paternalism will bolster the argument for acceptance of strong paternalism as well.

Bias of the Public Authority Contemplating Paternalism

As already noted, Mill's aim in *On Liberty* is to devise a policy regarding the restriction of liberty that will be optimific for actual conditions expected in modern societies, rather than for an imaginary ideal case. Mill's discussion does not abstract from the problems that beset a public

(14) See *Rational Choice : The Contrast between Economics and Psychology*, eds. Robin M. Hogarth and Melvin W. Reder (Chicago and London, The University of Chicago Press, 1986). See also Richard THALER, "The Psychology of Choice and the Assumptions of Economics", Working Paper RR-3, Center for Philosophy and Public Policy, University of Maryland.

authority contemplating paternalistic restriction due to the limited information at its disposal. Nor in trying to decide the merits of strict antipaternalism should we assume (a) that those subject to paternalist rules will comply perfectly with them, (b) that the administrative costs of paternalistic enforcement are insignificant or that the question of public funding of these costs poses no problems of efficiency or fairness, (c) that the officials charged with administering paternalistic rules will necessarily be motivated to carry out their duties in the spirit of the rules or that they will be competent to do so, or (d) that the public authority responsible for legislating paternalistic rules is itself motivated solely by impartial respect for principles of public morality.

Here I wish to raise the question whether or not the inappropriateness of making assumption (d) gives reason to think that strict antipaternalism is an expedient of fairness not perhaps in all abstractly possible circumstances but in anticipated conditions in the modern world. Very roughly the argument would go as follows. A government effectively controlled by an elite segment of the more able and the better off agents is unlikely to have much sympathetic understanding of the conditions of life and requirements for happiness of the less able and the worse off members of society⁽¹⁵⁾. If such a government enacts paternalistic legislation, wittingly or unwittingly it will legislate in a highhanded manner that is not likely to advance the supposed beneficiaries of the legislation. Strict antipaternalism is then a hedge against the benevolence of the crocodiles. In the event that a democratic government is effectively ruled by a majority of citizens, the suspicion of bias in the legislature now suggests legislative bias against the minority of more able citizens. Again, the conclusion of this line of thought is that strict antipaternalism in the circumstances for which Mill proposed it is recommended, not condemned, by fairness.

These conjectures do not succeed in establishing a case for the fairness of strict antipaternalism. First, if we assume a minority elite government biassed toward its own self-interest, this bias would appear to be just as likely to express itself in selfish laissez-faire as in selfish pseudopaternalism. No presumption in favor of antipaternalism emerges from this train of thought. Second, the introduction of majority rule political processes

(15) See John Stuart MILL, *Considerations on Representative Government*, in *Collected Works*, vol. 19., ed. J. M. Robson (Toronto and Buffalo, University of Toronto Press, 1977), p. 405.

in itself has no discernible tendency that I can perceive to increase the likelihood that any paternalist legislation will reflect a bias against a distinct minority of more able citizens. It all depends on the relation between the distribution of prudential ability among the citizens and the voting blocs that are likely to coalesce into majority coalitions of voters. Prior to investigation the conjecture that a majority of more able votes will vote against a minority of the less able down and out looks just as plausible. Third, even if a sophisticated political sociology did inform us that in a given society majority rule is likely to produce paternalist legislation that favors the interests of the less able over the interests of the more able, a welfare egalitarian may view this tendency with cautious approval so long as it works in practise to further equality of welfare at acceptable cost of overall welfare.

It's His Fault

Consider the commonsense view that we do no injustice to a person if we decline to aid him by coercive interference with his liberty against his will, at least when the harm he will suffer in the absence of our intervention will be due to his own fault. The varieties of welfare egalitarianism that I have deployed against Mill all systematically ignore the fault of the agent as a determinant of his welfare in recommending the fairness of equal distribution of welfare (or the greater priority of advancing the welfare of the worse off). Paying no heed to such an extreme conception of fairness in his argument for the liberty principle, Mill stands on the firm ground of common sense, one might hold.

My criticism of Mill can accommodate the common sense view just mentioned. I need not contest the notion that a person can come to grief through his own doing, in such a manner that society owes him no compensation and bears no responsibility for the mess that the individual has made of his life simply by virtue of having granted him freedom which he has abused. Suppose we accept a principle of equality which states, "Other things equal it is a bad thing if social arrangements render some persons worse off than others through no fault of their own" (16). Whatever conception of fault one adopts, inequalities of welfare that arise

(16) This formulation is borrowed from Derek PARFIT, *Reasons and Persons* (Oxford, Oxford University Press, 1984), p. 26.

through the individual's own fault as judged by that conception will neither violate the principle of equality nor count as unfair.

But on anybody's conception of fault the prudential disabilities that separate more and less able agents are surely in very considerable part due to accidents of genetic endowment and variously favorable early childhood circumstances that do not lie within the agent's control and for which he cannot be either praiseworthy or blameworthy. So even if we accept that it is sensible to attribute some prudential failings of individuals to personal fault, these attributions cannot reconcile us to regarding as fair the great bulk of inequalities of welfare that separate more and less able agents. Paternalism remains in the running as one morally appropriate response to some of these pervasive and disquieting inequalities. Nothing in *On Liberty* should assuage this sense of disquiet.

II. FEINBERG'S SOFT ANTIPATERNALISM

Voluntary Choice

So far I have not considered the common position that is usually labelled "soft paternalism" and that Joel Feinberg prefers to call "soft antipaternalism". Since Feinberg's recent volume, *Harm to Self*, in his magisterial work on *The Moral Limits of the Criminal Law* is in my opinion the best work on paternalism that we have, my discussion of this position focuses on his analysis (17).

Feinberg argues that what is wrong with paternalistic restriction of liberty is that it violates a compelling ideal of personal autonomy, but that paternalistic restriction of action proceeding from choice that is substantially nonvoluntary does not violate the agent's autonomy, hence may sometimes be permissible. We might say that interference with substantially nonvoluntary conduct does not constitute restriction of a person's liberty against his will. A "substantially nonvoluntary" choice of action is

(17) Besides *Harm to Self*, FEINBERG's *The Moral Limits of the Criminal Law* includes two other published volumes, *Harm to Others* (1984), and *Offense to Others* (1985). A fourth volume, *Harmless Wrongdoing*, is forthcoming.

My discussion of soft antipaternalism is heavily indebted to Dan BROCK, "Paternalism and Promoting the Good", in *Paternalism*, ed. Rolf Sartorius, pp. 237-260. See also Daniel WIKLER, "Paternalism and the Mildly Retarded", *Philosophy and Public Affairs* 8 (1979) : 377-392. I am also indebted to Feinberg's good criticisms of some of my own earlier thoughts on this topic, pp. 128-132.

one that departs too far from the ideal of a perfectly voluntary choice. Feinberg renders this ideal as follows : the choice of an adult person is perfectly voluntary if and only if (1) the chooser is competent (not insane, severely mentally retarded, or comatose), (2) the choice is not made under coercion or duress, (3) the choice is not made "because of more subtle manipulation" (such as posthypnotic suggestion), (4) the chooser is not making his choice because of ignorance or mistaken belief about the circumstances in which he acts or the likely consequences of the various alternative actions open to him, and (5) the chooser "does not choose in circumstances that are temporarily distorting" ("not impulsively [on impulse] ; not while fatigued ; not while excessively nervous, agitated, or excited ; not under the influence of a powerful passion., e.g. rage, hatred, lust, or a gripping mood, e.g. depression, mania ; not under the influence of mind-numbing drugs, e.g. alcohol ; not in pain, e.g. headache ; not a neurotically compulsive or obsessive choice ; not made under severe time pressures") [p. 115]. In short, a perfectly voluntary choice is one that is not marred by any of a miscellaneous set of features that tend to prevent choice from "faithfully representing" the agent "in an important way, expressing his settled values and preferences" (¹⁸).

To determine whether a particular choice falls too short of this ideal to be voluntary enough to be appropriately protected by antipaternalist principle, Feinberg urges that we apply a variable standard depending on the situation. To appreciate the rationale of a variable standard, note that the point of soft antipaternalism is not to prevent the doing of actions of low degrees of voluntariness as such, but rather "to prevent people from suffering harm that they have not truly chosen to suffer or to risk suffering" (p. 119).

Finally, Feinberg reminds us that sometimes restriction of individual liberty in self-regarding matters is needed to establish whether or not action is proceeding from choice that is voluntary enough in the circumstances. The principle of soft antipaternalism thus holds : "The state has the right to prevent self-regarding harmful conduct when but only

(18) But as Feinberg observes, a person may with perfect voluntariness make a choice that is out of character or that goes against her hitherto settled values and preferences. For a discussion of paternalism that defends strong paternalism when it protects the agent from choices that fail adequately to represent faithfully her settled values, see John KLEINIG, *Paternalism* (Totowa, N.J., Rowman and Allanheld, 1983), chapter 3.

when it is substantially nonvoluntary, or when temporary intervention is necessary to establish whether it is nonvoluntary or not" (p. 126)

Solving the Problem ?

To some extent this soft antipaternalist position assumes away the limiting conditions on choice of principle that persuade Mill to embrace (less than wholeheartedly, it is true) hard antipaternalism. Mill writes, "It is easy for any one to imagine an ideal public, which leaves the freedom and choice of individuals in all uncertain matters undisturbed, and only requires them to abstain from modes of conduct which universal experience has condemned. But where has there been seen a public which set any such limit to its censorshiup ?" (pp. 283-284). As Mill formulates the problem, we are to determine the best constitutional norm for the guidance of such legislators as we are likely to get, not a norm that would be best for ideal legislators of our imagining. We are also evidently meant to assume severe limitations on the extent to which the legislature can tailor its restrictions to the different decision-making and decision-executing abilities of its citizens. If we could exactly tailor restrictive social laws to individual deficiencies, we could do the best for each citizen, and the tradeoffs between the welfare of the worse off and the welfare of the better off that I have envisaged would not be necessary, so neither would fairness problems arise in resolving such tradeoff issues. We have to ask whether soft antipaternalism's way of assuming away the problem should count as solution or evasion of the difficulty.

Feinberg suggests two procedures for tailoring legal rules to individual levels of competence. One is to exempt agents who act soft-paternalistically with good reason from the ordinary civil and criminal law penalties to which they might otherwise be liable (pp. 154, 157). For example, the law might specify that a good Samaritan who intervenes forcibly to prevent an acquaintance from ingesting a drug that will harm him is exempt from liability to a criminal charge of battery. Another suggested procedure, urged in connection with the idea of placing persons who initiate or threaten self-harming actions under temporary restraint for the purpose of establishing whether the choice of these acts would be substantially nonvoluntary, is to institute commissions of inquiry and the like for deciding such matters (pp. 125-126, 128).

The first suggestion strikes me as sensible, though its impact in increasing the sensitivity of paternalistic restriction to individual cases is

likely to be limited. If the suggestion were implemented, the law would permit but not require soft-paternalistic intervention by persons who are well-placed to make a sound judgment about the case at hand. Surrounded by neighbours and strangers who would rather not get involved, the individual in need of rescue would be unlikely to get it. One's chances of rescue from substantially nonvoluntary self-damaging courses of action would depend largely on the astuteness, caring, and decisiveness of one's friends, relatives, and acquaintances. Notice also that due to assortative mating and more broadly the tendency for people to associate with others of similar accomplishment and status, the better off may stand to gain more than the worse off from criminal-law permission of private paternalistic interference.

In contrast, the proposal to establish voluntariness-determining boards of inquiry appears to be of dubious utility. Reliable evidence about such matters would be hard to come by — more difficult, I should think, than the problem that arises in criminal trials of determining whether the accused person's conduct is voluntary enough to establish his personal responsibility for it. Whether the reasons for doing what one proposes to do are confused or mistaken depends on the goal one hopes to achieve, so a person intent on establishing her right to do what she wants would have a strong incentive to dissemble about her intended goal and to present cooked-up, reasonable-sounding but sham reasons for her proposed course of action. Such dissembling would be difficult to detect. A negative verdict on the voluntary character of an individual's proposed course of action in a self-regarding matter would in the nature of the case often inflict a severely humiliating blow on the self-esteem of the individual. Such a finding by an official state agency would be bound to carry the sting of insult. Moreover, the fact-finding boards would be expensive to administer and would have difficulty convincing the general public of the reliability and consistency of their procedures. Enforcing the negative verdicts of these boards would be difficult and expensive, but necessary if the boards are not to arouse public contempt. Apparently what the soft antipaternalist envisages is selective prohibition of activities : whereas the use of dangerous recreational drugs, for instance, would not be generally prohibited, use of such drugs would be forbidden to people whose choice of drug usage was deemed nonvoluntary. One pictures a bureaucratic nightmare. Under current law we do selectively prohibit certain activities, as by FDA rules that forbid the purchase and consumption of certain

drugs except with a doctor's prescription, but in this example the standards used to decide who is prohibited from engaging in the activity and who is not are reasonably uncontroversial and application of the standards by medical experts does not (usually !) give rise to intractable dispute. It is unlikely in the extreme that either of these conditions would be met if the system of leaving the determination of individual voluntariness to state boards of inquiry were instituted. This proposal looks to be of doubtful utility, and to be very unlikely substantially to reduce the screening problem that sets Mill's agenda (¹⁹).

It should be noted that the extent of one's willingness to support the proposal to institute voluntariness-determining boards of inquiry despite its administrative cumbersomeness, likely high cost, and low prospect of delivering reliable verdicts will depend on the extent of one's commitment to the value of personal autonomy. The more one regards it as a terrible thing for the state or society to infringe the personal autonomy of an individual citizen acting voluntarily enough in a self-regarding matter, the more one will reasonably be willing to bear the moral costs of governmental procedures that attempt to discriminate voluntary and nonvoluntary self-harming actions for the purpose of leaving the one free while restricting the other. To settle accounts decisively with soft antipaternalism will require an evaluation of the ideal of personal autonomy that the soft antipaternalist is concerned above all to defend. (On this, see the final subsection of this paper).

There turns out to be no adequate reason for assuming that paternalistic rules can always be made variable in their application depending on individual determinations of decision-making competence. (To clarify : Feinberg himself does not make this assumption). The problem that I have raised for the hard antipaternalist must be faced by the soft antipaternalist as well.

Another Fairness Issue

The soft antipaternalist must face the issue of the fairness of the welfare-distributing impact of this proposed standard for justified pater-

(19) The doubts that I raise about the likely usefulness of state boards of inquiry for determining voluntariness do not extend to the ordinary practise of allowing state agents to intervene forcibly to determine whether apparently nonvoluntary action really is so — for example, checking to see if the person is acting on obvious and choice-determining factual ignorance, as in Mill's bridge-crossing example, or to see if the person is acting under the influence of alcohol or another choice-distorting drug.

nalism. But given that the soft antipaternalist need not be committed to the maximization of utility come what may, I have so far given no reason to think that he must be committed in advance to a policy that is tilted toward increasing the welfare of the already better off segment of society at the expense of the already worse off. When incomplete information and the unfeasibility of perfect screening force us to treat a group of bad choosers and good choosers alike, either restricting all or restricting none, the soft antipaternalist would appear to be entirely free to use an appropriately distribution-weighted consequentialist principle (or for that matter a nonconsequentialist principle) in deciding whether or not to impose a restriction (20).

This appearance strikes me as deceptive. The situation is confusing because soft antipaternalism is crucially vague. Once we try to pin down this vagueness, the position either reduces to single-party welfarist consequentialism (explained below) or it does not. If it does, I have no quarrel with the position except to note that the label is misleading advertising, because soft antipaternalism so construed dissolves into ordinary paternalism. If it does not, then I contend that soft antipaternalism rests on a notion of personal sovereignty or personal autonomy that is bound to have undesirable distributive implications. Whereas my objection to Mill's antipaternalism is that it is too utilitarian, my objection to soft antipaternalism is that it is not utilitarian enough !

Welfarism and Soft Antipaternalism

By *Single-party welfarist consequentialism* I understand the view that insofar as we have to choose action for situations that involve no conflicts of interest among persons, but simply involve one person whose good might be advanced by what we do, in deciding how our action might best advance the person's good we are to be guided entirely by that very person's own conception of the good. Insofar as we are aiming to act for the good of another, that person's values, tastes, and preferences (perhaps corrected by hypothetical ideal deliberation) entirely determine the goal we should seek (21).

(20) Feinberg discusses examples that turn on the unfeasibility of perfect screening of good choosers from bad choosers at pp. 18-21 and p. 128. He does not suggest that the liberty of the good choosers must take priority when protecting their liberty must be at the expense of the welfare of bad choosers, who could be helped by an inclusive paternalistic rule.

(21) On hypothetical ideal deliberation as determining the preferences that constitute

The reason that soft antipaternalism tends to melt into plain old-fashioned paternalism is that its core idea of a substantially nonvoluntary choice is elusive, and not clearly distinct from the straight utility-maximizing notion of a choice that ought to be forcibly interfered with for the agent's own good. In sorting out these notions, some distinctions will prove helpful. First, we have the idea of a *nonoptimizing choice* by an agent, a choice that (if acted upon) would fail to maximize the agent's own rationally expected good, provided that this shortfall would not be counterbalanced by likely good effects of the agent's act on other persons. Of course the class of nonoptimizing choices does not coincide with the class of choices that on single-party welfare consequentialist grounds ought to be interfered with. After all, there may be no reliable way to improve on an agent's imperfect choice. Second, let us define a *substantially nonoptimizing choice* as one that (if acted upon) would fail to maximize the agent's good (without producing counterbalancing benefits for others) to such an extent and in circumstances such that paternalistically restricting the agent from acting on that choice would be feasible and, from the standpoint of maximizing the agent's own good, desirable.

A substantially nonvoluntary choice, according to Feinberg, is one that is not "voluntary enough", where being voluntary enough is a function of the degree that the choice falls short of perfect voluntariness, the likelihood that acting on the choice will bring harm, the magnitude of the harm thus risked, and the extent to which any harm that might be done will be irreversible. The function is not specified ; this is evidently a matter to be left to the discretion of the reasonable legislator or other agent contemplating paternalistic intervention (p. 117). An imperfectly voluntary choice is one that fails to meet the standard of perfect voluntariness, but imperfectly voluntary choices may be optimizing and even if nonoptimizing, may not be substantially nonoptimizing.

The question I find it difficult to decide is whether the class of choices that are substantially nonvoluntary, hence not protected from paternalistic interference by the soft antipaternalist principle, coincides with the class of substantially nonoptimizing choices. Certainly the features of choices that by definition tend to render them substantially nonvoluntary are

a person's good, see Richard B. BRANDT, *A Theory of the Good and the Right* (Oxford, Oxford University Press, 1979), pp. 110-129 ; and David GAUTHIER, *Morals by Agreement* (Oxford, Oxford University Press, 1986), pp. 26-32.

features that in fact tend to render them substantially nonoptimizing. The situation seems interpretively indeterminate.

Feinberg leaves us in no doubt that in his opinion soft antipaternalism does conflict in a morally significant way with the ordinary propaternalist position that paternalism is justified just in case it maximally promotes the good of the intended beneficiary (without involving counterbalancing excess costs to others). The crux of the matter in his view is whether one takes personal sovereignty as a constraint never to be violated (see, e.g., pp. 57-62, 157, 184-186). Whence this confidence of judgment? The explanation may be simply that I have been assuming so far that we understand the reference to the good of the agent in the definitions of "nonoptimizing choice" and "substantially nonoptimizing choice" in a welfarist way, as determined entirely by the agent's own personal values and preferences (perhaps as these would be corrected by ideal deliberation). With that assumption in place, it might seem that any paternalistic acts we propose will be justified by some failure of the agent either to take effective means to achieve her goals or to identify her goals in a reasonable way, without factual error, reasoning error, or emotional instability (but see the discussion in the next subsection). But these are all voluntariness-reducing features of choice.

However, if we drop that assumption, then the propaternalist may be a perfectionist who believes he knows what goods are objectively worth achieving and that restriction of a person's liberty to enable him to achieve more of those goods may be justifiable quite apart from any showing that the person's choices are imperfectly voluntary. From a perfectionist standpoint, there may be good ground for paternalistic restriction even of an agent's perfectly voluntary choice. From a welfarist standpoint, such grounds will be difficult to identify; there may be none.

Insofar as Feinberg's soft antipaternalism takes perfectionism as its target of opposition I have no quarrel with it. Nor do I dispute Feinberg's assertion that soft antipaternalism is significantly antipaternalist. But all of this leaves completely open the question that interests me: Whether the paternalism that could be justified on single-party welfarist consequentialist grounds differs at all from that paternalism that could be justified on soft antipaternalist grounds. This is the question to which, I claim, soft antipaternalism as worked out by Feinberg permits no determinate answer.

A brief digression will clarify what divides the welfarist and her perfectionist opponent. If the welfarist simply identifies a person's good

with satisfaction of her actual preferences, the contrast with the perfectionist is obvious ; so let us consider an ideal welfarist who takes it that a person may misidentify her "true" values. According to the ideal welfarist, a person's good is maximal satisfaction of the personal values and preferences that she would have if she were to reflect about this matter with full pertinent information, while making no reasoning errors, and in a calm mood. But the ideal welfarist is completely open-minded about the outcome of this ideal deliberation procedure and does not assume it likely that all persons would converge on the same values and preferences as ideal deliberation was approached. (This is one way to represent Mill's espousal of diversity and individuality in chapter 3 of *On Liberty*. We have different individual natures and to some extent the good for each of us may well vary). In contrast, the perfectionist can be represented as committed to the belief that as the limit of ideal deliberation about one's good was approached, all persons would converge in agreeing on a certain conception of human flourishing, which constitutes the objective good for humanity. The perfectionist is willing to impose paternalistically on a person who with full voluntariness affirms a conception of her good that disagrees with the perfectionist's notion of objective good. Charitably construed, such imposition rests on the conjecture that there would be convergence after ideal deliberation and that the view converged upon would be substantially the same as what the perfectionist now upholds. So interpreted, the perfectionist seems to me to be exuding unwarranted self-confidence about epistemic matters to which at present we have only very restricted and doubtful access.

Welfare versus Autonomy

I want tentatively to explore further the issue of whether there would be any important practical disagreement between soft antipaternalism and single-party welfarist consequentialism. I describe below three cases in which it might be plausible to suppose such disagreement might arise.

Self-abasing Benevolence. Imagine that a person chooses with perfect voluntariness to sacrifice a great amount of her own welfare for a small net increase in the welfare of others⁽²²⁾. It is probably better to concentrate on examples of inefficient self-sacrificing behavior done for the purpose

(22) See Michael SLOTE, "Morality and Self-Other Asymmetry", *Journal of Philosophy* 81 (1984) : 179-192.

of benefitting strangers or acquaintances rather than close relatives or friends, because in some cases of the latter (e.g., a parent's sacrifice for a child) it can be very difficult to discern to what extent the agent's own self-interest encompasses the advantage of the person sacrificed for, so that in acting to benefit another one is also acting to benefit oneself. In cases of self-abasing benevolence the person seems to reveal a conviction that the welfare of others is more worthwhile than her own welfare, and this conviction need not be based on any factual error, error in reasoning, or the like. Nonetheless the fundamental welfarist idea is that, in Bentham's words, "Everybody is to count for one, nobody for more than one". No doubt a person's steady and abiding desire to benefit others even in ways that greatly diminish aggregate welfare typically could not be blocked from expressing itself except through draconian restriction of liberty that would itself do more harm than good. In some cases, however, the welfarist contemplating paternalistic intervention to stop self-abasing benevolence may correctly judge that forcible intervention, which strongly conveys to the restricted person the message, "We judge your welfare to be equally as worthwhile as anyone else's welfare", would itself shock the person in a way that imparts a greater belief in her own worth and destroys the propensity to benefit others at disproportionate cost to herself. In such cases the welfarist will recommend paternalistic restriction that the soft antipaternalist, I suggest, would on principle be required to eschew. Here I side with the welfarist against the soft antipaternalist (23)

Voluntariness-reducing Factors in the Choice Situation That Are Themselves Voluntarily Chosen in Advance by the Agent. Suppose that in a cool hour one decides that if the boss refuses one's request for an increase in pay, one will say in reply spontaneously whatever comes into one's head, regardless of one's emotional state at that moment. Or suppose that a climber deliberately sets out on an adventurous climb under what look to be treacherous conditions, without consulting an available weather report that would certainly have a bearing on the safety of the

(23) Notice, however, that a broader range of cases must be considered than is included in this discussion. Suppose that a person with perfect voluntariness chooses to dedicate herself to a goal, such as saving the whales from extinction, that is not sought as part of her own happiness or as part of the happiness of other people. I do not wish a welfarist principle to justify paternalism in such a case, but I cannot consider here the complications that must be introduced to attain this result. See my "Equality and Equal Opportunity for Welfare", forthcoming in *Philosophical Studies*.

planned enterprise. Her rationale for this apparent heedlessness is that she wishes to climb under conditions of extreme uncertainty, where some of the risks encountered are not known about in advance, so that she will have to exercise her wits and skill in a situation that will require quick and sure improvisation. A third sort of example occurs when gaining more information known to be relevant to a decision one is taking is itself painful for the agent. Suppose that you know that there is very little chance that learning about the safety hazards of your job will induce you to switch careers, but it is virtually certain that learning about such information and rendering it emotionally vivid to yourself will be distressing, perhaps to the point of inhibiting job performance. So one stops gathering job-safety information. It may be that if we fully describe these situations it will turn out that the supposition that the agent makes the advance decision with perfect voluntariness is contradictory. I am not sure about this, but I doubt it. Be that as it may, I note that we can imagine variants of the above stories in which a friend of the agent acquires information about the situation which renders it very probable that what the agent is planning to do will be substantially nonoptimizing. By welfarist standards, the friend would then be justified in intervening forcibly to restrict the agent's liberty for her own good. Here too there might be a conflict between the implications of soft antipaternalism and single-party welfarist consequentialism regarding the justifiability of paternalism in such cases.

Prior Restriction of the Choice Set That Partially Forms Individual Preferences (²⁴). Let us distinguish two kinds of paternalistic restriction of liberty : preventing someone from doing what he already wants to do, and preventing someone from ever entertaining the option of doing a thing and forming a desire for it by prior restriction of the choice set. For this distinction to have any practical significance, it must be assumed that people's basic preferences (what one prefers for its own sake, not as a means to further aims) are partially determined by the range of initial options available to them. A perhaps farfetched example illustrating the distinction would be this : Suppose the government by dint of great restriction of liberty of the present generation of smokers is able to enforce a completely effective ban on smoking, so the practise completely disap-

(24) This topic is considered from another viewpoint in Jon Elster, "Sour Grapes", chapter 3 in *Sour Grapes : Studies in the Subversion of Rationality* (Cambridge, Cambridge University Press, 1983), pp. 109-140.

pears after one generation. Ignore the possible unfairness to the first generation and consider the effect of this entirely successful ban on smoking on members of future generations. Let us suppose the idea of smoking effectively dies out. People read about the practise, as an historical curiosity, but nobody sees others smoking, so nobody ever forms the desire to smoke. I take it that with respect to the later generations of nonsmokers, the ideal of personal sovereignty underlying soft antipaternalism would deem unacceptable the coercive manipulation of people's choice sets in this fashion. Such coercion will count as a factor that reduces the voluntariness of people's nonchoice of smoking. Personal sovereignty is assaulted just as much by coercion that reduces the voluntariness of people's choices in self-regarding matters as by coercion that prevents people from acting on their voluntary self-regarding actions. If this is so, then we have here another possible instance of practical disagreement between the single-party welfarist consequentialist and the soft antipaternalist.

None of the cases described above seems very clear-cut to me. Hence I remain unsure whether soft antipaternalism really does end up being extensionally equivalent to single-party welfarist consequentialism or not. I want nonetheless to explore the possibility that the two views really do conflict and that soft antipaternalism condemns acts of paternalism that are justifiable on welfarist grounds.

Two Conceptions of Autonomy

The problem for the soft antipaternalist as I see it is this. We are to imagine that an individual's choice is substantially nonoptimizing, which means that paternalistic interference can improve on the expected outcome that is foreseeable if the individual is permitted to act on that choice. Moreover, paternalism can improve on the expected outcome as that would be judged from the self-interested standpoint of the individual whose conduct we propose to restrict. We propose to restrict the individual's self-regarding freedom in order to advance goals that the individual herself either is seeking to fulfill or would seek after ideal deliberation. In the name of personal sovereignty the soft antipaternalist opposes interference. Despite the fact that paternalism would advance the agent's own conception of her good, the agent's choice is voluntary enough, so the right of personal sovereignty trumps the agent's good. The individual's sovereignty over her own life gives her the absolute right that her liberty

to carry out a voluntary choice in a self-regarding matter should never be coercively restricted for paternalistic reasons — so says the soft antipaternalist.

But the agent's own conception of her good will presumably include a conception of autonomy or personal sovereignty and a weighting of this good against other goods that the agent values for herself. If the agent's choice really is substantially nonoptimizing, then paternalism can better advance the agent's good than her own execution of her own choice, where this good of the agent includes her own relative ranking of autonomy against other values. On what ground does the soft antipaternalist override the agent's own placing of autonomy in her personal scheme of values ? From a single-party welfarist consequentialist standpoint, this insistence that "autonomy trumps" is just another species of perfectionist imposition of values on the agent in defiance of the agent's own considered evaluation. To take an extreme case, an agent might conceivably have personal values and preferences that assign strict lexical priority to the desire never to be subjected to paternalistic interference not justified by soft antipaternalist standards. In this case, the welfarist and the soft antipaternalist will agree that if the agent's choice is "voluntary enough", so that this overriding desire for autonomy comes into play, no paternalism against the dictates of soft antipaternalism could be justified (²⁵). But if the agent does not happen to have this extreme weighting of preferences, why does respect for her autonomy plausibly require giving greater weight to her autonomy than she herself gives it ? An alternate way to put this point is to note that welfarist consequentialism has its own notion of autonomy — namely, insofar as one is trying to further an agent's good, be guided by that very agent's conception of good and nothing else (²⁶). It is far from clear that, having conceded so much to this conception, the soft antipaternalist can successfully defend a rival view of autonomy or sovereignty.

Feinberg skillfully characterizes the soft antipaternalist ideal of personal sovereignty in terms of a comparison to the sovereignty of states

(25) On this point I have learned from Danny Scoccia's Ph.D. dissertation, on file at the University of California at San Diego Library.

(26) Cf. John Harsanyi's notion of "preference autonomy" — "the principle that, in deciding what is good and what is bad for a given individual, the ultimate criterion can only be his own wants and his own preferences". See John HARSANYI, "Morality and the Theory of Rational Behavior", in *Utilitarianism and Beyond*, eds. Amartya Sen and Bernard Williams (Cambridge, Cambridge University Press, 1982), pp. 39-62 ; see esp. p. 55.

(pp. 47-57). A political metaphor is at work here. Just as a state is sovereign over its territory and has the right that its sovereignty not be infringed, so the individual is sovereign over his own life, or more exactly over his self-regarding actions, and has the right of self-rule that should be just as inviolate as state sovereignty. The political metaphor is supposed to clarify the concept of personal sovereignty even if we have doubts that states morally are entitled to the strong rights of sovereignty they have claimed. The ideal of personal sovereignty is that individuals should have the sovereignty over their own lives that states have claimed (rightly or wrongly) over their own territory (p. 51).

But soft antipaternalism as characterized by Feinberg actually backs away from this bold claim. The sovereignty that states claim is surely not the right that their substantially voluntary choices should not be subject to forcible external interference. Suppose the legitimate government of some country enacts some policy that will affect only its own citizens, and is enthusiastically supported by them, but is in fact a stupid policy that will defeat the goals that the government clearly announces and that the citizens foolishly believe will be served by it. Our ordinary notion of political sovereignty holds that a benevolent neighboring nation is not entitled in these circumstances to send its troops over the border in a surgical raid that will correct this self-defeating policy and inaugurate an imposed policy better calculated to achieve the goals the government and citizens so clearly desire. Political sovereignty entitles a nation to make its own mistakes and wallow in the consequences of its own stupidity at least where only its own citizens are directly harmed thereby. Personal sovereignty as explicated by Feinberg is different in this respect. According to Feinberg's idea of personal sovereignty, an individual's self-regarding choice that threatens harm to himself and falls "too far" short of perfect voluntariness may be subject to coercive interference on paternalistic grounds, consistently with personal sovereignty. But why not then let the individual's own values fully determine when paternalistic interference may justifiably be executed, consistently with personal sovereignty⁽²⁷⁾? Soft antipaternalism seems to be an unacceptable half-way house.

(27) This question is not rhetorical. We need to distinguish two cases. When the soft antipaternalist eschews paternalist intervention that would effectively promote the fulfillment of the person's actual preferences, the preference autonomy case for intervention is clear and, to my mind, persuasive. But suppose the person's actual preferences are not well considered, and the ideal welfarist proposes paternalist intervention in order effectively to

Moreover, to return at last to the issue of distributive fairness, it is easy to see that imputing to individuals a strict lexical preference never to suffer paternalistic interference in violation of the soft antipaternalist principle, regardless of individuals' own evaluations of this matter, will predictably work to the advantage of the haves and to the disadvantage of the have-nots. Good choosers will predictably fare better under a regime of soft antipaternalism than bad choosers, because the imposed value of sovereignty will be more likely to constrain the state or other would-be interferers from carrying out paternalistic acts that really will work to the benefit of the latter. The distributive dimension of the paternalism issue, hitherto largely ignored, is significant. Once acknowledged, it cuts against the advocacy of any form of strict antipaternalism (28).

University of California, San Diego.

promote the fulfillment of the preferences the person would have after ideal deliberation. These nonactual preferences may seem a more doubtful basis for intervention. I suggest that in this case the preferences that determine a person's good are the preferences she would have if she were to deliberate about her preferences in an ideal way with full information *including pertinent information about her actual resistance to advice regarding the rationality of her preferences, the likelihood that her actual preferences will ever approximate to her ideally considered preferences, the costs of bringing about preference change toward her ideally considered preferences, and so on.* Only in (the rare) cases where the person contemplating paternalistic intervention has a strong epistemic warrant for claiming to know the person's hypothetical ideally considered preferences will appeal to these preferences justify such intervention.

(28) This point has implications which cannot be developed here. I believe that any credible distributive egalitarianism will require assumptions that will tend to undermine any strict advocacy of antipaternalism. See in this regard Kai Nielsen's vigorous criticism of Robert Nozick's libertarianism, in *Equality and Liberty: A Defense of Radical Egalitarianism* (Totowa, N.J., Rowman and Allanheld, 1985), part IV. Nielsen's discussion does not challenge Nozick's antipaternalism.