

Copyright Notice

Staff and students of this University are reminded that copyright subsists in this extract and the work from which it was taken. This Digital Copy has been made under the terms of a CLA Licence which allows you to:

- Access and download a copy
- Print out a copy

This Digital Copy and any digital or printed copy supplied to or made by you under the terms of this Licence are for use in connection with this module or course of study. You may retain such copies after the end of the course, but strictly for your own personal use.

All copies (including electronic copies) shall include this Copyright Notice and shall be destroyed and/or deleted if and when required by the University.

Except as provided for by copyright law, no further copying, storage or distribution (including by e-mail) is permitted without the consent of the copyright holder.

The author (which term includes artists and other visual creators) has moral rights in the work and neither staff nor students may cause, or permit, the distortion, mutilation or other modification of the work, or any other derogatory treatment of it, which would be prejudicial to the honour or reputation of the author.

Title of extract: **Is it Wrong to Prevent the Existence of Future Generations?**

Name of Author: **Sikora, R.I.**

Title/Author of book/edition (if applicable): **Obligations to Future Generations.**

Sikora, R. I. and Barry, B. (editors)

Name of Visual Creator or other acknowledgement (if applicable):

Journal details –title/Vol/Year (if applicable):

Publisher: **White Horse Press**

Pages scanned: **112 - 166**

Course of Study (Module name and number): **Environmental Philosophy CSPH7010**

Name of Designated Person
authorising scanning:

E. L. Harris

Date authorised: 20.01.14.

Moodle/Powerpoint

Is It Wrong to Prevent the Existence of Future Generations?

R. I. SIKORA

Most people would probably agree that anyone deciding whether or not to do something that could conceivably lead to the destruction of mankind—for example, whether or not to start a revolution that might escalate into a nuclear war—should concern himself not only with the possible consequences of his action on those currently alive but should also consider the fact that his action might prevent the existence of future generations. It might seem that for practical purposes it doesn't matter whether or not it is wrong to prevent the existence of future generations because in order to prevent their existence you would have to do something drastic to everyone alive such as making them all sterile or destroying them all, and *that* in itself, aside from any other consequences, would certainly be wrong. But there could be a good reason for doing something that involved a *slight* risk of causing something terrible (such as, for example, the death of everyone alive); and a level of risk that would be tolerable if that was *all* that one should worry about might *not* be tolerable if the possible prevention of future generations had to be taken into consideration as well.

Though it seems eminently reasonable to hold that we have an obligation to refrain from preventing the existence of future generations (provided they have good prospects for happiness), it is a claim that some able philosophers reject. The obligation not to prevent the existence of future generations is supported by what is perhaps a more fundamental obligation—that it is *prima facie* wrong to prevent the existence of anyone with reasonable prospects for happiness. I am not

claiming that this obligation is always overriding: it is in fact commonly overridden by other obligations. But that we have such a *prima facie* obligation relating to possible future persons is the basic thesis I want to establish in this paper. I will call my theory (or any theory that holds that it is *prima facie* wrong to prevent the existence of a person with reasonable prospects for happiness) an "obligation theory," and any theory which denies this, an "anti-obligation theory."

Moral systems have commonly considered only our obligations to those who are currently existing. Once future generations are taken into consideration, new moral issues arise which are not just a problem for utilitarians (though they have generally been brought up in the context of some sort of utilitarian system); they are problems for any ethical system which purports to be complete. Because I am a kind of utilitarian I consider the issues within a utilitarian framework, but analogues to the kinds of utilitarianism I discuss could be given for other moral systems. For instance, there could be semi-consequentialist analogues in the style of W. D. Ross or Aristotelian analogues where the questions would be, not what it is right or wrong to do, but what it would be virtuous or vicious to do.

What I take to be the two main objections to saying that we have an obligation not to prevent the existence of future persons is that this obligation relates to someone whose existence is not actual but only *possible* and who may in fact never exist; and that it commits us to what Derek Parfit calls the repugnant conclusion: that if an extremely populous world with a happiness-average barely over neutrality would have a greater happiness-total than a much less populous world with a far higher happiness-average, we would be forced to choose the first. In its more plausible form, the obligation theory cannot escape Parfit's objection; but I will attempt to blunt it by arguing: (a) that the most plausible anti-obligation theories face the same problem; (b) that the possible-persons objection is incompatible with Parfit's objection because any theory that implies that we never have an obligation to create happy people even at the cost of some sacrifice to those alive—because their existence is only possible—could conceivably be confronted with a situation in which the low-average alternative world in the repugnant conclusion could only be avoided by the creation of those happy people; (c) that this more populous world with a low happiness level might in fact be quite similar to the actual world we live in; and (d)

that no currently defended version of an obligation theory would require us in actual practice to make the repugnant choice. I spend what may seem a disproportionate amount of space on the repugnant conclusion because it strikes some people, including, e.g., Peter Singer, as the most formidable objection to an obligation theory.¹

The main body of my paper is not about the repugnant conclusion, however, but attempts to make the following points:

(1) that timeless utilitarianism (the view that the utilitarian principle applies to everyone who exists in the timeless sense) is incomplete on the question of our obligations to future generations because it is incompatible with both obligation theories and anti-obligation theories;

(2) that the common view that it is at least permissible, taken in itself, to bring into existence a large group of people almost all of whom would be happy even if one of them would be utterly wretched can be justified only if it is admitted that it is a good thing to bring about or not to prevent the existence of happy people;

(3) that classical utilitarianism can be defended by a new argument which again relies on the intuition that we can only justify allowing groups of people some of whom will be wretched being brought into the world if it is admitted that it is a good thing to bring happy people into the world;

(4) that those who take the view that the preservation of mankind is not a moral but an aesthetic matter—that it would be a shame not to have people in the future mainly for the sort of reason that it would be a shame not to have elephants, giraffes, and other interesting and complicated forms of life, or for some other non-moral reason—are mistaken. It is, I argue, fundamentally a moral matter; though if one accepts my *meta-meta-ethical* theory, it doesn't really matter whether it is a moral question or not as long as the reasons that can be given for the preservation of mankind can prove persuasive to an epistemically rational person.

1. Anti-obligation Theories & the Repugnant Conclusion

It has commonly been thought that average utilitarianism can avoid both the possible-persons objection and the repugnant conclusion objection. Average utilitarianism, which seeks to maximize the happiness average rather than the happiness total, differs from classical

utilitarianism only on matters affecting population size: it would direct us, for example, to aim at a smaller population with a lower happiness total if this would result in a higher happiness average. Average utilitarianism avoids the repugnant conclusion because it allows the average to be raised by adding people whose existence at the time the decision is made is merely possible which, of course, makes it vulnerable to the possible-persons objection (though in actual fact this isn't a real problem).

Taking the problem of the repugnant conclusion first, consider the following case. There has been a nuclear war. Only a million people are left in the world all of whom, because of radiation, are sterile. Once they die, unless something extraordinary occurs, no one will exist. Because of the effect of radiation on their health, the death of most of their families and friends, and the tremendous destruction in the physical world, their prospects for happiness are very low. However, if they want to they can repopulate the world with test-tube babies whom previous experimentation has shown will have a high level of happiness. Producing and rearing these children will unavoidably require a good deal of time and effort from them which will lower their own level of happiness still further. There are then two possible worlds: one in which the average happiness level would be very low and the other, the world of the test-tube babies which their sacrifice could bring about, in which the happiness level would be very high.

Any theory that takes it to be wrong to require those alive to make sacrifices to bring about the existence of possible persons must choose the first of these alternatives, the world with a very low happiness level: that is, it must embrace the repugnant conclusion. Average utilitarianism, however, can choose the second alternative, the happier world: since its sole objective is to maximize the happiness average, it will tell us to choose the happier world even if it requires a sacrifice from actual persons for possible persons.² Average utilitarianism only avoids the repugnant conclusion because it allows that there can be obligations concerning possible persons. *Any theory which denies that there can be an obligation to add possible people is open to the repugnant conclusion objection so that in fact no theory can avoid both the repugnant conclusion and the possible persons objection.*

As I see it, the main problem for average utilitarianism is this. Suppose you could cause someone to exist (that is, you could "add" an

"extra person") whose net effect on others would not be harmful and who would enjoy a happy life though his happiness would be below the average. An average utilitarian has to assert not merely that we have no obligation to add this extra person but that it would be *wrong* to add him, and that it would be wrong no matter how happy he would be so long as he would not be as happy as an average person in that world. And it would be wrong to add him even if, besides being happy himself, he would have a beneficial net effect on others so long as that effect was not as great as the difference between his own happiness level and that of other people because if it were not, he would bring down the average. This can go further: it would be wrong to add him even if adding him satisfied the conditions of Pareto optimality, that is, even if each of the other people whose condition he affected was the better for it. In fact, it could even be wrong to add a happy person or group of persons who would satisfy what might be called the "Pareto-plus optimality requirement" by having a beneficial effect on *everyone* already there so long as doing so would lower the overall average. To take a simple case, suppose you have a world with a population of two people each with a happiness level of ten, and that you add another person with a happiness level of five who through his actions would make each of the others happier so that each of their happiness levels would go up to eleven. The Pareto-plus optimality principle would tell you to add him, but an average utilitarian would have to say that it would be wrong to add him: you would be reducing the average from ten to nine (i.e., 27/3). Sometimes showing that an ethical view has counter-intuitive consequences is not totally decisive because our initial intuitions may prove to have been mistaken, but this is surely not the case with the Pareto-plus optimality objection. It is enough by itself to demolish average utilitarianism.

2. *Is the Repugnant Conclusion Really Intolerable?*

Since any theory which denies that we can have an obligation to future generations (as well as the most plausible theories which do not deny this) entails the repugnant conclusion, it behooves me to see whether it is not something we can live with after all. It strikes me that the world we actually live in may be no better than the low-average alternative in the repugnant conclusion. All that would have to obtain

for this to be the case is that the average person in the world should not have *significantly* more happiness than unhappiness, and it is certainly arguable that this is so. Furthermore, although one tends to think of the low-average alternative world as one in which everyone's happiness level is always just barely over neutrality, it doesn't *have* to be like that to get the required average; and even if it were, it wouldn't be as bad as it seems for if no one experienced more than mild pleasure, there wouldn't be much room for pain either; otherwise, the average wouldn't be above neutrality.³ And many people would, I think, take a life in which there wasn't much pleasure but in which there wasn't much pain either to be an acceptable sort of existence. In actual practice, one wouldn't expect a world in which the average person's pleasure barely exceeded his pain to be either a world of continuous mild pleasure, or a world of hedonically neutral experiences with just enough mild pleasure to yield a positive net balance. One would expect instead a world rather like ours in which there was, among other things, a lot of intense suffering and even more moderate suffering and unpleasantness; and in order to balance this painful experience and come out with a slightly positive hedonic balance there would have to be some intensely pleasant experiences—joyful, even ecstatic experiences—or, lacking that, a great many moderately pleasant experiences.

If hedonistic utilitarianism is true, then in order for the low-utility-average alternative of the repugnant conclusion to contain more utility than disutility, that world must contain at least somewhat more pleasure than pain. Such a world may be no worse off hedonically than the world we actually live in. But if ideal utilitarianism or an interest theory is true, a world with considerably more pain than pleasure could still contain more utility than disutility because intrinsic value and disvalue in those systems is not solely a matter of pleasure and pain. Consequently, an ideal utilitarian or a philosopher who holds an interest theory could assert that the highly-populated low-utility-average alternative world of the repugnant conclusion although it contains more utility than disutility might nevertheless contain considerably more pain than pleasure, and considerably more pain than pleasure than is the case in the actual world. Though I would like to claim that the low-average alternative of the repugnant conclusion would not be worse than the actual world, showing that this is so is not essential to my case.

3. *The Repugnant Choice in the Actual World*

While both obligation and anti-obligation theories are open to the repugnant conclusion in theory, it might be claimed that while this is only a logical possibility for an anti-obligation view, there is a real chance that any obligation theory will be forced to it. In order to consider the merits of this contention, I need to look at some particular obligation theories. Classical utilitarianism requires actual people to make sacrifices to add extra people whenever doing so will maximize happiness (or whatever is taken to be the basically valuable thing). If one didn't want to go this far—and this is, I think, the commonsensical stance—one might take the position that while the happiness of extra people should be given *some* weight, it should not be given equal weight with the happiness of actual people. This is what I will label compromise principle 1, since another way of looking at it is as a compromise between a theory attributing no value to adding happy people and classical utilitarianism.⁴ The question is, given the world we live in, is it likely that either theory would force us to add great numbers of people till we have a large drab world instead of a much smaller, much happier world? Consider the sorts of effect that adding population could have on the average and the total happiness in the world: it could (1) increase the total without lowering the average; (2) increase the total but lower the average; or (3) decrease both the total and the average. It is only in the second case that there is any danger of the repugnant conclusion, but the world as a whole is now obviously so crowded that adding extra people is likely to lower both the total and the average happiness.⁴

If it is indeed unreasonable to think that the world's happiness will be increased by adding extra people, it is unlikely that the moral beliefs of a classical utilitarian or someone holding compromise principle 1 will force either of them to make the repugnant choice. Actually, there is a reason why it is even more unlikely. To be forced to the repugnant conclusion, you must not merely advocate a policy that will decrease the average in order to increase the total but *the decrease must be from a high average to an average barely above neutrality*. If there is to be such a decrease, the original average would have to be many times as great as the new average for, to be high, the average must surely be many times as great as an average that is barely above neutrality. Thus

the new population would have to be many times as large as the old one. Consequently, in order for you to be confronted with the repugnant choice it must not only be possible for you to raise the total by increasing the population: *you must be able to raise the total by increasing the population to many times its current size*. But that is clearly out of the question for the world as a whole and if it is attempted in some part of the world with a relatively small population such as Canada, there is the further fact that both space and natural resources desperately needed elsewhere would be used up, not to speak of the enormous increase there would be in world pollution problems.

Jonathan Bennett has suggested that since Parfit's objection runs into these difficulties, we should look for some other form of the objection which does not. It might be urged that it would not only be repugnant to choose a world with an average barely above neutrality over a world with a high average, but it would also be repugnant (though perhaps less so) to choose a world with a lower happiness average than some other possible world even if the disparity isn't as great as that in Parfit's example. But this is average utilitarianism again—which is, of course, open to the Pareto-plus optimality objection. In order to avoid it, the principle could be modified to: it is always wrong to increase the total by lowering the average except when doing so would not be detrimental to the interests of some privileged group (such as those currently alive). The repugnance would then consist not in choosing a world with a lower happiness average than some other possible world, but in lowering the average in order to increase the total happiness by adding extra people at the expense (however little that might be) of those in that privileged group. A crucial assumption underlying this seems to be that there is no, or only marginal, moral worth in adding happy people to the world. But that is just what the objection seeks to establish so it suffers from circularity. Still this doesn't show that the intuition is mistaken. I will try to show that in §§ 8, 9, & 10.

4. *No Plausible Ontological Preference Theory Can Avoid the Repugnant Conclusion*

I have argued that a strict anti-obligation principle is no better off with respect to the repugnant conclusion than the most plausible obligation theories: they are all open to it in theory but not in practice.

At this point I ask whether it is possible to find a plausible moral theory for my opponents which is in the main against adding extra people if this would require any sort of sacrifice from what might be called the ontologically privileged, but which nonetheless avoids the repugnant conclusion even in theory. A reader who is satisfied that there is no such theory can skip this section. In order to cover the most promising alternatives, I mean to examine systematically the main theories based on different answers to the question, "How much weight ought to be given to the interest of the ontologically privileged, and how much to the interest of those who are not members of that group?" (By the ontologically privileged, I mean that class of people whose ontological status is thought to make the promotion of their happiness of greater moral importance than that of "extra people," i.e., those not included in that group. The ontologically privileged might, for example, be thought of as the class of all actual persons.

1. *It is always obligatory to add happy extra people when the ratio of happiness gained through adding them compared to the happiness loss of the ontologically privileged is satisfactory.* Obviously, a ratio of less than 1/1 is unsatisfactory: no one supposes that the happiness of extra people should count more than that of the ontologically privileged.
 - 1.1. *Any ratio higher than 1/1 is satisfactory.* (the position taken by classical utilitarianism)
 - 1.2. *Only some ratios over 1/1 are satisfactory.* (compromise position 1)
2. *There is no ratio of gained happiness for added people versus loss to the ontologically privileged such that whenever that ratio is attainable it is obligatory to add happy extra persons.*
 - 2.1. *It is never obligatory to add happy extra people but it is not wrong to add them unless their addition would be detrimental to the interests of the ontologically privileged.* (That it is not wrong to add happy extra people whose addition will not be detrimental to the interest of the ontologically privileged is to be taken as included in all of the following views.) This avoids the main objection to average utilitarianism: it permits us to add extra people who will be happy if adding them is not detrimental to the interests of the privileged group. However, it entails

the repugnant conclusion because it prohibits raising the overall average at the expense of the ontologically privileged.

- 2.2. *It is sometimes obligatory to add happy extra people but only when it is compatible with the maximization of happiness for the ontologically privileged.* This clearly entails the repugnant conclusion. In addition, it is what Rawls calls a lexicographic principle (or, for short, a lexical principle), i.e., a principle which treats two distinct sorts of consideration as morally significant but gives one strict precedence over the other. When two sorts of consideration rather than just one appear to be morally relevant, lexical principles embodying the two considerations seem to have a decided advantage over principles which allow room for only one of them; but the problem is that if two distinct things are really morally significant, it seems unlikely that one of them should always take complete precedence over the other. For example, it would seem that if we have any obligation at all to add happy people, the obligation, say, to visit your elderly ailing aunt should not take precedence over your obligation to add a very large group of very happy people. A second problem is that in actual practice the second-rank obligation is usually given no real role. In the present case, for example, the obligation to add happy extra people is only meant to be a tie-breaker when two alternate policies would work out equally well for the ontologically privileged. It seems unlikely that there would ever be any such ties to break.
- 2.3. *It is sometimes obligatory to add happy extra people but only when doing so involves no net loss to the ontologically privileged or is needed to avoid the repugnant conclusion.* This is 2.2. modified by the addition of the boldface clause. It is an ad hoc modification of a strict ontological preference theory, hence, the "ad hoc theory." Since it avoids the repugnant conclusion without involving the consequences of an average theory, it is worth serious attention. Unfortunately, it is open to stiff objections:
 - (a) In a world with a high happiness average there would be no obligation to add extra people no matter how happy they would be and no matter how small a sacrifice would be needed to add them. On the other hand, in the sort of

world with an average barely over zero envisaged by the repugnant conclusion, major sacrifices could be required, and these extra people for whose existence major sacrifices could be required might be far less happy than extra people for whose existence members of a happier state would be required to make no sacrifices whatever. But:

(i) This runs counter to our ordinary tendency to think that it is more appropriate to ask for sacrifices from those who are well-off than from those who are badly off; and

(ii) if we are required to make sacrifices to bring about the existence of happy extra people, it would seem that the happier those extra people would be, the greater the obligation would be: instead, the ad hoc theory calls for a sacrifice to bring about the existence of extra people with even a modest degree of happiness if they are needed to avoid the repugnant conclusion, and there is no obligation whatever to bring about the existence of even extravagantly happy extra people if they are not needed to avoid the repugnant conclusion.

- (b) It is not just that people in an unhappy world would have a greater obligation to make sacrifices to add happy people than those in a happy world; the obligation would be infinitely greater because those in a happy world have no obligation to make any such sacrifice at all. It is simply something they should do if no sacrifice is required.
- (c) According to the ad hoc principle, there is no obligation to make sacrifices for slightly less improvement than just enough to avoid the repugnant conclusion, and no obligation to bring about more improvement than is necessary to avoid it (even if such improvements could be achieved at the cost of proportionally far less sacrifice than those that the principle would require.)
- (d) The ad hoc principle is, like the last, a lexical principle.
- (e) The ad hoc view is an ontological preference theory (a theory that favors the ontologically privileged); but in abandoning the claim that we are *never* obligated to make sacrifices to bring about the existence of happy extra people, it gives up the main attraction of ontological preference theories.

2.4. *It is obligatory to add extra people whenever the ratio of gained happiness for those extra people versus the loss to the ontologically privileged exceeds some specified ratio higher than 1/1, except when doing so will lead to the repugnant conclusion.* Although this is a lexical principle in that avoidance of the repugnant conclusion is given strict priority over compromise principle 1, it is atypical in that it allows ample room for the application of the second-ranking consideration. In fact, if I am right, a situation will never arise in which the top-ranking consideration comes into play because compromise principle 1 will never in fact require us to make the repugnant choice. If so, this principle is extensionally equivalent to compromise principle 1.

Thus if anyone objects that compromise principle 1 leads to the repugnant conclusion in theory, this principle is an alternative that avoids the objection but is still for all practical purposes equivalent to compromise principle 1. However, I cannot accept it myself. I cannot because, if one is a utilitarian and if one admits (as the principle does) that there is some finite good in adding an extra person with even a slightly positive happiness balance, one will have to hold that it is infinitely *bad* to shift to the "repugnant" average, in order to justify the claim that it would be better to avoid the repugnant conclusion than to add even an infinite number of extra persons at a low but still positive happiness average. In fact, since the principle would allow one to accept a decrease in the overall average for a suitable increase in the total (so long as the average did not decline all the way to a "repugnant" average), infinite importance would have to be attached to the last bit of the decline, and that I can't accept.

2.5. *One should follow the classical principle except when it would lead to the repugnant conclusion.* Here again the original principle and its lexical counterpart designed to avoid the repugnant conclusion are extensionally equivalent. Again, I am bothered by its being a lexical principle although I am happy with it for all practical purposes.

I realize that I've given these seven possible theories in perhaps tiresomely complicated detail. Let me characterize them once more in a

brief way, which will, I think, show more clearly the relationship they have to each other:

1. The classical principle;
2. Compromise principle 1;
3. It is not wrong to add happy extra people when it is not detrimental to the interests of the ontologically privileged;
4. It is obligatory to add happy extra people unless it is detrimental to the interests of the ontological elite;
5. Principle 4. modified to avoid the repugnant conclusion;
6. Compromise principle 1 modified to avoid the repugnant conclusion;
7. Classical utilitarianism modified to avoid the repugnant conclusion.

5. All Ontological Preference Theories Are False

One of the first objections to any theory that holds that it may be obligatory in some cases for us to make a sacrifice for future generations is, put first in its crudest form, that it is never justifiable to require a sacrifice from an existing person for someone who does not exist. People who are currently existing constitute a kind of ontological elite whose interests are to take strict precedence over the interests of everyone else. The commonsensical root of this lies, perhaps, in something like this: suppose you are thinking of having another child. In weighing the pros and cons, you consider the interests of people who are currently living, the sacrifice that would be required, for instance, from yourself and from the other members of your family. You do not consider the interests of the unborn child one way or the other. Why should you? He or she is just a possibility.⁵ But the case is not as simple as that, it turns out. As soon as much thought is given to the matter, it is clear that the ontologically privileged class must be extended to include not only existing but inevitable people, though just what it is to be an inevitable person is not immediately clear. Some writers seem to think that the class of inevitable persons is the class of all those who exist in the timeless sense except for those who have existed in the past and those who exist now. But this won't do because it is also supposed that you may in fact bring about the existence of a person (thereby making him exist in the timeless sense) who does not belong to the class

of inevitable persons because you could have refrained from bringing about his existence. Furthermore, an inevitable person is not one who will exist no matter what *anyone* does: that would result in there being virtually no inevitable people. An inevitable person must instead be someone who will exist regardless of what a particular agent or group of agents (smaller than all mankind) does. Thus the notion of an inevitable person is always relative to some agent or agents. For example, a person whose existence is inevitable from my point of view may not be inevitable from his prospective parents' point of view: they could, for example, use a birth-control device or copulate at a different time. This means that there is no single class of inevitable persons. Therefore, if in framing our population policies the interests of inevitable people are to be set above the interests of merely possible people, different individuals should favor different population policies.

Peter Singer (following Derek Parfit) has quite a different objection to having the privileged class contain only actual and inevitable people. From the point of view of a population planner, it is inevitable that there be some minimum world population at a given time in the future—say, 1985—regardless of which of the policies open to him is put into effect.⁶ On the other hand, the number of *particular* individuals whose existence is inevitable from his point of view is much smaller than the minimum world population because his policy will affect the lives of most of the world's population either directly or indirectly, and even a minute effect on the time of impregnation will result in a different individual being born. Suppose that from his point of view, the smallest possible population for 2025 is five billion, but that only three billion of them will be inevitable from his point of view. Clearly, it would be wrong for him in framing population policy to be concerned only, or even primarily, with those three billion so that the class of the ontologically privileged must clearly include not only actual and inevitable persons. It needs to be supplemented at the very least by the class of those filling what I will call "inevitable slots." Consider the following case. I know a couple who plan to have a child. I can alter their lives slightly so as to change the time of conception, thereby causing a different child to be born. Accordingly, any particular child they have will not be an inevitable person from my point of view. Still, because there is no morally acceptable way in which I can completely prevent them from having *some* child, there can be said to be a slot that will

inevitably be filled though the particular child who fills it will not be an inevitable person. He or she will be the occupant of an inevitable slot. Will it do then to say that the top priority in population policy should be the welfare of actual and inevitable persons and the occupants of inevitable slots, and that the interest of others should count only as tie-breakers? Is this at last the combination of classes of persons whose interests can justifiably be placed above those of extra persons? No. The reason is again that the minimum possible population size for some given time in the future would exceed the number of people who would fall in the favored group of ontological categories for the population planner. In this case, the time in question would be somewhat farther into the future because we now have the occupants of inevitable slots as well as actual and inevitable persons. It still doesn't have to be so very *far* into the future, however, because there would soon come a time when a comparatively small part of the population was made up of the descendants of those alive at the time of the population planners' decision, and only those people could qualify as the occupants of inevitable slots.

The notion of an inevitable slot can be expanded in two ways. First, the class of occupants of inevitable slots could be expanded by allowing inevitable as well as actual persons to count as prospective parents of the occupants of slots; and secondly, the requirement for a child to be an occupant of an inevitable slot could be lowered so that it is only necessary that one rather than both of his parents is bound to have a child. But these changes wouldn't really make a significant difference; they would simply move the time at which the ontologically preferred persons would fall short of the minimum possible population somewhat farther into the future.⁷

It should be clear by now that any ontological preference theory will have consequences that none of us would for a moment be willing to accept. This applies not merely to relatively restricted ontological preference theories which restrict the ontologically privileged to actual and inevitable persons but to the various extensions of such theories with their increasingly far-fetched ontological categories. One consequence of this is that any theory that claims that there can be no obligations concerning possible persons, happy as well as unhappy ones, can be dismissed out of hand.

6. Peter Singer's Minimum Number Theory

At least some anti-obligation theories avoid the repugnant conclusion; namely, average theories and what I have called the ad hoc theory; but average theories are decisively refuted by the Pareto-plus objection, and the ad hoc theory is excluded for other reasons—among them, the fact that it is an ontological preference theory. Peter Singer's anti-obligation position while neither an ontological preference nor an average theory, still seems to avoid the repugnant conclusion. Following Parfit, Singer sees that it will not do to be concerned only with actual and inevitable persons:

If a possible future state of affairs is a world of P people at an average level of happiness A , it is wrong to bring into existence any greater number of people, $P + N$, such that no sub-group of $P + N$ contains P people at an average level of happiness equal to or higher than A .⁸

This is not in any way an ontological preference theory. It doesn't matter to Singer whether or not the persons whose average is to count in deciding on an ideal population size are actual or inevitable, and although he doesn't use the notion of occupants of inevitable slots in any of the ways I've discussed, it is clear that someone would not be entitled to special consideration on his theory simply because he occupied one of my slots. It might be argued that Singer's concern with the minimum number enables him to use "occupant of an inevitable slot" in a way I haven't considered: aren't all the people in whose happiness he is interested (except insofar as they are tie-breakers) occupants of the minimum number of slots? Perhaps so, but a more appropriate term would be "*degree of happiness slots*" because, for Singer, it is the degree of happiness rather than ontological status that gives someone priority in a population planner's consideration. By rejecting ontological preference theories, Singer avoids the troubles they entail; but he also gives up their supposed advantage—denying that there can be obligations concerning persons who are neither actual nor in some sense inevitable.

What Singer wants (as he makes clear earlier and as is only reasonable) is a principle that will ensure the highest possible happiness aver-

age for the group of P persons; but his formula does not meet this requirement because it would allow population planners to limit the population to P persons even if adding additional people would make it possible to have a sub-group of P persons whose happiness level was higher than the happiness level of P persons would be if there were only P persons. To avoid this, his principle should be modified to:

If a possible future state of affairs is a world of P people at an average level of happiness A , it is wrong to bring into existence any greater number of people, $P + N$, such that no sub-group of $P + N$ contains P people at an average level of happiness equal to or higher than A , *and it is wrong to fail to bring into existence additional people if they would increase the happiness of a sub-group of P people.* (My change in his principle is in italics.)⁹

This is equivalent to:

If the smallest possible population at some time in the future is P , we should aim at a population size such that the happiness level of a group of P members of it will be as high as possible.

I have objected¹⁰ that, as it stands, this principle tells you to add an extra person even if he would be extremely wretched as long as his addition will increase, no matter how little, the average for the privileged group. Singer could of course add a suitable proviso to avoid this, though this would, I think, run counter to the spirit of his work. In the event of a tie where two alternate population plans would result in the same average happiness level for the privileged group, Singer thinks that the welfare of those not in this group should be weighed simply as a tie-breaker. In other words, he is using a lexical principle where the welfare of extra people is the second-ranking consideration.

Though I have a high regard for Singer's theory, which is the logical next step after an ontological preference theory, it is ultimately untenable for the following reasons:

Objection 1. Suppose that you could have either (i) a relatively small population with a high happiness average (its size will provide the minimum number); (ii) a population twice as large with half of them, the minimum number people, at a slightly higher average than in

alternative (i) but with the other half at a happiness level only slightly over neutrality; or (iii) a population three times as large with a third of them (the minimum number people) with a higher happiness average than in alternative (i) though a bit lower than they would be on alternative (ii), and the remaining two-thirds only slightly less happy than the minimum number people. On Singer's view, we must choose (and this, I think, is counter-intuitive) the *second* rather than the third alternative. He is forced to this because, in the second alternative, the minimum number people would have a slightly higher average than in either of the others. This is hard to swallow but there is worse to come.

Objection 2. Suppose that the minimum possible population for a given time in the future is one, and that you will get the happiest sub-group of one in a fairly large world, World A , in which the average for the rest will be barely above neutrality; but that in a slightly smaller world, World B , everyone would be just a bit less happy than the minimum number person in World A . Singer would obviously be forced to choose World A . In choosing it, he would not only be opting for a world with both a lower happiness average *and* a lower happiness total (and one in which these characteristics could not be excused by considerations of ontological preference); he would have fallen into the repugnant conclusion: he would be choosing a world with a happiness average only slightly above neutrality (for the single happy person of course wouldn't significantly increase the average in his world) rather than a world with a high happiness average. And if it is argued that he would never be confronted in practice with a situation that would require him to make this objectionable choice, the same holds for the classical and the compromise principles. The minimum number theory lacks the very feature which had seemed one of its prime recommendations.

Objection 3 (the ABC objection). If an ethical system tells you to bring about state of affairs A in preference to state of affairs B , and then tells you that once you have A you should abandon it for B , the system is in trouble. If over and above this, circumstances could arise in which once A was brought about, B would no longer be possible but the system went on to tell you to bring about state of affairs C , which is in fact inferior to B , the system is in even worse trouble. Singer's minimum number principle combined with his traditional utilitarianism for actual persons is open to this objection—for short, the *ABC* objec-

tion. Singer's principle (as I have amended it) is that if the smallest possible population size for a given time is P (the minimum number), you should have just that size population that will give you the happiest possible sub-group of P persons. Thus, if the alternatives for some given time in the future are: *state of affairs AA*: 10 billion people with a happiness average of 9; *state of affairs A*: 20 billion people with the top ten billion at a happiness level of 12 and the other 10 billion with a happiness average of 1; and *state of affairs B*: 21 billion people with 10 billion at a happiness level of 11 and the other 11 billion at a happiness level of 10; it is clear that Singer must pick *A*. Suppose now that when *A* has been brought about, *B* is no longer a possible alternative but that one could instead (say, by drastically redistributing the wealth of the country) have *state of affairs C* with 20 billion people, 10 billion of them with a happiness level of 7 and the other 10 billion with a happiness level of 6. Singer's traditional utilitarianism for actual people would force him to choose *C*, despite the fact that he would initially have regarded *C* as the *worst* of the four alternatives. Thus, the *ABC* objection.

It might be argued that if you know that your moral convictions will force you to change from *A* to *C* later, you should not have chosen *A* originally. To take care of this, a proviso could be added such as: in determining what population size will result in the happiest minimum number, take into account any redistribution of means of happiness that utilitarianism for actual people will eventually require. Reasonable as this sounds, it leads to paradoxical consequences:

(a) The augmented rule would still tell you to bring about *A* rather than *B* or *C*, provided that you couldn't later reduce the happiness of the minimum number people by changing to *B* or *C*. But it would surely be odd for a rule to tell you to bring about *A* rather than *B* or *C*, despite the fact that it would also tell you that once you had brought about *A* (if only you could) you should abandon it for *B*, or, barring that, for *C*.

(b) In cases where you would choose *A* only if you knew that state of affairs *B* or *C* would *never* be future alternatives which your principle would force you to choose, you might sometimes be able in some way to tie your own hands—say, to pass a law so that once you had brought about *A*, alternatives *B* and *C* were no longer open to you. But it is surely peculiar to say that you should tie your own hands so that in

the future you won't be able to do what you ought to do if only you could.

Objection 4. Derek Parfit provides the principle which is the basis for this objection.¹¹ He says that if a woman means to have a baby in any event but if a baby conceived at time t would be happier than one conceived at time t^1 , then she should conceive the baby at time t . The interesting thing about this is that instead of the usual situation where the choices are between two alternatives in one of which a person would be happier than he would be in the other, the choices in this case are between two entirely different people, the reason being that if conception occurs at a different time, a different individual results.

The underlying assumption here seems to be that if you are going to create one of two possible people, you should create the happier one. If there is no conflict of interest this seems intuitive enough, but in a case where having the happier child would cause a loss to an actual person, it is not clear how much weight should be given to the interest of the actual person as opposed to the difference between the interests of the possible persons. I can formulate three possible versions of this principle, each giving progressively less weight to the interest of the actual person. First, the mother could have an obligation to conceive a baby at one time rather than the other only if it would involve no loss to herself. But this is too weak for Parfit's purposes. Consider next a case where the mother must make some sacrifice to conceive the baby at time t rather than at time t^1 . Let us say that her loss would be 10. This must be weighed against the difference in happiness between the two possible babies (let's say that this difference is 11). The moderate version of the principle would say that the difference between the babies' welfare should count *somewhat* less than her own welfare but not infinitely less. The strong version would say that the mother should count her own welfare as on a par with the difference between the respective welfare of the possible babies and that since this difference is 11 which is greater than her own loss, she should conceive the baby at time t .

The last two alternatives are both incompatible with the minimum number theory. Suppose that the minimum number is two and that the two persons are prospective parents who will benefit slightly more from having a child at time t^1 than from having a child at time t but that the child conceived at t would be a good deal happier than the one con-

ceived at t^1 . Both versions, strong enough to be plausible, would tell them to have the child at time t but (provided that neither of the possible children would be happier than their parents) the minimum number theory would tell them to have a child at time t^1 . (Parfit holds the strongest version of the theory. I agree with him, but I consider the other two because, when added disjunctively to the strong version, they are much harder to deny but still strong enough to be useful in attacking a number of important positions.)

Objection 5 (the timeless persons impartiality principle). Suppose that we know that roughly 10% of the population in the future will be born with some sort of congenital disease which will make them relatively unhappy and that we plan to make the population larger than it needs to be so that these people will not fall in the minimum number category. In terms of Singer's principle, this means that the interests of living persons are to take strict precedence over those of persons born in the future with the congenital disease. Suppose further that there is a scarce natural resource which, if it is not consumed by those now alive, could be used to relieve the distress of those future people with the congenital disease; and that, though those now alive would get some pleasure from the resource, it would do them far less good than it would do the diseased people. According to the Singer theory, it is perfectly justifiable for those alive to use up the resource. This violates what I will call the timeless impartiality principle, that we should be impartial in weighing the respective interests of everyone who exists in a timeless sense.

Or imagine another case: where the more those now alive suffering from this congenital disease use the natural resource, the better for them; but where at some point the law of diminishing returns sets in dramatically, so that they derive progressively less and less benefit from larger doses. Since they fall into Singer's privileged category, his theory directs us to use up the resources on them; whereas the timeless principle would tell us to save some for more efficient use on future persons with the same disease.¹²

7. *Timeless Utilitarianism Is Incomplete*

The timeless impartiality principle coupled with the core of utilitarianism (that the utilitarian principle applies at least to everyone currently alive) yields timeless utilitarianism (that the utilitarian principle

extends to everyone existing in the timeless sense). Because most utilitarians are likely to find timeless utilitarianism hard to reject, any version of utilitarianism which can be derived from it will be equally hard to reject. I believe that Narveson, Sumner and Warren think that their particular kinds of utilitarianism can be derived from timeless utilitarianism, despite the fact that Sumner takes a pro-sacrifice, and Narveson and Warren, anti-sacrifice positions. I mean to show that these derivations are not possible because timeless utilitarianism does not give an answer one way or the other to the question of whether there is intrinsic worth in adding happy people.

First off, for an action to be obligatory, it is necessary not only that it not be wrong itself but that no alternative action be permissible. Given this, it can be shown that it is neither obligatory to add happy people in order to increase the total happiness at the cost of a sacrifice from those currently alive nor obligatory *not* to add them. If you add them, they will always have existed in the timeless sense so that it will not have been wrong by adding them, to have treated their happiness on a par with the happiness of currently existing persons. If you don't add them, they will not have existed in the timeless sense so it will not have been wrong, by adding them, not to have treated their happiness on a par with the happiness of those alive. But if both adding them and not adding them are permissible, neither is obligatory. Consequently timeless utilitarianism is incomplete: it doesn't help us to decide whether or not there is intrinsic worth in adding happy people to the world.

Timeless utilitarianism is, for a similar reason, incomplete in Parfit's baby case. If the happier baby is conceived at a time which involves some inconvenience to his parents, it will not have been wrong, provided the loss is not too great, because the happier baby will have existed in the timeless sense so that it will not have been wrong to have given his happiness full weight. If on the other hand, the less happy baby is conceived at the more convenient time, this will not have been wrong either because the happier baby will not have existed in the timeless sense so that his happiness did not have to be counted.

8. *The Eclectic Position and the Transitivity Argument*

At this point, I am tempted to make a final try to take the bits and pieces that are left and to construct a plausible position for those who

hold that there is no intrinsic value in the preservation of mankind. The theory I have come up with is a minimum number theory, and it seems to me that it should have the following characteristics, for reasons I'll make clear as I go along:

(i) Because timeless utilitarianism seems right to utilitarians as far as it goes (even though it is incomplete), it should contain the principle of timeless utilitarianism.

(ii) It needs to be supplemented by either a moderate or a strong version of Parfit's principle.

(iii) To avoid the objections to ontological preference theories that led Singer to reject them, it must require us to take the welfare of at least the minimum number of persons into consideration even though many of them will not belong to an ontological elite.

(iv) As a minimum number theory, it must require us to be concerned with the future welfare of many prospective persons who don't belong to any ontological elite. Otherwise it would be vulnerable to the manifold objections to ontological elitism.

(v) To avoid conflict with timeless utilitarianism, it must require that if we add people in order to benefit the minimum number, the happiness of all the added people must be counted on a par with that of actual persons.

(vi) In order to be an anti-sacrifice view, it must hold, of course, that there is no intrinsic value in adding happy persons (unless they will be happy enough to be included in the minimum number group).

The resulting eclectic position can't be supported by arguments based on either the repugnant conclusion objection or the possible persons objection. I suspect that whatever attraction the eclectic position may have derives from the fact that it is a philosophical descendant of theories which were supported by arguments on which it can't itself rely. Mongrel as this conglomerate of principles may be, it deserves attention since it seems to be the best that can be done for those who don't want to hold that there is intrinsic value in the preservation of mankind.

Both (ii) and (v) can require us (though in different ways) to make sacrifices when we add people in order to get the particular people who will be happiest.¹³ Since there might be other ways in which we could be called upon to make sacrifices for the same sort of objective which the eclectic position doesn't say anything about, it is not yet complete. For a philosopher with the consequentialist orientation of a utilitarian,

the next move would be to make some general rule of which (ii) and (v) would be corollaries. But in order to do this one must first choose between the two versions of Parfit's principle. In fact, one must choose the stronger of the two to get a principle that will match (v) in its indifference regarding whether the total happiness should be raised by increasing the happiness of those alive or by arranging that the happiest possible people should be born. Still, whatever version of Parfit's principle you choose, you will be led inevitably to a pro-obligation theory. If you select the moderate version, it will be a compromise principle; if you accept the strong version, it will be the classical principle. I mean to show this first with the moderate version because I suspect it is the one that most people would find hardest to reject.

I am not claiming that either version of Parfit's principle implies a pro-obligation view. It is rather that there is no satisfactory utilitarian rationale for accepting either version unless you hold that there is intrinsic value in adding happy extra people. Narveson accepts Parfit's principle as part of an overall principle that is like the eclectic position I've just constructed except that it lacks (iv) and (v) (those pertaining to Singer's minimum number theory) and includes the weak rather than the moderate version of Parfit's principle. Narveson takes his overall principle (including the part from Parfit) to be implicit in timeless utilitarianism, which he regards as providing a utilitarian rational both for an anti-sacrifice view and for Parfit's principle. But timeless utilitarianism is incomplete with regard to both.

I can think of only two ways in which a utilitarian could justify Parfit's principle (or rather, its moderate version). The most obvious and, I think, the correct way is to claim that a prospective mother should make a small sacrifice to get a considerably happier child because: (1) other things being equal, a world that contains a happier person will be better than a world that contains a less happy person in his place; (2) happiness added to the world through adding the happier rather than the less happy person should be discounted in relation to the effect it will have on the happiness level of those currently alive; and (3) one should always do the thing that will result in the best of the possible worlds one's action might lead to. If one can improve the world by adding a happy rather than a "neutral" person (i.e., a person, so to speak, with equal amounts of utility and disutility) and if (as a utilitarian must hold) adding a neutral person does not make the world better or worse than adding no person at all, then (given the normal

transitivity relation) one can improve the world by adding a happy person. How much the world is improved by adding someone with a given happiness level is a function of the discount rate chosen in step 2. A similar line of reasoning starting from the strong version of Parfit's principle would lead directly to classical utilitarianism because with the strong version, no discounting is required.

A second rationale could conceivably be given for accepting Parfit's principle: (1) given that the world is to contain a certain number of people, following Parfit's principle will indeed make the world better but (2) one can't make the world better by adding extra people no matter how happy they will be unless they will have a beneficial net effect on the people already there.

The transitivity argument would seem to show that given (1), it is implausible to hold (2). It could conceivably be claimed that transitivity doesn't hold in this case, but how could such a move be justified? It won't do to object that it may be obligatory to add extra people on the grounds that there can be no obligations concerning possible people, because the denial that there are such obligations leads to totally unacceptable conclusions. Nor will it do to say that there can be no obligation relating to what Mary Warren calls "merely possible persons," where a "merely possible person" is a possible person who will never exist; it won't do because there *can* be an obligation relating to a merely possible person who would be wretched if he existed—namely, the obligation not to add him. Nor will it do to say that a world can't be made better or worse by the addition of someone who would have no net effect on those existing. Clearly, the world *would* be worse if the person added was unhappy. Instead the claim will have to be that there is an asymmetry here: that although the world can be made worse by adding unhappy people, it cannot be made better by adding happy ones. This claim is the subject of § 9.

9. An Argument for the Obligation Theory

In actual practice, it would never be justifiable *per se* to add any large group of people to the world unless there is at least some intrinsic moral worth in adding happy people to the world. The reason for this is that one can be virtually certain that at least *some* members of any large group that might be added to the world will, for one reason or

another, have wretched lives, which of course counts negatively; so, unless adding happy people is to count *positively*, no matter by how much the happy outnumber the wretched and no matter how happy they are, it would be as wrong to add the whole group, happy members included, as it would be if it contained only its most wretched members.¹⁴

An alternative form of this argument: it would always be wrong, taken in itself, not to prevent the existence of any large group of people unless it is *prima facie* wrong to prevent the existence of happy people; otherwise, there would be nothing to offset the *prima facie* wrongness of failing to prevent the existence of the wretched people who would inevitably be members of any large group.¹⁵

William Anglin objects¹⁶ that the most I have proved is that it is a supererogatory good deed to bring into existence (or not to prevent the existence of) happy people. My response¹⁷ is that a utilitarian cannot admit that bringing about a given outcome is a supererogatory good deed unless we have at least a *prima facie* obligation to bring about that sort of outcome.

Bennett makes three objections to my argument,¹⁸ the second of which is by far the most interesting:

Someone who thinks that there is a moral case for the continuation of mankind, over and above one stemming from the (dis)utilities of actual people, need not immediately fling himself into the arms of a principle enjoining the maximization of happiness. There are other bases he might appeal to, some of them more plausible than any amount-of-happiness principle; e.g. the principle which enjoins the completing of unfinished business. . . .¹⁹

(Although the core of Bennett's objection is contained in this passage, I will draw on the rest of his paper as well to make his case as strong as possible.) He thinks that we can get something adequate to offset the wrongness of bringing the wretched people into existence without holding that it would be *prima facie* wrong to prevent the existence of the happy members of the prospective group. He is right in holding, in effect, that an action that is *prima facie* wrong may sometimes be permissible even if it would not be *prima facie* wrong *not* to do it. For example, suppose that I have a life preserver which someone else needs but which I also need. It is *prima facie* wrong for me to keep it because he needs it. It would *not*, however, be *prima facie* wrong for me not to

keep it but to give it to him. Even so, it is *permissible* for me to keep it because of my own need. My own need makes it *prima facie* permissible (and indeed permissible *tout court*) to keep it without making it in any way wrong for me not to keep it. My need is a factor in the situation which tends to make keeping the life preserver permissible without making not keeping it in any way wrong.

Typically, such characteristics of actions are egoistic. Is there a characteristic of actions, some sort of balancing factor, which is egoistic, which tends to make the act morally permissible, and which would allow the bringing about of the existence of a large group of people despite the fact that some of that group will lead miserable lives? It is conceivable (though unlikely) that such a characteristic can be found if the number of miserable people the act would cause to exist is very small. Bennett thinks, however, that there is a balancing factor strong enough to offset even the bringing-into-existence of a great *many* miserable human beings—as would be the case, for instance, if the future of humanity were involved. Let me construct a case with an incredibly large number of possible people. Jones is the sole survivor of a nuclear war. People who are now dead and toward whom Jones has no obligation have started a batch of test-tube babies. Because they will be extremely small even when they are full-grown, the earth will be able to support a truly *enormous* number of them, in fact, 1,000,000², all of whom but a very unhappy million will have rich and happy lives. If Jones chooses, he can totally prevent their existence, and the prospect of the very unhappy million little people makes it *prima facie* wrong for him not to prevent the existence of the whole lot. Suppose that Jones has some sort of interest in these possible people, an interest which has nothing to do with the well-being of the enormous number of happy ones—perhaps an aesthetic interest in having mankind continue. Surely it is obvious that if there is only Jones's own aesthetic interest to counterbalance the great moral weight of the *prima facie* wrongness of not stopping the birth of a million wretched people, that interest wouldn't even *begin* to be weighty enough.

However, Bennett makes it clear that his own "pro-humanity stance" is not aesthetic but "intensely practical." Furthermore, his central stance is that it is not a matter of principle: he wishes to avoid principles because they may commit him to unacceptable conse-

quences. He offers as his reason for being prepared to make sacrifices and even to fight for the preservation of mankind, not the principle that one should not prevent the completion of unfinished business (or even business of some particular kind) but his interest in the particular unfinished business of the human race, an essential factor being the fact that *he* is a member of that race.²⁰ He urges that "each person is morally entitled to give some special weight to his own wants and needs and interests, just *qua* his." Suppose then that my man Jones, like Bennett, has an intense interest in seeing certain human projects completed and that he has reason to suppose that if the existence of the test-tube babies is not prevented they will complete those projects. Would this interest, even granting Bennett's point that special weight should be given to the fact that it is *his* interest, be of sufficient weight to offset the evil of causing a million wretched lives? Obviously not.

Alternatively, Bennett can take the line that he *is* appealing to a principle to get his counterweight, a "principle that enjoins the completing of unfinished business." Previously, his proposed counterweight was egoistic. This time, it is non-egoistic, which makes it more promising. However, it still won't do. Let's change the previous case slightly and have the test-tube babies initiate a new set of projects of their own rather than choosing to complete mankind's unfinished business. Despite the fact that there is now no unfinished business in this situation which could serve as a balancing factor, it still would be morally permissible not to prevent their existence even though there would still be a million wretched lives.

But perhaps some other non-egoistic reason can be found to outweigh the moral reason for preventing the existence of these future generations. It cannot be a consideration having to do with the pleasure, wellbeing, or fulfillment of interest of the almost 1,000,000² happy people because that would be a moral consideration and of no use to Bennett. But if the balancing factor is meant to be some value in the completion of the various projects that is *independent* of the contribution to the interest-fulfillment,²¹ pleasure, or wellbeing of the 1,000,000² people those projects may make, I am at a loss to see how this could counterbalance the weight of the million miserable lives. I would be equally at sea if the counterweight were supposed to consist in something like the value sometimes attributed to preserving particu-

lar species—apart from the wellbeing of their members, apart from their ecological role, and even apart from the aesthetic pleasure that some people might take in looking at them or knowing that they were there. Here again, I can't for the life of me see how it could be enough to make it morally acceptable to let a million wretched people be born.

10. A New Defense for Classical Utilitarianism

In §9, my argument was for the obligation view in general; here I will give a new argument for classical utilitarianism, based on the same underlying intuition.

(1) To prevent the existence of a person who would have, by a given amount, more unhappiness than happiness in his life is as important morally as it is to prevent an equal amount of unhappiness for a person who is alive.

(2) Other things being equal, if it is expected that the total happiness and unhappiness of a given group of people will be equal, it is not wrong to add them.

(3) But (2) would not hold unless there is at least as much positive value in increasing the happiness total of mankind by adding happy people as there is negative value in adding unhappy ones.

(4) Given (1) and (3) plus timeless utilitarianism, we can infer that classical utilitarianism is correct.

Some of those who disagree with me on whether or not it can sometimes be obligatory to make sacrifices for future generations may attack premises (1) and/or (2) under the mistaken impression that if those premises are weakened, the argument will result in some sort of anti-obligation theory. But the only sort of conclusion to which plausible forms of (1) and (2) can lead is still an obligation theory.

Perhaps the most fragile of these steps is (2). Though I think that (2) is in fact correct, one could assert instead that a group of people ought not to be added unless their total happiness will considerably exceed their unhappiness. (2) could be modified to avoid this objection but that would weaken the argument's conclusion. The argument would then be:

(1) To prevent the existence of a person who will have, by a given amount, more unhappiness than happiness is as important morally as to prevent an equal amount of unhappiness in a person who is alive.

(2a) Other things being equal, it is wrong to add a group of people unless one expects the happiness of the group to exceed its unhappiness by a ratio of x to 1.

(3a) But (2a) would not hold unless there is at least $1/x$ times as much positive value in increasing the happiness total of mankind by adding happy people as there is negative value in adding unhappy ones.

Given (1) and (3a) plus timeless utilitarianism, we can infer (4a): there is at least $1/x$ times as much positive value in increasing the happiness total of the world by adding happy people as there is in preventing an equal amount of suffering (or adding an equal amount of happiness) for those alive (compromise principle 2).

Though he is bound to regret losing the simplicity of the classical principle, a utilitarian might tolerate (4a) provided the ratio isn't *too* far from the one-to-one ratio of classical utilitarianism. But might we be forced to accept such a high happiness-excess ratio in (2a), say, 10 to 1, that it would result in a conclusion really quite far removed from classical utilitarianism? I think not, because such a ratio would lead to counter-intuitive results. If it were taken to be wrong, other things being equal, to add a group of people unless it is probable that their happiness total will be at least three times (let alone ten times) their unhappiness total, one would be forced to regard it as wrong *per se* to add *any* future generations unless one were extremely optimistic, because it is highly unlikely that they would on the average have three times as much happiness as unhappiness.

Alternatively, the first premise might strike some people as mistaken on the grounds that it is not, in fact, always wrong to add a person whose unhappiness would exceed his happiness—it is only wrong when the ratio would be fairly high—let's call this level Φ . If one wishes to go along with this, my original premise could be replaced by another according to which it is *prima facie* wrong to add a person only if his ratio of unhappiness to happiness exceeds level Φ . If, say, one held that there was negative value in adding a person only insofar as he would have more than five times more unhappiness than happiness and a given person would be likely to have six times more unhappiness than happiness, $5/6$ of his unhappiness could be disregarded in deciding whether it would be wrong to add him. This would provide a basis for another attempt to reconcile our normal intuitions about which groups of prospective people it would not be wrong to add with a high happiness-

excess demand in the (2a) ratio: we would have such a little excess of unhappiness to count negatively that a high ratio of good to bad could be required in (2a) without violating our intuitions as to which groups it would not be wrong to add. With these revisions the argument would lead to compromise principle 2.

It is doubtful that utilitarians will embrace the resulting argument, however, because not only is compromise principle 2 unwieldy when compared to the elegance of classical utilitarianism, the choice of a particular ratio for (2a) will inevitably be somewhat arbitrary as well—which will be reflected in the conclusion. Besides, taking account only of a fraction of the unhappiness excess of a terribly unhappy person, in deciding how much negative value there would be in adding him, is inconsistent with a hedonistic version of timeless utilitarianism.

It has been claimed that hedonistic utilitarianism is implausible because it would result in the outrageous requirement that people who have more unhappiness than happiness should be killed. Surely this is not the case because, among other things, any such policy would not only evoke violent and widespread revolution but would cause such enormous dread and misery that it could not be defended on hedonic grounds. Even so, there is no question that a hedonistic utilitarian must hold something that many people could not countenance: that is that it is irrational to regard one's life as having intrinsic worth if it contains more unhappiness than happiness. I am willing to go along with this myself in that I think that although we shouldn't go around killing unhappy people, we also shouldn't knowingly bring them into the world.

The problem for a hedonistic utilitarian is that his system forces him to hold that a life with more pain than pleasure, even a little bit more, is not worth living and is not worth bringing into existence. A non-utilitarian can avoid the problem by holding that a life with more dis-utility than utility can be worth living and can be worth bringing into existence, but a utilitarian can't. If a philosopher cannot swallow this stiff consequence of hedonism but wants to remain a utilitarian, there are two things he can do—be an ideal utilitarian or hold an interest theory. If he does the first, he can hold that unpleasant as well as pleasant experiences have intrinsic value, and that the intrinsic value of a pleasant experience is not solely a function of how good it feels but can far exceed the value of the pleasure alone. Could an ideal utilitarian then argue that, with these new sources of intrinsic value, there

would be so much excess intrinsic value (even in the case of groups of people whose addition was taken to be only marginally permissible) that we could justify the decision to add them even if the happiness-excess ratio in (2a) was so high that it would lead the argument far from classical utilitarianism? In order to do so, some method of measuring intrinsic value must be found, a notorious headache for ideal utilitarians. If one gives up the whole job of measurement as hopeless, one will not be in a position to decide between (2) and (2a). Nevertheless, if my argument for classical utilitarianism is sound in other respects, one will still be forced to accept some form of obligation theory. The particular form will depend on whether one accepts (2) or (2a): if (2), classical utilitarianism follows; if (2a), some other pro-obligation view. But without a way of measuring intrinsic value, there is no reasonable way of choosing between (2) and (2a). If on the other hand, one tries to come up with a workable method of measurement, the best I can think of is something like asking people at the end of each day whether that day has been worth living; or, putting this in more technical terms, whether they felt that their experiences contained more that was intrinsically good or more that was intrinsically bad. I suspect that those who want a high ratio for (2a) so that they can hold the compromise position instead of classical utilitarianism, won't find much support in the answers they get. Nor do I think it will do them much good to claim that the goodness of the good days exceeds the badness of the bad days sufficiently to get the kind of overall excess of intrinsic value that their strategy requires.

The other possibility is an interest theory. Sumner argues that, on an interest as opposed to a hedonistic version of utilitarianism, one would very rarely find a life in which there was an excess of negative over positive value. If one holds such an interest theory and one also holds (as Sumner would) that most lives have far more positive value than negative value, one can again adopt a very high ratio in (2a), so that although one's conclusion will attribute some value to adding happy people, it will be relatively little. Sumner's claim is that most lives have more positive than negative value; but for a large number of people, life is more a matter of the non-fulfillment of negative interests than the fulfillment of positive ones. That is, their life is more a matter of occurrences that they would like to avoid but can't than of occurrences which they positively desire. Work occupies a large part of most people's lives and many people, perhaps most, seem unfortunately to

find their work something that they hope at best to get through with relatively little pain rather than as a source of positive fulfillment.²² Think how many people would rather sleep than work, not to speak of all the physical and mental illness in the world, the misery, injustice and loneliness, a list as grim as it is long. One might argue that even people whose lives are more a matter of having their negative interests frustrated than of having their positive interests fulfilled usually have an extremely intense though irrational interest in avoiding death as such, i.e., an interest over and above their desire to avoid it because it will prevent them from fulfilling any more of their positive interests. Given this, it can be urged that even though living involves a great deal of frustration of their negative interests, dying would involve even *more* because of the sheer intensity of the desire not to die and that therefore we should try to keep even such unhappy people alive. Except for extreme cases, this conclusion may well be true; but though it provides a reason for preserving the life of a person in this unhappy position, it does not provide a reason for denying that, taken in itself, it is a bad thing to add him to the world.²³

When I look back at the intricate series of arguments and counter-arguments in this section, I am consoled to note that at the very least they show that some sort of theory which allows for the value of adding happy people must be correct. However, this was already quite clear. As for the case I've tried to make for classical utilitarianism, I think if one begins with a favorable attitude toward hedonistic utilitarianism, the argument is hard to resist. But for those who hold an interest theory or ideal utilitarianism, the case for classical utilitarianism doesn't look quite so strong. A revision of (2) that would allow them to reject it in favour of a compromise principle quite close to it is somewhat plausible, and a revision that would allow them to have a compromise principle quite far removed from it would not be *totally* implausible. But who wants a compromise principle anyway? Furthermore, a form of the argument which is neutral as to what is intrinsically valuable can be given:

(1) To prevent the existence of a person who would have, by a given amount, more of what is intrinsically bad in his life than whatever is intrinsically good is as important morally as it is to prevent an equal amount of what is intrinsically bad for a person who is alive.

(2) Other things being equal, if one expects the total amount of

whatever is intrinsically good to equal the total amount of whatever is intrinsically bad for a given group of people whom one could add, it is not wrong to add them.

(3) But (2) would not hold unless there is at least as much positive value in increasing the total of whatever is intrinsically good for mankind by adding people with an excess of whatever is good as there is negative value in adding people with an excess of whatever is bad.

(4) Given (1) and (3) plus timeless utilitarianism, we can infer that classical utilitarianism is correct.

The new version of the argument should seem much more compelling than the old one for a non-hedonistic utilitarian. The second premise is still the crucial one. Isn't it in the last analysis implausible to deny it now that it is neutral as to what is intrinsically valuable? If one expects equal amounts of positive and negative value (whatever one takes them to be) for a group of possible people, isn't it hard, other things being equal, to deny that it is not wrong to create them? And why resist this highly plausible premise when one's only reward is a compromise principle one would rather do without? Given a choice between the classical principle coupled with premise (2) and the unwieldy compromise pro-sacrifice principle coupled with premise (2a), surely not many philosophers would choose the latter. If what one really wants is neither of them but a theory that assigns no worth whatever to adding happy people to the world, the price is preposterously high—it is to wreak havoc with our other intuitions.

11. *Meta-ethics, Meta-meta-ethics and the Obligation Principle*

Up to now my strategy has been to bury anti-sacrifice views under a heap of moral intuitions that go against them, and my discussion has remained strictly within the bounds of normative ethics. Now I turn to meta-ethics to consider various analyses of the concept of obligation; also, to what might properly be called *meta-meta-ethics*, to ask how much, if anything, in practical deliberation about future generations hinges on which meta-ethical theory turns out to be correct.²⁴

Brave New World. Smythers and Green are two of the few sterile survivors of a nuclear war. Smythers is a musician with a bad heart. He would like nothing better than to spend the rest of his life with his

Steinway, just where he is, working on the late piano works of Beethoven. Green is a scientist who has a formula for making test-tube babies to repopulate the world, babies who will have excellent prospects for happiness and achievement. They will, among other things, be able to take full advantage of the cultural resources left behind by past generations. Because Smythers has enormous respect for Green, he will do whatever Green tells him. If they didn't have the babies to take care of, they could have lives of ease and pleasure with the other survivors whereas they know that taking care of the children will require major sacrifices from both of them. Smythers in particular realizes that the effort will shorten his life. Despite this, Green decides to make the babies and Smythers accepts his decision. The babies grow into a god-like race of intelligent, strong, happy and loving people who not only fully appreciate the works of the great composers, writers and artists of the past but who create even greater works themselves. Although Smythers enjoys seeing the results of his sacrifice, the enjoyment he derives does not equal the sacrifice he made. Still, in retrospect, he is glad that Green persuaded him to make the decision he did and he would be ready to make the same decision again. (It seems to me, incidentally, that Smythers could well have been glad to have made the sacrifices he did for the new race, even if they ignore the unfinished projects of mankind and start their own instead.)

There is, I believe, a broad sense of "obligation" such that, given Smythers' attitude, a strong case²⁵ can be made that he had not just an aesthetic or other non-moral preference for the preservation of mankind but a moral *obligation* to make the sacrifices he did to repopulate the world. There is perhaps a narrower sense of the word as well in which he did not; but I will argue that even if there is, the broader sort of obligation takes precedence. I hope that most of my readers will find my account of Smythers' response plausible, and that they would go so far as to say that, in the same situation, they would be glad if they had done the same thing he did. Even so, I am sure that some people will think that I am wrong, that there is no sense of "obligation" such that Smythers had an obligation to make the sacrifices he did; or that even if there is such a sense, it lacks the precedence I claim for it. In order to meet this objection—which I regard as formidable—I need to make a considerable detour into meta-ethics. Thirty years ago, philosophers paid too much attention to meta-ethics rather than normative ethics, but we are now doing the opposite. With the exception of Hare,

none of the main writers on obligations relating to possible persons deals with the meta-ethical aspects of the problem. This is, I think, a mistake.

To begin. Non-naturalism, the theory that ethical judgments cannot be translated into non-ethical terms and that ethical terms therefore refer to distinctively ethical qualities, has been shown to be unsatisfactory because there are no distinctively ethical qualities. Ethical naturalism can be divided into two kinds, "subjective naturalism" and "objective naturalism." Subjective naturalism is the theory that ethical judgments can be translated into non-ethical terms and that the rightness of an action is a function of some actual or ideal person's attitude towards it, while objective naturalism takes the rightness of an action to be independent of any such relation. Moore's open-question argument is not enough to show that all forms of ethical naturalism are false: it only shows that ethical terms are not overtly synonymous (i.e., obviously synonymous) with non-ethical terms, not that they are not covertly synonymous with them (i.e., synonymous but not obviously so).²⁶ The same would hold for tests in which we try to show that a given non-ethical term is *not* synonymous with an ethical term on the grounds that it does not strike most normal users of the language as contradictory to say that one of the terms applies to an action but not the other. For example, you can't show that "right" is not covertly synonymous with "conducive to the maximization of happiness" by noting that most people don't regard it as contradictory to say than an action is right but that it will not maximize happiness. However, you can show that "right" is not even covertly synonymous with "conducive to the maximization of happiness" by showing that many normal users of the language do not even take the two terms to be entirely or even approximately coextensive. You may use two terms as synonyms without realizing that you are doing so and therefore fail to pass the open-question test, but you can hardly use two terms as synonyms and yet firmly and repeatedly contend that one of the terms applies to a given action and the other not. The same attack may be repeated against any of the other objectivist candidates for synonymity with ethical terms. More important in meta-meta-ethics: even if some objectivist term did turn out to be synonymous with an ethical term, it wouldn't make any practical difference.²⁷ Though I am very sympathetic to hedonistic utilitarianism, I wouldn't feel I had done much for it if I managed somehow to show that "right" really is synonymous

with "conducive to the maximization of pleasure." The fact that the word "right" happened to have that meaning might be interesting but it wouldn't give me another reason to decide to maximize pleasure. Imagine two worlds, in one of which the word "right" is not used at all but only the expression "conducive to the maximization of pleasure," while in the other the two expressions are used as synonyms. A person in the second world would have no more reason to maximize pleasure than someone in the first.

The simplest form of subjective naturalism is the view that what it means to say that an action is right is that the speaker favours it. This view is clearly wrong because it doesn't allow for the possibility of mistaken ethical judgments (except in those cases where the speaker doesn't know what he favours). Simple emotivism, the view that to say that an attitude is right is not to make a statement about one's attitude but simply to express it, is open to the same objection. C. L. Stevenson's attempt to save emotivism from this objection by saying that ethical judgments have descriptive meaning which varies from one group to another as well as emotive meaning is unsuccessful. It enables us to account for some but not all of the ways in which ethical judgments can be mistaken.

Ideal observer theories. In an ideal observer theory, an action is taken to be right if it is the action that an ideal observer would choose to have performed. Very generally, the ideal observer is said to be benevolent and calm and it is stipulated that his choices are not to reflect cognitive defects. More precisely, it is stipulated that his choices will not be affected by logical errors or by failure to make the logical inferences relevant to (i.e., capable of influencing) them; he will have all the conceptual apparatus relevant to those choices; and he will have the relevant factual information as well as a vivid idea of the relevant consequences of his choice.

Such a person is said to be epistemically rational, calm and benevolent. (The qualification "epistemically" is necessary because without it, "rational" might be taken to imply something over and above epistemic characteristics.) Having a vivid idea of the consequences of an action is part of epistemic rationality, because one would like the observer to have knowledge by both acquaintance and description of the consequences of an action; and a vivid idea of the consequences is what thoroughgoing knowledge by acquaintance would give him. An advantage of acting in the way you would if you had knowledge by acquaint-

ance of the results of your action is that it would keep you from being sorry for what you had done when you saw what it led to. Smythers acquired knowledge by acquaintance of a good part of the results of his action and the fact that this led him to be pleased with what he had done would provide an advocate of the ideal observer theory with the beginnings of a case that he had made a good decision.

I am happy with the requirement that the ideal observer be calm and epistemically rational, but serious problems arise from the stipulation that he be fully benevolent. First, it means that we would sometimes be required to perform actions that are beyond the call of duty: it would require us to categorize behaviour that we usually regard as supererogatory, such as that of Albert Schweitzer, as obligatory. Secondly, it is, I think, desirable that ethical terms be defined in such a way that most people would have strong motivating reasons for being moral. Suppose that in order to persuade someone to make a genuine sacrifice for someone else (who will gain more than the agent will lose) I tell him that he ought to do it. On the ideal observer theory, this is the same as saying that an ideal observer would choose to have him do it. Because the ideal observer is by definition benevolent, it follows analytically that he will make the benevolent choice; but the person I am trying to persuade could well reply that what a benevolent person would choose is of no significance to *him* because he is not benevolent himself.

Another problem. It is by no means clear that we have been told enough about the ideal observer to be able to decide what he would say about such questions as whether it would be wrong to prevent the existence of future generations. If additional characteristics were ascribed to him, there would be a risk of circularity as well as a repetition of the problem I just found in attributing benevolence to him. Suppose that the ideal observer is defined as a person who, among other things, cares as much about adding happy people as about the happiness of already existing people. If I try to persuade someone to make a sacrifice to create a happy person on the grounds that that is what an ideal observer would want, he can say, "But I'm not like your ideal observer. I'm just not the kind of person who cares about adding happy people." This problem cannot be avoided by stipulating that the ideal observer, besides being epistemically ideal and benevolent, be normal in other respects, because one would then have to decide what the normal person's attitude towards happy extra persons would be; and even if a

normal person would have the attitude you wanted, someone could say, "Well, yes, the ideal observer may have the attitude of a normal person but I don't, and I don't care if I am not normal. And to this there is no answer.

My theory avoids this impasse by dropping the stipulation of benevolence, and by substituting the agent himself in an epistemically rational state for the ideal observer. To say that I ought to do a given thing is to say that I would choose to do it if I were calm and epistemically rational.²⁸ Now there is a good answer to the question, "Why should I be moral?" or "Why should I do the moral thing?" "Because it's the thing I would do if I were epistemically rational." Most people would regard this as a strong reason for doing it. Sometimes, of course, this won't persuade us to do what we ought to do but this enables the theory to account for *acrasia*: the theory would make the conviction that one ought to do a given thing about as persuasive in getting one to do it as such convictions usually are. Some readers may worry that, with the stipulation of benevolence removed, the theory won't require fully benevolent behavior from any typical human being because typical human beings, even if they were epistemically rational, would probably not choose to be fully benevolent. This is true, and consequently my meta-ethical theory cannot support straight utilitarianism.

Actually, I take this to be an advantage, because traditional utilitarianism with its insistence on complete benevolence is forced to construe supererogatory behavior as obligatory. My meta-ethical theory would serve as a basis for a kind of utilitarianism without supererogation where, while it is never wrong to do the optimific thing, one is not *obligated* to do it when doing so requires a personal sacrifice, unless other people would gain proportionally more than the agent would lose: the ratio of the gain to others versus the loss to oneself is the one an individual would choose if he were epistemically rational.²⁹ I believe that most of us, if we were epistemically rational, would be willing to make sacrifices for others. I am relying heavily here on the fact that although epistemic rationality does not logically require that we be sympathetic towards others, the requirement that we be vividly aware of others' experiences would, when combined with the other aspects of epistemic rationality, contingently involve for almost everyone at least some degree of sympathy towards them. Some people can, of course, become aware of the suffering of others and not care about it or even be glad. Such attitudes are frequently the result of conscious or

subconscious errors about the attitudes of others towards oneself. A person who became epistemically rational would discover the influence of such errors on his attitude and even when the error-based attitudes lingered, he would try to discount them in making his decisions.³⁰ Of course we confront a great deal of real malice, hostility and envy in the world; but even here epistemic rationality can have a softening effect if, as Ayer and others have held, it is logically impossible that anyone should be ultimately responsible for his actions and character and therefore that no one deserves the torments of hell (or any other torments) if he is bad any more than if he is good. Still, there may well be some people who would, even if they were epistemically rational, more or less totally lack sympathy for others; if so, they would, on my theory, have no obligation to make sacrifices for them. If, on the other hand, "obligation" is defined to include benevolence, even such people would have an obligation to make sacrifices for others; but with "obligation" defined this way, these people would have no motivating reason to do what they were obligated to do, so nothing would really have been gained.

It might be thought that my sort of utilitarianism without supererogation would result in egoistic public policies (including policies regarding future generations) rather than the altruistic policies prescribed by traditional utilitarianism. Surprisingly, it would not. One of the main strengths of the view is that it would not only require us to favor traditional utilitarian public policies but would give us a strong motivating reason for doing so, without having to assume that we are fully benevolent or even that we would become fully benevolent if we were epistemically rational. The demands for true personal sacrifice in the private sphere would be much less great with my system. But this is not particularly worrisome because, even if one were restricted to nothing more than sophisticated egoistic considerations, a reasonably high degree of morality in the private sphere could still be justified. Furthermore, I suspect that for most people the difference in the private area, between the demands of my system and the demands of traditional utilitarianism, could be largely a matter of supererogatory demands.

But how can this theory justify a traditional utilitarian position on matters of public policy (or, to be strictly accurate, something extremely close to it)? When a public policy that would have the best results for mankind would require personal sacrifice from one—in the form of

higher taxes, say—almost invariably the public gain would be immensely greater than one's own sacrifice. Consequently, unless one were extremely egoistic and would remain so even if one were epistemically rational, one would have an obligation on this theory to favor the policy in spite of the higher taxes one would have to pay. For example, suppose that it's a matter of Canada giving wheat to India where following the altruistic public policy would require a dollar from you as well as a dollar from each of the twenty million other Canadians, and that spending the money on India would do roughly twice as much good there as spending it at home. The sacrifice/gain ratio in this case would be roughly one to twenty million: in having the higher taxes imposed, an average taxpayer would lose only a twenty-millionth of what mankind would gain. With a ratio like this one would have to be extremely selfish and one's selfishness would have to be firmly entrenched against the effects of epistemic rationality for one not to have an obligation on my theory to favor the altruistic policy in spite of the higher taxes one would have to pay.

Voting is one of the main ways in which an ordinary citizen can support or fail to support a public policy. The chance that one's vote will prove decisive is extremely low but that does not affect the sacrifice/gain ratio because the chance of one's vote causing a sacrifice to be imposed on oneself would be exactly as low as the chance of that vote leading to a gain for mankind. Sumner and Anglin have objected that even if one's vote in a case like the wheat to India example is decisive, the gain to mankind for which one is responsible should be determined by dividing the total gain by the number of people voting for the policy, that my theory should only count the gain for which one individual is responsible in determining the sacrifice/gain ratio, and that this would yield a ratio that most people would find inadequate to merit their sacrifice. This criticism clearly assumes not only that what a utilitarian should care about is the amount of goodness or badness for which he is responsible but that if a number of people are equally responsible for a given outcome, each one is responsible for a fraction of that outcome equal to one over the number of people involved. Suppose you are faced with two alternatives: either you must, by yourself, confine a prisoner unjustly for two hours or else you must cooperate with 9,999 other people in confining him unjustly for 10,000 hours; and that, in the second case, although your contribution would be no greater than that of anyone else, it would still be essential. In my

critics' view, utilitarianism would tell you to take the second alternative because it would only make you responsible for $1/10,000 \times 10,000$ hours of confinement; where the first alternative would make you responsible for $1/1 \times 2$ hours of confinement. But this clearly is altogether alien to the spirit of utilitarianism. Furthermore, although utilitarianism is commonly described as holding that we should always do the thing that will have the best consequences, this interpretation fails also, I believe, to capture the true spirit of the doctrine. Surely Henry Sidgwick (to my mind the paradigm of the utilitarian) wouldn't have wanted us to leave out of our calculation events which, although they are not properly speaking *consequences* of our actions, would not have occurred if we had acted differently. Surely what we are to do if we want to act morally is to consider the various states of the world that could occur depending on what we decide to do and to choose the best of them. Put another way, what the utilitarian wants an agent to do is to pick the best of all contingently possible worlds available to him.

One of the main objections to an obligation theory is that it can't consistently give as a reason for not preventing the existence of future generations, that it is *prima facie* wrong to prevent the existence of happy people, without claiming as well that couples likely to have happy children usually have an obligation to have them. Sumner attempts to answer this by arguing that such couples could usually spend the time they would devote to rearing children in other ways which were at least equally beneficial to mankind, so that in terms of traditional utilitarianism they would not have an obligation to have children. But if many of these couples could in fact be *more* useful to mankind by *not* having children, a traditional utilitarian would be forced to say that they have an obligation not to have children; and this is counter-intuitive. My modification of utilitarianism gets me out of this awkward corner: it allows me to hold that it is usually beyond the call of duty—i.e., supererogatory—to benefit humanity either by having children you don't want or by not having children you do want. It would normally be commendable rather than obligatory to place the interest of society above your own interests in this sort of decision.

My meta-ethical theory has an enormous advantage in that whenever other theories and mine result in different prescriptions, the other theories tell you to do something that you would choose to do only if you were in some way epistemically irrational. And, unlike many subjecti-

vist theories, my ideal subjectivist theory allows ample room for the criticism of ethical judgments: it is easy to make mistakes about what you would choose to do if you were epistemically rational. However, it does allow for ultimate irresolvable moral disagreements, particularly on the sacrifice/gain ratio. One way to avoid this would be to set the ratio at a level such that it would not be too demanding for any epistemically rational person except a few whose selfishness was both extreme and immune to the effects of epistemic rationality. If the standard were set in this way, almost everyone would either be willing to make the required sacrifices or would only be unwilling because he was not epistemically rational. This alteration has two advantages: in the same circumstances, everyone would have the same obligations, and almost everyone could be given effective motivating reasons for doing at least as much for others as they had an obligation to do. Furthermore, optimistic actions would still always be morally acceptable; the theory would still come up with much the same demands as traditional utilitarianism in public matters; and prudential egoistic reasons could still be provided for moral behavior in private matters. There is the disadvantage, however, that there would no longer be the guarantee that you ought to do the thing that you would choose to do if you were epistemically rational. Except for exceptionally selfish persons, the altered theory would never demand greater sacrifices than one would be willing to make, but it would frequently require less sacrifice for the welfare of others than one would be willing to make if one were epistemically rational.

There is some similarity here to Hare's theory that ethical judgments are universal commands. Consider the matter of epistemic rationality: although Hare can't say that ethical judgments are properly speaking *mistaken*, he would presumably be prepared to say that if you issue a command that you would not have issued if you were epistemically rational, your command would be subject to criticism. Here, I think we are reasonably close. We both take obligation to be ultimately a function of attitude. But Hare's theory leads him into difficulty over supererogation. Complete benevolence is built into his theory: in order to make a satisfactory moral judgment, the interests of everyone affected must be considered, and the interests of others are to be counted as heavily as one's own.³¹ Consequently, his theory is bound (like an ideal observer theory) to require supererogatory good deeds.

As Hare recognizes himself,³² it is difficult to accept his meta-ethical view without supposing that we ought to behave like traditional utilitarians; and of course traditional utilitarians require supererogatory good deeds.³³

Some readers may object that even if there is a broad sense of "obligation" in terms of which obligations are a function of what you would choose to do in an epistemically ideal state, still whenever there is a conflict between obligations in this broad sense and obligations in a narrow sense (where they are purely objective), the latter takes precedence. If this were so it could be argued that, in the case of any obligation to add happy extra people, there would almost always be some purely objective obligation that would take precedence over it.

I have three objections to this:

1. There *are* no purely objective obligations. Factual considerations alone, independently of one's attitudes (including one's attitudes in an epistemically ideal state) cannot force one to accept an ethical judgment. Searle has offered what is perhaps the most plausible attempt to refute this claim.³⁴ He holds that (a) one can be forced by the facts to admit that a person has made a promise; and (b) that once one has made this admission, it is inconsistent to deny that he has at least a *prima facie* obligation to keep his promise. Though I agree with (b) I have argued elsewhere³⁵ that (a) is mistaken. To say that a person has made a promise is (1) to make the factual assertion that he has said certain words in a certain sort of situation and (2) to make the ethical presupposition that there is an institution that enables us to acquire obligations by saying such words in such situations. Because (2), the ethical part, is a presupposition, one can't challenge it by saying that that person didn't make the promise: to say that is to deny the objective part of the assertion which may well be correct. But one can admit the objective part without admitting that there is even a *prima facie* obligation if one is prepared to deny that the institution does anything more than make people think that they have acquired obligations. Such a denial would be an evaluation of the institution. Searle is prepared to say that evaluations may be purely objective but he can't use that claim as a premise in his argument, without circularity, because it is by that argument that he hopes to convince us that evaluations can be purely objective.
2. If ethical judgments and/or evaluations were purely objective, they

would simply be *shorthand* ways of conveying factual information; as such, they would deserve no more respect in decision-making than judgments that presented the same information using more words. For example, if "right" happened to mean "conducive to the maximization of the gross national product," that would not provide a reason for altering our attitude towards that curious objective.

3. Suppose that I admit that there are objective obligations and that my objective obligations conflict with what I would choose to do if I were epistemically rational. I can't think of any effective rational means of persuading me to be "objectively" moral so long as I am firmly convinced that I would not choose to be "objectively" moral if I were epistemically rational. In fact, it is an analytic truth that if I were epistemically rational, *any* attempt to persuade me to give precedence to "objectively" moral considerations would be bound to fail. Only if I were *not* epistemically rational, could I be persuaded to do something that I wouldn't do if I were epistemically rational. Someone trying to persuade me would be in a difficult position, to say the least, if he had to admit at the outset that he could only succeed if I were epistemically at fault. He would be rather like a car salesman who had to admit at the outset that if one knew all the relevant facts, one would buy his competitor's car.

What I've been trying to accomplish by rather a complicated route in this section is this: I have argued that there is a broad sense of "obligation" such that if one is glad that he has made a particular sacrifice (in my example, in order to create a happy race of people) and if one is given knowledge by acquaintance of the outcome of one's action (in my case, with those happy people), one would be prepared to do the same thing over again; if one's attitude didn't reflect logical confusion or factual ignorance, one would have had an *obligation* to have made the sacrifice in order to bring about that consequence (i.e., to have created that race of happy people). Suppose now that I am wrong and that there is no such sense of "obligation"—that I have, in effect, simply given a stipulative definition. It would still hold that one would have to be epistemically irrational not to do what one had an obligation to do in my stipulated sense; one would therefore have a strong reason for doing it rather than anything else that one might be obligated to do, given some other sense of "obligation." Thus if my original hypothesis as to how most of us would react to seeing the consequences of creating

a happy new race is correct and if our reactions do not reflect epistemic irrationality, most of us would have a strong reason for creating such a race. We would, of course, have an equally strong reason for *not preventing* the existence of such a race or for not preventing future generations if they would be equally happy. They won't be, but so long at least as we think that they will have more happiness than unhappiness we have *some reason* for not preventing their existence. But this is exactly what those who oppose the obligation view wish to deny.³⁶

Notes

1. A third objection (which I don't take up here) is that the obligation theory is incompatible with the commonly held view that normally it is morally permissible for couples either to have children or not as they like. In §10, I develop a meta-ethical position designed to support the obligation theory. This position requires a revision in traditional utilitarianism, a revision which allows me to hold that although we have an obligation to preserve mankind, typical couples have neither an obligation to have children nor not to have them.

2. Rawls is the main philosopher whose name has been associated with average utilitarianism. He is not an average utilitarian himself, but he does regard it as decidedly preferable to classical utilitarianism. (*A Theory of Justice*, Cambridge, Mass.: Belknap/Harvard University Press, 1971, pp. 161-165.)

3. Some philosophers claim that even rough cardinal comparisons of pleasure and pain are either meaningless or totally unjustifiable. But this would prevent us from saying many things that are (I think) obviously true. Take the following. "Some prisoners who had at least some pleasure during their stay at Dachau still had far more pain than pleasure while they were there." This judgment clearly presupposes cardinal comparisons of pleasure and pain but we can nonetheless be morally certain that it is correct.

Or suppose that Bertha has a mild headache from 9 A.M. till 9 P.M. while her husband Sam is perfectly comfortable from 9 A.M. until 8 P.M. when he is picked up by the secret police and tortured for an hour. We would all suppose that Sam suffered much more pain than Bertha during the twelve hours, something we are clearly not entitled to suppose unless we believe that Sam's pain from the torture was more than twelve times as severe as Bertha's mild headache pain.

Two other points should be made: (1) It is often the case that my opponents must also assume in their objections—this is true, e.g., of the repugnant conclusion—that cardinal comparisons of utility and disutility are possible at least in theory. (2) If the notion of a life that is neutral in that it contains equal amounts of pleasure and pain (or some other utilities and disutilities) is found objectionable, a stretch of a person's life could be described as neutral iff he would, if he were epistemically rational, be indifferent between living through it again or being unconscious during that stretch of time.

4. Not only is extra population likely to lower the total amount of happiness both for those now alive and for future generations (if there are any), but it could lead to the destruction of all mankind through the more dangerous forms of pollution or through a nuclear war brought on by competition for increasingly scarce resources.

5. Mary Warren contends ("Do Potential People Have Moral Rights?" in this volume) that not only do we not have any obligations relating to possible persons who have not yet been conceived but that even after conception, we have no obligations to the fetus. She claims that although the fetus has the potential for having experiences, this is not enough to give us even a *prima facie* obligation to help realize this potential, or even a *prima facie* obligation to refrain from abortion, i.e., from preventing its realization. It has been argued against her view that an unconscious person also has only a potential for having experiences but that in the case of an unconscious person this is enough to give us *prima facie* obligations towards him and that consequently the same should hold for fetuses when they have only the potential for having experiences.

One possible response to this objection (which I will consider later) is that taken by Engelhardt ("The Ontology of Abortion," *Ethics*, Vol. 84, 1974.), that the crucial moral difference between an unconscious person and a fetus is that the unconscious person has a *history of past experiences* while the fetus does not (what I will call the "history requirement"). Warren argues rather that the crucial difference is that between something which has an "actual and present capacity" for consciousness, and something which has a capacity for consciousness that is not actual and present (what I will call a "plain capacity"). Warren argues that since an unconscious person has a present and actual capacity for consciousness, we may have an obligation to make him, or to allow him to become, conscious; whereas if something—like a fetus—has only a plain capacity for consciousness, we cannot have any such obligation.

The question is what exactly the distinction is between these two kinds of capacity. I shall consider a number of possible interpretations of her distinction, none of which will serve her purpose:

(1) The requirement that the capacity be *present* could mean that in order for something to have this "higher" sort of capacity for consciousness, it must be likely that if it is left alone it will become conscious or that it can be made conscious within a short time. But if this is the crucial difference between the two sorts of capacity and if it is accepted that if something lacks the higher sort of capacity we can have no obligation to make it conscious, we get the awkward result that if someone were in a coma from which he could only be roused after a long period of time we would have no obligation to rouse him.

(2) The requirement that the capacity for consciousness be actual could be construed as the requirement that it not merely be a capacity to *acquire* that capacity but that it be that capacity itself. But this also leads to counterintuitive results when it is applied, for example, to a person in a coma.

(3) For something to have a present and actual capacity for consciousness, it must be the case that relatively little needs to be done to make him conscious. This interpretation would again have an unfortunate consequence for someone in a coma: we would have no obligation to bring him out of it if doing so would require a good deal of work.

(4) If something is of the same order of complexity as a human being, it may have a present and actual capacity for consciousness; otherwise not. This would allow us to have an obligation to animate robots provided they were sufficiently complex and it was possible to do so. Warren would clearly find this unacceptable and furthermore, it is hard to see why *complexity* should be morally relevant here.

(5) In order for something to have a present and actual potentiality for consciousness, it must resemble conscious human beings in all physical respects (or this category could be broadened to include animals as well) but to distinguish this from the history requirement, it need never have been conscious. But this would have the unacceptable consequence that if an unconscious creature had a history of consciousness but didn't have the specified physical characteristics, we wouldn't even have a *prima facie* obligation to allow it to become conscious again.

Back to the history requirement. Suppose that we had the alternative of either restoring to life a long-forgotten person—say, the maternal grandfather of the caveman who painted the bison pictures in the Lescaux caves—or of creating someone who has never lived before who

would be much happier than the caveman (who would be only moderately happy being restored to life in the contemporary world). Suppose, further, that the expected effects on other people choosing either alternative would be roughly equal. Using the history requirement would force us to say that it would be wrong to create the new person rather than the caveman despite the fact that the new person would be far happier. This, I think, would go against most people's moral intuitions.

I have chosen a long-forgotten person as my example rather than, say, one's own maternal grandfather because our intuitions in the caveman case are not influenced by factors that usually (but need not) accompany a history of consciousness (such as gratitude for good deeds the person might have done, obligations created either by implicit or explicit promises to preserve the person's life, or the obligations which one might naturally feel toward friends or relatives). Considerations such as these are highly relevant, of course, and they bring me in the end relatively close to Warren on the actual rights and wrongs of abortion. Although I think that it is *prima facie* wrong to destroy a fetus, I also believe that far more weighty considerations are needed to justify the destruction of an actual person than to justify an abortion.

6. "A Utilitarian Population Principle," in *Ethics & Population*, edited by Michael Bayles (Cambridge, Mass.: Schenkman, 1976), pp. 81-99.

7. There is still another objection to attempts to save the ontological preference theory by adding the occupants of inevitable slots to the ontologically privileged. Though there is some intuitive appeal in saying that the welfare of inevitable persons is more important than that of possible persons, there is little intuitive appeal in the corresponding claim about the occupants of inevitable slots, particularly when the notion of an inevitable slot is attenuated as it must be to get anything like the right number of occupants of these slots. And there is another matter that I have passed by rather hastily before: the class of persons whose existence is to be regarded as inevitable from the point of view of the population planner should obviously not be made up of those who will exist whatever he does regardless of whether it is morally acceptable or not, but rather of those who will exist regardless of which *morally acceptable* thing he does. Thus the category of inevitable persons is not purely ontological but ethical as well, which muddies the water still more.

8. *Ibid*, p. 89.

9. My objections are addressed to this modification of Singer's principle. His principle clearly needs *some* modification, and this strikes me as the most reasonable move he could make. Singer is aware of the change I have made and has not objected to it.

10. "Utilitarianism: The Classical Principle and the Average Principle," *Canadian Journal of Philosophy*, Volume V, No. 3, November 1975, pp. 409-419.

11. Derek Parfit, "Rights, Interests & Possible People," partially reprinted in Samuel Gorovitz *et al.*, eds., *Moral Problems in Medicine* (Englewood Cliffs, N. J.: Prentice-Hall, 1976), pp. 369-375, at pp. 373-5.

12. The timeless impartiality principle is also of course incompatible with ontological preference theories because they tell us that in adding extra persons we should weigh the welfare of the added persons less heavily than that of the ontologically privileged.

13. Timeless utilitarianism on the other hand only requires that, in adding people, the particular people added should be made as happy as possible. (ii) and (iv) deal with the question of who should be added, once it is granted that someone is going to be added. Timeless utilitarianism ignores this question entirely and bears instead on what should be done once it is decided that some particular person is to be added.

14. One objection to this argument (suggested to me by A. J. Ayer) is that it is always wrong taken in itself to bring about the existence of any group of people regardless of how many of them will be happy if *any* of them will be wretched or perhaps even very unhappy. But this won't do because it entails the absurd conclusion that it would always be wrong *per se* to bring *any* large group of people into existence, even in a world with enormously better prospects for happiness than ours—that it can only be justified as something which, though bad in itself, is unfortunately needed to satisfy the then selfish desires of most people for children.

15. This particular argument for an obligation theory may be confused with others which rely on the claim that there is no plausible principle that can provide a rationale for saying that we have obligations relating to possible wretched people but not relating to possible *happy* people. I think that the following principle would provide such a rationale, and it is not obviously false: *an action is prima facie wrong iff it harms someone or someone exists who would have been benefited by some alternative action*. This principle implicitly denies that it is morally important to bring happy people into existence. A possible

happy person whom you haven't caused to exist *doesn't exist* so your failing to cause him to exist won't be wrong even if it would have been a benefit to him if you had caused him to exist. This is a variation of the principle Parfit attacks with his birth-timing principle (*ibid.*, pp. 373-375).

16. "Repugnant Conclusion," forthcoming in the *Canadian Journal of Philosophy*.

17. "Utilitarianism, Supererogation and Future Generations," forthcoming in the *Canadian Journal of Philosophy*.

18. Bennett's first objection is that we can account for its being permissible to add a group with some miserable people in it without relying on the counterbalancing effect of the happiness of the happy people in it on the grounds that phasing out mankind even over a 150-year period would "involve ghastly horrors; and the disutilities also include events which, though not intrinsically dreadful, serve to frustrate hopes, abort plans, disappoint expectations, and so on (as when a happy man suffers a disutility by dying quietly in his sleep)." This sort of objection was relevant to the argument I gave in my earlier article ("Utilitarianism: The Classical Principle and the Average Principle"), because I neglected to specify there that the good and bad effects on persons other than those being added were to be roughly equal. (I had meant to take care of that problem by specifying that there was no population problem in the world in question.) But Bennett's objection does not apply to the version of the argument in this paper.

Bennett's last objection is that he has a rule that provides a "clear coherent basis" for the balancing role of the happy people as against the miserable ones, but which does not imply that we have any obligation not to prevent the existence of happy people; namely, that "facts about personal utilities are relevant to moral obligations only through arguments of the form, 'It is obligatory to do A because if A is not done then there will be people who . . .'" Clearly, Bennett means to complete his rule by some such expression as "will be worse off than they would have been if A had not been done." When so completed, this rule is unacceptable because it violates even the weakest form of Parfit's baby-timing principle. The rule implies that, other things being equal, a prospective mother would have no obligation whatever to make any sacrifice, no matter how small, in the timing of a child's conception, even if this sacrifice would result in a *very* happy baby rather than resulting in another baby who because of birth defects would have a far less happy life, as long as this second possible baby would have a life that was worth living. It has this implication because,

given my case, even if she conceives at the less propitious time there will be no person who is less well off than if she had conceived at the more propitious time. The happier baby is the only conceivable candidate, and he or she won't exist.

Bennett might try to avoid this by allowing Parfit's principle as an exception to his rule or, what amounts to the same thing, by filling in the blank in a way compatible with Parfit's principle. But it is difficult to reconcile Parfit's principle with an anti-obligation view. (See §8.)

Besides the fact that Bennett's rule generates counterintuitive consequences, the thing he is attempting to do by means of this rule—balancing the creation of unhappy people with that of happy ones while at the same time not allowing *any* positive value to the creation of happy people—seems basically misguided. Why should the bringing-into-existence of happy people be of great moral importance when they are part of a package along with some unhappy people but of *no moral importance whatever* when they are all by themselves?

Let's imagine that there is one person left in the world (let's call him Z), who shares Bennett's interest in the completion of certain human projects. He has arranged to re-populate the world with test-tube babies but he discovers that although they are likely to complete the projects the rest of humanity has left unfinished, there are more unhappy babies than happy ones. As it stands, Bennett would probably agree that Z had done something for which he is not blameworthy but which nevertheless is objectively wrong. However, he would, I should think, suppose that Z could rectify the situation if he could make enough extra happy babies so that there would be a high enough ratio of happy babies to unhappy ones, and that if it were within his power, Z would have an obligation (even at great sacrifice to himself) to make enough happy babies to achieve that ratio. But let's suppose that after Z had made enough happy babies to reach that ratio, he could go on at relatively little cost to himself (because by now he has the hang of the thing) to make a great many more, even happier, babies. Bennett would have to say that while Z would have a pressing obligation to make the first batch of extra babies, he would have *none whatever* to make any more, no matter how happy they would be and no matter how easy it would be for him to do it; and Bennett would have to account for the difference solely in terms of the existence of the unhappy babies. This seems counterintuitive. It implies that if I want to see how many happy people I have an obligation to bring into existence, the thing to do is not to see how happy they would be or how much sacrifice it would cost me but rather how many unhappy people

I have already brought into existence. And if it is urged that that is the right thing to look at if I want to find out at what point creating happy people would be supererogatory rather than obligatory, I must refer my reader to my discussion of this point in "Utilitarianism, Supererogation, and Future Generations" and to that part of §10 dealing with supererogation. Although my account is primarily directed to utilitarians and Bennett is no longer a utilitarian, I suspect that the difference between utilitarianism and his version of consequentialism is, for present purposes, of no real significance.

19. "On Maximizing Happiness", in this volume.

20. To be fair, I should point out that Bennett introduces this notion in a context in which it would be reasonable to expect that it (or similar interests) would be shared by a great many other people so that their combined interests could constitute a powerful reason for preventing the destruction of the human race. Like Kavka, I think that while this is a good reason for not preventing the existence of future generations, another good reason is that it is *prima facie* wrong to prevent the existence of happy people; though, unlike him, I think that arguments *can* be given for that principle. (One of those arguments, incidentally, is not that possible people are in some odd sense waiting in the wings.)

21. The projects which these interests fulfill could have practical application (such as medical theories) or not (such as the completion of some philosophical or mathematical theories).

22. See *Working* by Studs Terkel (New York: Pantheon/Random House, 1972) for an account of how people in a great variety of jobs feel about their work.

23. Even if an interest theory could be made compatible with a version of (2a) that a classical utilitarian would especially want to avoid, this would not, I think, be a serious problem for the classical position. Any theory which assigns intrinsic value to the fulfillment of each and every interest is committed to ascribing it not only to states of consciousness but to physical objects as well, and surely this is a mistake. (This in spite of Nozick's arguments to the contrary in *Anarchy, State & Utopia* [New York: Basic Books, 1974]: Part I, Chapter 3, pp. 42-45.) If on the other hand one ascribes intrinsic value only to the fulfillment of interests directed towards states of consciousness, the best method of measurement would be essentially similar to that I proposed for ideal utilitarianism, a method which offers little support for the extremely optimistic view of existence needed to justify a high ratio in (2a).

27. Singer, *ibid.*

28. It could be objected that if my ethical theory were correct, the grammar of moral discourse should be such that if two people disagree on whether a given action is obligatory, they could conceivably both be right because each is using "obligatory" relative to his own ideal attitude whereas in practice we act as if "obligatory" is not a relativistic term. First, this is only roughly true of actual practice, and secondly, even if the ultimate test for moral judgments is the speaker's own ideal attitudes, so long as they are cognitively ideal attitudes there is more than enough room for mistakes as to what those attitudes would be to account for the impression that all differences in ethical judgments reflect cognitive errors.

I could, however, afford to retreat to a position like Mackie's ("A Refutation of Morals," *The Australasian Journal of Philosophy*, 1946) in terms of which, although the ultimate test for ethical judgments in my version is attitudinal, to say that an action is obligatory is to ascribe a distinctively ethical quality to it, although in fact there are no such qualities. Since all I am concerned with is the question of what the ultimate test is, the difference between my theory and a Mackiean counterpart is not, for my purposes here, of any importance.

29. Although my theory deviates from classical utilitarianism in allowing us to discount the interests of others, the discounting has nothing to do with their ontological status.

30. I am assuming here that a person who has become epistemically rational would want to do what he would choose to do if he had the attitudes that he would have had if he had the same hereditary make-up but had always been epistemically rational.

31. Cf. *Freedom and Reason*, (Oxford: Clarendon Press, 1963), Chapter 6.

32. *Ibid*, Chapter 7.

33. It is possible to construct another sort of universal imperative theory of ethical judgments that could lead to supererogatory demands. Such a theory would stipulate that: (1) the agent be thoroughly egoistic but (2) in order to insure impartiality in making his universal demands, he is supposed to imagine that he is ignorant of his own particular situation; and (3) the agent is supposed to believe (or rather, to pretend to believe) that his command will be universally accepted. In these circumstances, the egoistic agent would probably find that he would maximize his own chances for happiness by commanding everyone to be a utilitarian. He would run only a very slight risk of having to make major sacrifices for others because if everyone complied with his command and acted as a good utilitarian, situations would seldom arise in which extreme altruistic behaviour would be required. Thus this sort

of prescriptivism would also seem to lead to traditional utilitarianism and to the supererogation that goes with it.

It might seem that any universal command theory of ethical judgments necessarily leads to problems with supererogation but this is not the case. It could be claimed that obligation-statements are universal commands and that such a command is open to criticism if the agent issues it only because he is in some respect epistemically irrational, is not prepared to obey it, or is not prepared to accept (though not necessarily to *like*) compliance with it on the part of others. If the language were used in this way, one would expect commands in which we would sometimes be required to make sacrifices but only when those others would gain proportionally more than we would lose. (Situations in which one is bound by a promise to help another person are an apparent exception but the effect of the promise would not be to alter the ratio at which one would be required to make sacrifices for others; it would rather be to alter the situation in such a way as to provide one with additional reasons for doing the thing in question. Failing to do so would be likely to have a detrimental effect on one's own reputation and character; it would frustrate the legitimate expectations of the other persons, and it would have a minute effect on the extraordinarily important institution of promising.) This would eliminate supererogatory demands. It also has the consequence that we would not be entitled to expect the complete benevolence from others that we would like: it would commit us to being prepared to accept without protest treatment from others which was only as good as we would be willing to give them in corresponding circumstances.

One trouble with the other two sorts of prescriptivism is that it is difficult to see how one could persuade an epistemically rational person to issue the respective sorts of command unless he were completely benevolent or really believed that everyone else would follow his commands. On this prescriptivist counterpart of my theory, neither of these would be required because the individual is only commanding himself to make those sacrifices that he would choose to make if he were epistemically rational.

34. "How to Derive 'Ought' from 'Is'," *Philosophical Review*, 73, 1964, pp. 49-58.

35. "Facts, Promising and Obligation," *Philosophy*, 50, 1975, pp. 352-355.

36. I am indebted for helpful discussion and criticism to my co-editor Brian Barry and to Jonathan Bennett, John Black, D. G. Brown, Howard Jackson, Peter Singer, Brian Stouffel, and David Zimmerman.