Northwestern University Political Philosophy Conference

James Sherman

Graduate Student, Dept. of Philosophy, University of Texas at Austin

1 University Station C3500, Austin, TX 78712

1.512.569.3597

j-sherman@mail.utexas.edu

ABSTRACT:

John Stuart Mill presents his harm principle in On Liberty as a strict limitation on state interference in the lives and actions of individuals. The state, he claims, may only properly interfere with the actions of an individual in order to prevent harm to others. The utilitarian argument with which he supports the principle, however, gives to the term "harm" an alarmingly wide range of meanings. The argument leaves open the possibility of advocating extensive state control over individual life while still legitimately claiming to endorse Mill's liberal political philosophy. This possibility runs counter to the spirit of liberalism, and excluding it is one of the goals of Joseph Raz's alternative autonomy-based argument in "Autonomy, toleration, and the harm principle." Raz's perfectionism, however, prevents him from giving autonomy any independent value, and his argument remains vulnerable to a commitment to extensive state control, if only this control can be achieved in a sufficiently efficient and subtle way. After discussing the problems in Mill's argument, I will elicit this "efficiency problem," as I shall call it, from Raz's argument. Raz briefly indicates in a few places how he might respond to this problem, but none of these indications is sufficient to diffuse it. I will then suggest that a Razian perfectionism does contain the resources to deal with the problem.

Utility, Autonomy, and the Harm Principle

John Stuart Mill presents his harm principle in On Liberty as a strict limitation on state interference in the lives and actions of individuals. The state, he claims, may only properly interfere with the actions of an individual in order to prevent harm to others. The utilitarian argument with which he supports the principle, however, gives the term "harm" an alarmingly wide range of meanings. The argument leaves open the possibility of advocating extensive state control over individual life while still staying true to Mill's liberal political philosophy. This possibility runs counter to the spirit of liberalism, and excluding it is one of the goals of Joseph Raz's autonomy-based argument in "Autonomy, toleration, and the harm principle." Raz's perfectionism, however, prevents him from giving autonomy any independent value, and his argument remains consistent with a commitment to extensive state control, if only this control can be achieved in a sufficiently efficient and subtle way. After discussing the problems in Mill's argument, I will elicit this "efficiency problem," as I shall call it, from Raz's argument. Raz briefly indicates in a few places how he might respond to this problem, but none of these indications is sufficient to diffuse it. I will then suggest that a Razian perfectionism does contain the resources to deal with the problem.

I

Mill's harm principle does not allow for state interference whenever an individual's action will result in harm in the ordinary sense. His rule-utilitarianism demands that the action in question belong to that class of actions that generally tends to cause harm. The harm must, moreover, be felt by someone other than the agent. Mill defines this class of actions in two ways. He first tells us that a harm is an action that

"violate[s] a distinct and assignable obligation to any other person or persons" (Mill, 148). An obligation of this special sort has been violated whenever an action leads to "a definite damage, or a definite risk of damage, either to an individual or to the public" (Mill, 149). The term "definite damage" requires explication. Whenever someone acts, the consequences of the action fall into one of five possible categories: utility-maximizing, imperfect utility-increasing, neutral, imperfect utility-decreasing, or utility-minimizing. Mill believes that one has the right to waste one's own life, and thus the Millian definition of "definite damage" must fall into either the fourth or the fifth category (usually the fourth, given the rarity of someone's doing an action which, given the circumstances, causes the most pain that could possibly be caused). Actions that fail to produce any utility, or produce very little, do not count as harm, even when these actions fail to prevent easily avoidable disutility to other people.

The main threat to a utilitarian defense of the harm principle is the suggestion that an individual might be coerced into living a productive form of life. The initial coercion would then result in both the individual's eventual happiness and whatever happiness the individual brings to society. And if the individual were to truly embrace his new life, he would be just as productive as someone who had freely chosen it. From a utilitarian standpoint, then, the coercion would be unimpeachable. In his argument from the significance of choice, Mill tries to exclude this possibility. He insists that enthusiasm and choice are connected, claiming that we only exercise our faculties when we choose freely (Mill, 122). A project can only be pursued energetically if it engages our faculties, because "if the inducements to an act are not such as are consentaneous to his own feelings and character…it is so much done towards rendering his feelings and character

inert and torpid instead of active and energetic." (Mill, 123) Thus, we can only put our energy into plans we have chosen. When one is coerced, one fails to deploy one's faculties "other than the ape-like one of imitation" (Mill, 123) in pursuing one's project. An activity that one has not chosen will necessarily fail to engage one's nature. For Mill, if a person lacks initial enthusiasm, then he cannot embrace the project later on. An individual human nature must develop "according to the tendency of the *inward forces* which make it a living thing" (Mill, 123, italics added). To develop as people, the projects we pursue must be set by our own internal natures. In asserting that "in proportion to the development of his individuality, each person becomes more valuable to himself, and is, therefore, capable of being more valuable to others," (Mill, 127) Mill shows that he does not believe that one can come to value a project that one has not chosen. Since a coerced project does not help to develop one's individuality, one cannot value one's contribution to it. If the project is not valued, we cannot expect it to be pursued enthusiastically. This increases the likelihood that the cost of implementing it will outweigh the increase in utility gained from pursuing the coerced project.

Mill's problematic psychological assumptions undermine his argument. He must assume that the activity could not eventually come to spark the interest of the one coerced, leading to a willful acceptance of the project. There is no reason to assume this kind of obstinacy in human psychology. Mill thereby fails to block the state's justification for interfering in purely personal behavior in some cases. The principle of utility not only justifies state interference in some (albeit specific and limited) kinds of personal behavior, it also allows for a very broad definition of what constitutes a harm to another person. His definition of harm, moreover, yields a wide range of acceptable

cases of state interference. To say that no assignable obligation has been violated, or no definite damage done, is to say that the action of an individual "neither violates any specific duty to the public nor occasions perceptible hurt to any assignable individual except himself," (Mill, 149) nor does he pose any clear risk of having one of these consequences. We must add this point about risk, since Mill has already identified the risk of definite damage as a condition that legitimates state interference. But there are many actions that pose a risk of causing perceptible hurt to one or more individuals. Mill cites gambling, drunkenness, idleness and uncleanliness as acts that fall *outside* the realm of what should be prohibited by law (Mill, 147). But drunkenness brings with it the possibility of injury to other people, and excessive gambling by someone with a family to support constitutes a violation of an assignable obligation. Severe uncleanliness in public can contribute to the spread of disease. Even idleness can cause perceptible harm to the public, if the state has any kind of welfare system. Mill makes explicit that he views certain causes of taxation and state expenditure as perceptible harms to the public. He asserts that if one has a child, one should be compelled by the state, as far as is possible, to feed, clothe and educate her at one's own expense. Anything less is a crime against society as well as the child (Mill, 176).

Mill's theory shrinks the liberty to plan one's life, and implies that coercion is acceptable even in some cases that he explicitly wants his theory to place outside the realm of state authority by allowing the state to interfere based on the risk of harm, and counting the burden of taxation as a harm to the public when it results from actions that could have been avoided. One might object to my extension of Mill's remarks on the duties of parents into the cases of the idler, drunk, and gambler. The parent has a duty to

another person, and if he or she fails, the state has the right to intervene. But in the other cases, although there is the risk of perceptible harm to others, only duties to self are being violated. This objection misses the point. For Mill, the duties we have to ourselves are the duty to preserve our dignity and the duty to develop our natural abilities (Mill, 145). He does claim that his harm principle prevents the state from forcing anyone to perform these duties. I have not, however, been arguing that the harm principle permits coercing the idler, private drunk, and gambler on account of their lack of dignity or unfulfilled potential. If we suppose that they are independently wealthy, and thus do not need societal support, there is no way to advocate coercing them based on the risk they pose to others. But once they become the source of financial harm to the public, the fact that this harm results from a violation of purely personal duties is irrelevant. For Mill's principle, only the resulting public harm matters. As soon as the risk of "perceptible hurt" is present, interference is justified.

II

Joseph Raz refashions the argument for the harm principle, aiming to avoid the difficulties inherent in Mill's defense of the liberty to plan one's own life. Rather than argue from the principle of utility, Raz begins with the reality of competitive moral pluralism and the value of autonomy. After examining how he develops this argument, I will comment on the ways in which it improves on Mill, and then discuss the problems that remain for it.

To be autonomous is to be part author of one's life. This requires that "many morally acceptable, though incompatible, forms of life be available to a person" (Raz 1988, 158). The available forms of life must be incompatible, because being part author

of one's life requires having a range of choices concerning the kind of life one will lead. Without incompatible options, there is no need to make a choice about one's life, and one cannot be an author of one's life if one never has an opportunity to choose one way of life over another. Autonomy is more than a lack of coercion. It is possible for someone to have only one form of life available to him and, as luck would have it, to be enthusiastic about leading that kind of life. Such a person has not been coerced, but he is not autonomous. Everything has been set for him whether he likes it or not. Raz stresses that autonomy requires an *adequate* range of options, but does not require any one option. The key is that some of the forms of life available to an individual be lives he actually wants to lead. But one can be autonomous even if the way of life one wants most of all is not available.

The range of available lives must be morally acceptable as well as adequately diverse. A range of options that does not include any morally acceptable lives is obviously inadequate, since many people will fail to find in such a range any life that they want to lead. The real point of this moral requirement is in what it excludes. The presence of morally repugnant options is not a requirement of autonomy (Raz 1988, 169). Raz acknowledges that morally good individuals are "able to cope with the temptations and pressures normal in their society," and thus that "vices and moral weaknesses are logically inseparable from the conditions of a human life that can have any moral merit" (Raz 1988, 168). He does not, however, think that these problems can be eliminated, and so respect for autonomy does not require that they be available. We shall return to this point when we examine the difficulties that Raz's argument faces.

Raz does not believe that autonomy is valuable in itself. An autonomous pursuit of evil or vice is more worthy of blame than a similarly evil but coerced pursuit. The value of autonomy lies in choosing to pursue a good form of life (Raz 1988, 169). A good life is a life of self-realization, a life that develops some of one's valuable capacities. Autonomy is not required for self-realization, and being autonomous means being able to choose or reject a life of self-realization. Autonomy itself has no value if the wrong choice is made. An autonomous life of self-realization, however, is better than a non-autonomous self-realizing life. There must be many different, incompatible ways of self-realizing because the condition of autonomy requires that incompatible forms of life be available. We cannot develop all of our capacities fully, so being forced to develop some instead of others must be both morally acceptable and consistent with autonomy. This requirement connects autonomy and moral pluralism. A morally acceptable form of life exhibits some virtues to a high degree and others to a lesser degree. For there to be incompatible morally acceptable forms of life, there must be incompatible virtues. Choosing to lead a life that exhibits one set of virtues to a high degree means giving up a life that exhibits others virtues. Moral pluralism is the view that there are several incompatible forms of life, each of which is equally worth choosing. The different virtues exhibited by these forms of life provide the competing reasons for choosing them. Autonomy requires that available forms of life be *equally* morally acceptable, because if one could rank-order one's options then choosing any but the best option would be defective in some way. An autonomous person can make one of several choices, none of which is worse overall than the others. Otherwise, there would be a

sense in which the situation made the choice for one, provided that one was committed to leading the most morally valuable life one could.

The link between respect for autonomy and toleration is made through *competitive* moral pluralism. This view "admits the validity not only of distinct and incompatible moral virtues, but also virtues which tend, given human nature, to encourage intolerance of other virtues" (Raz 1988, 164). Raz claims that in addition to tolerating behavior we view as bad, we can also tolerate people's limitations, when these limitations result from the choice to develop certain virtues at the expense of others. If I have chosen a life that allows me to cultivate the virtues of decisiveness and expedience in action, I will probably be tempted toward intolerance of those whose lives have cultivated the virtues of cooperation and careful deliberation. My awareness of the different but equal value of this other form of life, which is incompatible with mine, gives me a reason to tolerate the other person's limitations, and to expect that my own limitations will be tolerated.

If we respect autonomy as one component of a good life, we must value moral pluralism. The diversity of human virtues makes it likely that this pluralism will be competitive. Our respect for autonomy thus provides a reason to value toleration. The final step in Raz's argument is to show that the appropriate principle for limiting the scope of autonomy-based toleration is the harm principle. Raz distinguishes his version of the harm principle from Mill's defining it as one which "regards the prevention of harm to anyone ([the agent] included) as the only justifiable ground for interference with a person" (Raz 1988, 169). Raz gives both narrow and broad senses to the notion of harm. In the narrow sense, someone is harmed when his prospects are limited or his

efforts toward a project he has already begun are frustrated (Raz 1988, 169). In the broad sense, harm includes physical injury that is not incapacitating, and reasonably endurable offense (Raz 1988, 170).

The argument from autonomy is an argument for the narrow understanding of the harm principle. Harm in the narrow sense just is restriction on autonomy. This is one great improvement over Mill's argument from utility. Many of the acts that count as harms to the public based on Mill's argument will not so count on Raz's, because they do not meet the more stringent requirement of restricting the autonomy of others. The argument from autonomy thus preserves more of the spirit of liberalism that the harm principle is supposed to embody. Raz's argument also justifies a broader system of welfare than Mill allowed. Raz identifies three duties essential to promoting autonomy (Raz 1988, 166). The first is the negative duty to refrain from coercing others. The other two are positive duties: we must help others cultivate the capacities that a good life requires, and help make an adequate range of good lives available to others. Basing the harm principle on respect for autonomy means "establish[ing] that the autonomy-based duties never justify coercion where there was no harm" (Raz 1988, 171). Raz considers the objection that the existence of positive autonomy-based duties requires the state to coerce its citizens into contributing to the autonomy of others (most likely through compulsory taxation), and that the failure of someone to make such a contribution does not constitute a harm. His reply is that this objection relies on an unjustifiably narrow definition of harm, and he asserts that "one can harm another by denying him what is due to him" (Raz 1988, 171). Because autonomy is valuable when it is part of a life of self-realization, we have a duty to promote it by making acceptable options available to

others. If we fail in this, other people will have fewer prospects than if we had succeeded—fewer prospects than they *should* have had, since the value of autonomy makes promoting it a *duty*. The central meaning of harm is limiting someone's prospects. We harm others when we fail to fulfill our autonomy-based duties, because our actions result in some people having an insufficient range of valuable prospects, and thus in those people being denied the possibility of good lives. The existence of state-enforced autonomy-based duties, therefore, does not contradict the harm principle.

III

I will accept Raz's defense of the claim that the harm principle and the duties of autonomy are compatible. I want to focus on whether his conception of autonomy is adequate to exclude any justification for coercion. Two of Raz's claims about autonomy give us *prima facie* reason for thinking that his conception is defective: his claim that respect for autonomy does not require preserving any morally repugnant options, and his claim that autonomy is not valuable in itself, but only as part of a life of self-realization. Suppose someone has chosen to reject a life of self-realization, but without doing anything that significantly limits or frustrates the autonomy of others—we may use the examples from Mill of the idler, drunk, or gambler. It would be consistent with a general respect for autonomy to eliminate the circumstances that allow for these worthless choices, so long as we eliminate them in the right way. If we completely eliminate alcohol, or gambling, or try to enforce a universal prohibition on idleness of any kind, we will likely violate the valuable autonomy of many people who are leading lives of self-realization. Why not, then, make it public policy to coerce these non-realizing individuals into leading different lives, lives that include the pursuit of some worthy

project? Doing so would allow us to continue respecting the autonomy of those who make moderate use of alcohol, or occasionally participate in gaming or recreation, as part of a good life (and these activities certainly can be part of one.) Raz claims that "a moral theory which values autonomy highly can justify restricting the autonomy of one person for the sake of the autonomy of others or even of himself in the future…But it will not tolerate coercion for other reasons" (Raz 1988, 173-4). But the "high value" Raz's theory gives to autonomy is strictly conditional: if one pursues a life of self-realization, then one's autonomy is very valuable. That kind of life is much more valuable when it is autonomous than when it is not. But the cases we are considering are cases of *wasted* lives. The autonomy of these individuals is not valuable. Raz must give some reason for us to respect it beyond the mere fact of its being autonomy.

Raz has two responses to this problem. One appeals to epistemic utility. He asserts that the state's considering something to be valuable or valueless is no reason for anything. Only its being valuable or valueless is a reason. If it is likely that the government will not judge such matters correctly then it has no authority to judge them at all (Raz 1986, 412). The problem with taking this claim as a response to the present problem is that in the cases we are considering, it is implausible that the judgment that these lives are wasted is wrong. This was the same problem Mill had in making a similar response. For these lives to be valuable, there must be some valuable capacity which they cultivate. But clearly there is not. It is a fact that we can waste our lives, and the harm principle is not supported by any position that cannot produce an argument against coercing those who choose to waste their lives.

The second response focuses on the efficiency of coercion. Raz claims that "forms of coercion…all invade autonomy, and they all, at least in this world, do it in a fairly indiscriminate way. That is, there is no practical way of ensuring that the coercion will restrict the victims' choice of repugnant options but will not interfere with their other choices" (Raz 1988, 173). If we coerce the drunk to change his way of life by locking him away, we have done nothing to improve matters. Imprisoning him will do just as good a job, if not better, of preventing him from leading a life of self-realization as he was previously doing on his own. We now have some reason for thinking that the costs of implementing coercion in this kind of case are unjustifiably high. Let us even assume that at present all our practical means of coercion suffer from this inefficiency. It is still possible that in the future we will develop more sophisticated and subtle means. If we do, how can Raz's argument justify withholding those means from individuals who have chosen to waste their lives?

This objection is what I referred to above as the "efficiency problem" for Raz's argument. He does consider and respond to this objection, but his response is far from adequate. He asks "what if it became possible to coerce people to avoid immoral but harmless conduct without limiting them in any other way," and then insists that "it is an advantage of my argument that it does depend on contingent features of our world…I do share the reluctance of supporters of the harm principle to say that in the imagined circumstances the enforcement of harmless immorality is justified," however, "it is impossible for us to say how the change would affect the merits of the issue" (Raz 1986, 419). One of the relevant contingent features of our world is that we do make progress in

developing subtler, more efficient, more focused means of coercively affecting behavior[1]. It does not require such a disorienting stretch of the imagination to consider a world in which the possibility Raz considers is realized. If we believe in the harm principle we believe that it should guide the actions of the state even in circumstances such as this. Otherwise, it is no more than a practical requirement of the moment, which should be abandoned as soon as we have the means to implement a superior principle. Raz's admission of reluctance shows that he leans toward the former position. To support the harm principle from a Razian framework, then, we must find a way to defend it even in this hypothetical situation.

IV

I shall now briefly sketch some reasons for thinking that a basically Razian view does have the resources to cope with this objection. The first thing to note is what we cannot do. If we stick to the claims that undesirable options are not required by respect for autonomy, and that autonomy is only conditionally valuable, then we cannot exclude the possibility that coercion of the harmlessly immoral will be justified in some cases. But this is more than a Razian view requires. We should, rather, argue that although coercion could be justified in these cases, it would not be the best option. There are superior reasons favoring non-coercively promoting lives of self-realization[2]. The ideal situation is one in which everyone autonomously pursues a life of self-realization. This is a useful political goal, even if it is impossible to attain. We must decide what to do about those who have chosen to waste their lives, and we have two options. We may use our

---

[1] See for example the work of social psychologist Richard M. Perloff, especially The Dynamics of Persuasion: Communication and Attitudes in the 21st Century, Hillsdale, NJ: L. Erlbaum Associates, 2002.

[2] I will not give any detailed consideration to the kinds of non-coercive promotion the state might use. They might include aspects of the public education system (including opportunities for adult continuing education, tax incentives, etc.

sophisticated means of coercion on them, or we may use sophisticated (we may assume) means of non-coercive promotion[3]. The right question to ask is: at what point does coercion become preferable? Coercion has the advantage of certainty. If we coerce someone into pursuing a valuable project, and we are good at coercion, we know that he will in fact pursue the project. But an uncertainty attaches to it as well, the uncertainty that Mill noticed, but exaggerated. We do not know if the person will ever come to embrace the project. He may, but we do not know for sure. Non-coercive promotion is not certain to succeed, but we know that if it does, the project will be genuinely embraced. The result of promotion is an autonomous choice.

Thus far, the cases for each option may seem comparable—there is some certainty and some uncertainty attached to each side. We know, however, that the end result of successful promotion is better than that of coercion, because an autonomous life of self-realization is better than a non-autonomous one. We should, therefore, at least begin with promotion in each instance. The next question is, at what point are we justified in switching to coercion? We must remember that the value of autonomy, though conditional, is still very high. Resorting to coercion means preventing someone from leading an autonomous life of self-realization, the best kind of life she could lead. Here, another kind of uncertainty becomes important. For any given case, we have no way of knowing how long it will take someone to respond to the state's promotion of good lives. There can be no point at which we can say definitively that *now* is the time to switch to coercion, because this particular person is a hopeless case. If we really do believe that

[3] Since the hypothetical allows for the existence of improved means of coercion, it seems only fair that in evaluating the situation we posit the existence of sophisticated means of promotion as well.

the value autonomy adds to a life of self-realization is great, we have reason for never deciding to resort to coercion.

Mill's utility-based argument for the harm principle fails to justify the ban on state interference in the case of wasted lives, and gives an alarmingly broad meaning to the term "harm." Raz's autonomy-based argument narrows the meaning of harm, and thus avoids many of the problems of Mill's view. This argument as Raz makes it still leaves open the possibility of justifiably coercing those who have wasted their lives, albeit only within a counterfactual situation. A Razian view, however, does contain the resources to respond to this objection. A fully developed version of this response would contribute much to the project of liberal perfectionism.

Bibliography

Mill, John Stuart. On Liberty. Middlesex, England: Penguin Books Ltd., 1978.

Raz, Joseph. "Autonomy, toleration, and the harm principle." In Susan Mendus, ed. Justifying Toleration: Conceptual and Historical Perspectives. Cambridge: CUP, 1988.

Raz, Joseph. The Morality of Freedom. New York: Oxford UP, 1986.