

Parfit's Moral Arithmetic and the Obligation to Obey the Law

GEORGE KLOSKO
University of Virginia
Charlottesville, VA 22901
U.S.A.

Though consequentialist theories of political obligation have been widely criticized in recent years, a series of arguments presented by Derek Parfit, in *Reasons and Persons*,¹ are now believed to have given this position new life. Parfit's presentation, which is intended to correct 'five mistakes in moral mathematics' (*RP*, 66), can be extended to account for the most central of our political obligations, the obligation to obey the law.² This paper takes a closer look at Parfit's moral arithmetic and attempts to demonstrate the continuing difficulties of consequentialist theories of political obligation.

I PD Problems

Before we turn to Parfit's arguments, a few preliminary matters must be touched upon. For ease of discussion, I will concentrate on one particular theory, act-utilitarianism (AU), which is the most familiar form of consequentialism. With relatively slight modifications, my discussion could be made to apply to other forms of consequentialism as well. Act-utilitarians believe that a given act is right if no other act could be performed under the circumstances that would have better conse-

1 D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press 1984); this work cited hereafter as *RP*, in the text.

2 C. McMahon, 'Autonomy and Authority,' *Philosophy and Public Affairs* 16 (1987), 315-19; see below, Section III; also A.D.M. Walker, 'Political Obligation and the Argument from Gratitude,' *Philosophy and Public Affairs* 17 (1988), 206-7.

quences.³ The main problem with AU theories of political obligation is that in many important cases it appears that disobeying the law would have better consequences than obeying it. These cases are often analyzed in terms of the 'prisoner's dilemma' (PD).⁴ One of a citizen's main obligations is paying his taxes. But if Greene is a member of a large society with an annual budget in the hundreds of billions of dollars (that of the U.S. is currently more than one trillion dollars), his few thousand dollars would make no perceptible difference. If he did not pay, his payment would not be missed, but he would have more money to spend on his family and friends, perhaps to donate to worthy causes. Thus it could be argued that more good would come from his not paying taxes than from paying them. Similarly, assume that Greene's city (X) is afflicted by drought and so various conservation measures are enacted, including a law requiring that citizens wash dishes no more than once a day. But if X is sufficiently large, the few extra gallons of water that Greene consumed with extra washings would not be missed. Because Greene is uncomfortable when he has dirty dishes in the sink, and the washing restrictions would increase the risk that he and his family would become ill, again it could be argued that more good would be created if Greene violated the restriction than if he adhered to it. Similar cases could be multiplied; additional cases are examined below.

The cases that are of greatest interest to political philosophers have the following features.

(a) They concern the provision of benefits that are generally characterized as 'public goods.' For our purposes, the main feature of public goods is 'non-excludability': if a given public good is provided to certain members of some community it is either prohibitively difficult or impossible to deny it to other members.⁵ Prime examples of public

3 This is a comparative, as opposed to an absolute, form of AU; I believe it is the stronger doctrine and so more worthy of attention. Reasons of space prevent discussion of this and other complex issues in utilitarianism. For discussion of the consequences of actions, see below, Section V.

4 Stimulating recent discussions of PD problems are Parfit, *Reasons and Persons*, Part I; and E. Ullmann-Margalit, *The Emergence of Norms* (Oxford: Oxford University Press 1977), Chap. II. It should be noted that, as D. Regan points out (*Utilitarianism and Cooperation* [Oxford: Oxford University Press 1980], 62-3) the cases discussed throughout this paper are not PD cases in the technical sense. In PD cases the different participants have different maximands; in the cases under discussion we can assume common interest.

5 As the term is generally used, 'public goods' are characterized by non-excludability and 'non-rival consumption,' i.e., that one individual's consumption of a given good will not affect the amount available for others. The latter is not directly relevant to the concerns of this paper and so is not discussed.

goods are national defense, the rule of law, and protection from environmental hazards, all of which are frequently provided by governments.

(b) Provision of the benefits under discussion requires large scale but not universal cooperation. If the benefits noted in (a) are to be provided in some society, the efforts of large numbers of people will be necessary. But provision of the benefits is not incompatible with the failure of a certain number to cooperate.

(c) The burdens that any individual must bear to provide the goods in question (and the share of the goods that he receives) do not perceptibly affect the burdens and benefits of other individuals. This condition can be referred to as 'independence,' in that the burdens and benefits of any one individual are independent of those of others. In the cases that interest us independence is insured by the large size of the societies. In the tax case, for instance, the fact that Greene does not pay neither requires that the payments of other individuals be increased nor forces government to cut its services. Of course Greene's action could also have indirect effects, e.g., if it became widely known that he was not paying his taxes and his example inspired numerous others to behave similarly. Under these circumstances, his non-payment could affect other individuals. But there are many cases in which an individual could avoid some burden of citizenship without attracting such attention. Throughout this paper we will concentrate on cases in which this condition can be assumed to hold.

(d) In the cases that interest us the contributions required for the provision of public goods are generally viewed as burdensome.

(e) The public goods are worth their costs. The benefits that any individual receives from general cooperation outweigh the costs of his own cooperation, while society as a whole also benefits. If the goods were not provided all would be worse off.

If conditions (a) – (d) exist, then even if (e) holds and the benefit Greene receives from some public good outweighs his costs, his interest could be said to lie in non-cooperation. In the conservation case, because of (a) Greene will receive the benefits of general conservation whether or not he cooperates. Because of (b) and (c) his failure to conserve water will not affect the likelihood of the benefit's being provided. Because cooperation is burdensome (d), he would prefer not to cooperate. Under these circumstances, Greene will receive the benefits of general conservation as long as a sufficient number (N) of his fellow citizens cooperate. There are two possibilities: either N others will cooperate or they will not. In the former case, Greene will receive the benefits whether or not he cooperates, and so his self-interest favors non-compliance. If, on the other hand, N others will not cooperate, the benefits will not be provided whether or not Greene does. In this

case cooperation would be foolish. If all the X-ites are similarly self-interested and follow similar reasoning, they too will decide not to cooperate, with (e) possibly disastrous results for X as a whole.

PD-logic does not require that individuals be self-interested. We will assume that the individuals discussed throughout this paper are act-utilitarians and so wish to bring about the greatest good for society as a whole. On this assumption, too, because Greene's contribution is burdensome and has only imperceptible effects, he could easily conclude that his non-compliance is best for society. Again, if his society is populated by similarly rational act-utilitarians who follow similar lines of thought, important public goods will not be provided and all will be worse off. These cases, moreover, tell against AU as a moral theory. Because our moral intuitions strongly indicate that Greene should cooperate, that his failure to do so would be unfair,⁶ AU appears to recommend against performing actions that we believe to be right.

II Parfit's Moral Arithmetic: SC Cases

For AU the difficulty in PD cases is justifying an individual's obligation to cooperate when his cooperation will not have perceptible effects. Arguing against a number of 'mistakes in moral mathematics,' Parfit notes two kinds of cases in which imperceptible effects can make actions wrong.⁷ Mistake #3 is 'to ignore very small chances when they would either affect very many people, or would be taken very many times' (RP, 75). The fourth and fifth mistakes are 'to ignore *very small* and *imperceptible* effects on very large numbers of people' (RP, 75; his italics). Cases that fall under the third mistake can be termed small chance (SC) cases; cases that fall under the fourth and fifth mistakes are similar and can be discussed together as small effect (SE) cases.⁸ The most important SE cases for our purposes concern the obligation to obey the law.

The logic of SC cases is clear. As Parfit notes, the third mistake is often made in discussions of elections, where there are large numbers

6 This conclusion follows from the principle of fairness; see G. Klosko, 'Presumptive Benefit, Fairness, and Political Obligation,' *Philosophy and Public Affairs* 16 (1987) 241-59.

7 For the sake of simplicity, throughout this paper I discuss only what makes actions wrong. Only verbal changes would be required to expand the discussion to encompass what makes actions right.

8 In subsequent discussion, I generally ignore slight effects, discussing only imperceptible effects. But this has little bearing on my arguments.

of voters (RP, 73). Grey can argue that she should not bother to vote in a Presidential election; if roughly one hundred million people are likely to vote (the vote in the 1984 U.S. Presidential election was about 93,000,000), the chances of her vote determining the outcome are infinitesimal. But if an SC case involves substantial stakes or affects very many people, Grey's reasoning can be shown to be wrong. We can assume that although voting entails some trouble for Grey, this is not a great deal. We can also assume that there are substantial differences between the candidates A and B, that A would be a much better President. According to Parfit, the figures involved are as follows (74):

$$\frac{\text{The average net benefit to Americans from A's election} \times \text{The number of Americans} - \text{The costs to Grey and others of her voting}}{\text{one hundred million}}$$

In many cases the outcome will be positive and so in favor of voting.⁹ Over the four years of his reign President A will make numerous decisions that will greatly affect the entire population – indeed, all the inhabitants of the world. If the additional effects on non-Americans were factored in, the argument would be strengthened. Thus Parfit's analysis of SC cases helps to clear up important problems that beset AU.

III SE Cases

SC cases turn on Pascalian logic. According to Pascal's wager, the consequences of God's existence are so large as to justify faith, even if the chances of His existence are extremely small. The consequences of A's or B's winning the election are also large, and so voting is justified. An additional example Parfit presents concerns a nuclear engineer (RP, 74-5). By neglecting some precaution, the engineer could very slightly increase the chances of a catastrophic accident. Because the consequences would be so grave, in most cases he should take the extra trouble. However, Parfit's analysis of SC cases is not readily applicable to general questions of political obligation. The general obligation to obey the law should be construed as an SE rather than an SC case, and so is not amenable to Pascalian logic.

⁹ Parfit is anticipated here by B. Barry, *Political Argument* (London: Routledge and Kegan Paul 1965), 328-30.

SC and SE cases are sharply distinct in two respects. In discussions of consequentialism, the point at which an agent's actions pass from having imperceptible to perceptible effects, or harmless to harmful effects, is generally referred to as a 'threshold.'¹⁰ We can use the term in a fairly broad sense, for any point at which a significantly different state of affairs (in terms of harms or goods) is brought about by some action[s]. SC and SE cases involve (i) different sorts of thresholds, and (ii) differences in how their respective thresholds are attained.

SC cases are characterized by what we can call (i) 'absolute thresholds.' In these cases there are severe differences between possible outcomes: either A or B wins the Presidential election, which could be decided by a single vote; either there is a nuclear accident or there is not. There are no series of gradations between the non-existence and existence of the relevant consequences. As for (ii), the thresholds present in such cases can be crossed by the particular acts of particular individuals. Though one hundred million people can vote in an election, the outcome could conceivably be different if there were one vote more or less.¹¹ The chances are of course very slight, but the possibility exists. The fact that a single action of a single individual can trigger an absolute threshold is crucial to the Pascalian logic of SC cases. The fact that SE cases contain neither this sort of threshold nor the possibility of reaching it in this way makes their moral mathematics more complex.

The differences between SC and SE cases can be seen by looking at the recent discussion of Christopher McMahon. McMahon does not adequately distinguish the two, and so reveals some problems that can arise in applying SC logic to questions of political obligation.¹² McMahon approaches political obligation through Parfit's analysis of SC cases (AA, 315). He begins with the familiar example of deciding whether

10 On thresholds, see D. Lyons, *Forms and Limits of Utilitarianism* (Oxford: Oxford University Press 1965), 63-91.

11 It should be noted that the results of an election are susceptible to analysis in terms of SE logic as well. In addition to who wins or loses the size of the win, e.g., whether or not it is a landslide, can have significant implications for a country's future, while in a large society the vote of a single individual will make an extremely slight contribution to the outcome, either way. Thus it could be argued that absolute thresholds are less absolute than they appear. Though this is true, however, it has little bearing upon the argument of this paper. Though *aspects* of elections can be analyzed in SE terms, elections differ from SE cases in having other significant aspects that can be analyzed in SC terms. It is upon these aspects Parfit focuses; for present purposes other aspects can be passed over.

12 McMahon, 'Autonomy and Authority,' cited hereafter as 'AA,' in the text. For our purposes it is not necessary to discuss the wider context in which McMahon's discussion of political obligation is set.

to cross a beautiful lawn. Though the effects of Haddock's crossing would be imperceptible and he would save time and so benefit from crossing, he should not do so. There is a small chance that he is on the threshold 'where crossings are so frequent that the lawn is not able to repair the damage done by a particular crossing' (AA, 315). If Haddock multiplies the small chance that he is on the threshold by the large potential loss of aesthetic satisfaction to many people that would result if the lawn were ruined, he could well conclude that he should not cut across it. McMahon finds confirmation of this line of reasoning in Parfit's examination of whether to vote in a U.S. Presidential election, as discussed above (AA, 317-18). His analysis of these two cases is then employed to argue that a rational anarchist should obey the law, so as not to bring about the descent of his society into a state of nature. In McMahon's words:

[O]n even the most optimistic assumptions concerning life in a state of nature, the anarchist should judge the average net gain from avoiding the state of nature to be considerably greater than that from electing the better of two (typical) presidential candidates. Thus the probabilities that figure in Parfit's example seem approximately right to justify an American who is a rational altruist in obeying the law. Even if the chance of being on the threshold where cooperation to avoid the state of nature collapses is as little as one in one hundred million, obeying the law will still be the indicated course. (AA, 318)

McMahon notes that assigning a numerical value to being on the threshold of social collapse is purely speculative, but he concludes: 'Parfit seems to have shown that cases of this sort are more amenable to consequentialist reasoning than might have been supposed' (AA, 319).

As we have seen, McMahon analyzes grass crossing and political obligation according to the logic of the election (an SC case). However, the former are SE cases. An election involves (i) an absolute threshold, (ii) the attainment of which can turn upon a single act. Circumstances are different in the other two cases. The thresholds in SE cases can be termed 'step thresholds.' Exactly how a lawn is damaged by repeated crossings is a factual question, and doubtless varies from lawn to lawn. But we can assume that as people repeatedly cross lawn L, the degree to which it is damaged will gradually increase from imperceptible effects to a point at which it is completely ruined, with all grass gone. In a case such as this we can posit a series of points along the line between L's perfect and ruined conditions. Since each point will be reached by the crossings of a certain number of people, *each of these different points* can be viewed as a step threshold. It is to a certain extent arbitrary where we draw the lines between different degrees of damage, and so different step thresholds, but the differences in damage

represented by successive step thresholds could well be relatively small. In addition, if we assume that a single crossing by a given individual, Haddock, will have imperceptible effects and be independent of the actions of other individuals, by himself Haddock will not be able to cause even a single step threshold to be reached. In order to damage the lawn, he must act in conjunction with a number of other individuals. There is accordingly *no* chance that he will cause the lawn to be ruined, or even to be perceptibly damaged. If N others cross along with him the lawn will be visibly damaged, and if $2N$ or $3N$ or however many others that it takes cross with him it will be completely ruined. But in a case such as this, the effects in question depend almost entirely on the actions of these others; they would be imperceptibly less severe were Haddock not to cross. Because Haddock's crossing does imperceptible damage and is assumed to be independent of the others' actions, he cannot be said to cause the attainment of any particular step threshold, or the lawn's demise. Though he undoubtedly bears some responsibility for ruining the lawn, there are difficulties in assessing this responsibility in consequentialist terms, a problem to which we will return in Section V.

Exactly how the rule of law in society X breaks down, from a situation of complete peace and harmony to a Hobbesian state of nature, is again a factual question, which will also vary between societies.¹³ But the situation would generally be far more similar to the lawn case than to the election case. Again we can assume a lengthy course of disintegration, the identification of certain points along which as step thresholds is to some extent arbitrary, and that the differences in damage represented by different step thresholds can be relatively slight. As in the lawn case, any one individual's role in the attainment of step thresholds is sharply different from that in SC cases. If X is a large society, Haddock's disobedience will have imperceptible effects and

13 It is worth noting that, if a consequentialist theorist wishes to argue from complex and/or long range effects of actions, then it is up to him to present convincing accounts of these. The fact that in many cases detailed descriptions of long range effects are highly speculative tells strongly against consequentialism, when it is extended beyond intuitively clear and simple cases. One possible way of dealing with apparently imperceptible consequences is 'probabilistic risk assessment,' as suggested by K. Shrader-Frechette, 'Parfit and Mistakes in Moral Mathematics,' *Ethics* 98 (1987) 50-60. According to Shrader-Frechette: 'if one looks at allegedly imperceptible harms with fine enough medical and scientific know-how and instrumentation, then it is questionable whether there are any genuine effects of nonmental acts which are imperceptible' (60). Even if this claim is true, however, the practical difficulties in producing convincing figures in many of the cases that interest us severely lessen the value of the approach.

be independent of the actions of other individuals. Thus there is *no* chance that his disobedience will bring about the attainment of a single step threshold, let alone the collapse of society.¹⁴ In conjunction with the actions of *N* others, his action could well cause a step threshold to be reached. As the number of these disobedient others increases, additional step thresholds will be crossed, until a war of all against all results. But again, because social disintegration is caused almost entirely by the actions of the others, whose behavior is independent of Haddock's, and conditions would be only imperceptibly less severe were he to obey rather than disobey, Haddock cannot be held responsible for the attainment of any particular step threshold or for society's collapse. The assessment of his responsibility here will clearly resemble that in the lawn case and will be discussed below.

It is clear, then, that McMahan is not justified in applying the Pascalian logic of SC cases to SE cases. Consequentialist arguments that justify political obligation must be very different.

14 This is not to deny that specific acts of disobeying the law can differ in their consequences, or that certain acts could well stimulate large numbers of other individuals to behave similarly, thereby leading to social collapse. The disobedient acts that most clearly fall into this category are public and dramatic. Examples are acts of civil disobedience by well-known figures, e.g., Martin Luther King, or flouting of the law by highly visible people in positions of authority, e.g., the behavior of Richard Nixon during the Watergate years. (For a discussion of 'stimulatory' actions see, J. Narveson, 'Utilitarianism, Group Actions, and Coordination,' *Nous* 10 [1976] 173-94, esp. 191ff.; and J. Glover, 'It Makes No Difference Whether or Not I Do It,' *Proceedings of the Aristotelian Society, Suppl.* 49 [1975] 171-90.)

Along similar lines, dramatic crimes, such as the bank robberies of John Dillinger, can inspire numerous imitators, or a riot can be set off by a single act of flagrant disobedience under tense conditions, such as an inflamed situation of racial conflict. Under these and other, similar circumstances, the SC argument can justify obedience, so as not to risk setting off a slide into anarchy. However, acts of this sort are exceptional; though the SC argument can support obedience under such conditions, this does not approach a justification for general political obligations. In the large majority of cases the particular laws with which a citizen is asked to comply will neither offer opportunities for dramatic disobedience nor be under powder-keg conditions. In the PD cases that interest us, the citizen is asked to pay taxes, register for military service, comply with anti-pollution laws, etc. Not only will his compliance or non-compliance not dramatically affect the behavior of numerous others, but it will not be noticed by them.

McMahan's discussion, too, concerns general political obligation. His rational anarchist considers whether to obey the law *simpliciter*, without reference to the kinds of special circumstances that would make SC logic applicable to his deliberations.

IV Parfit's Analysis of SE Cases

Though Parfit's logic in SC cases appears to hold in regard to the duty to vote, and perhaps some other requirements of citizenship, a general obligation to obey the law resists this approach. Questions of political obligation, like other SE cases, cause obvious difficulties for consequentialist theories. There is something inherently odd about saying that action A is wrong *because of its consequences*, when its consequences are imperceptible.¹⁵ There are of course laws, the disobedience of which has perceptible consequences. Violations of laws against murder and assault cause palpable harm and are clearly wrong on consequentialist grounds. Such cases will not concern us. Our main concern is with laws that coordinate the actions of large numbers of people in order to provide important public goods, most notably the rule of law, though we should note that obedience to murder and assault laws also bears upon the rule of law. In addition, one sort of imperceptible effect is not controversial. If A contaminates B with a deadly virus, A clearly harms B, though B might not notice that he has been contaminated, and the virus might not show up for several years. The cases on which we will concentrate concern the production of effects that are imperceptible not only because they are not noticed, but because they must be performed in conjunction with large numbers of other, similar actions before their harms become perceptible.

We can begin by looking at simple cases. Should White conserve water in accordance with governmental regulations if compliance would inconvenience her and would have no perceptible effects upon the community's water situation? If society X is beset with air pollution and all inhabitants are required to install catalytic converters on their cars, should Haddock comply if no one would ever know if he did, his non-compliance would have no perceptible effects upon the quality of the air, and catalytic converters are expensive and troublesome to maintain? According to our moral intuitions compliance is clearly required, at least under certain circumstances, and so we must see whether consequentialist arguments can support this conclusion.

Parfit defends a consequentialist analysis of SE cases by attacking what he calls the fifth mistake in moral arithmetic:

(The Fifth Mistake) If some act has effects on other people that are imperceptible, this act cannot be morally wrong *because* it has these effects. An act cannot be wrong because of its effects on other people, if none of these people could ever

¹⁵ See below, n. 17.

notice any differences. Similarly, if some act would have imperceptible effects on other people, these effects cannot make this act what someone ought to do. (RP, 75)

Now Parfit's arguments as to why the fifth mistake is a mistake are not entirely clear. His main argument appears to rest on the presentation of ingenious examples of SE cases that involve actions that are obviously wrong, and then attempting to account for this. One celebrated example is the 'Harmless Torturers.' Imagine a situation in which one thousand victims are tortured simultaneously by one thousand torturers. Each victim is connected to a machine that allows his pain to be increased in imperceptible gradations as a switch is pushed. In the bad old days each torturer repeatedly pushed the switch for a single victim. But now each torturer pushes the switches of all one thousand victims once each. As a result, the victims are all in agony, though no torturer makes any victim's pain perceptibly worse (RP, 80).

Even though they do not perceptibly increase the suffering of single victims, Parfit believes that the torturers' actions are wrong. Their wrongfulness can be accounted for according to a principle such as the following:

(C12) When (1) the outcome would be worse if people suffer more, and (2) each of the members of some group could act in a certain way, and (3) they would cause other people to suffer if *enough* of them act in this way, and (4) they would cause these people to suffer *most* if they *all* act in this way, and (5) each of them both knows these facts and believes that enough of them will act in this way, then (6) each would be acting wrongly if he acts in this way. (RP, 81; his italics)¹⁶

(C12) would explain the harmfulness of the harmless torturers' actions. But (C12) is flawed. Though it is clear that the torturers acting in concert commit grievous wrongs, it is not clear that the actions of a single torturer, which increase the suffering of each victim only imperceptibly, are wrong *because of their consequences*.¹⁷ Indeed, this is what Parfit's analysis of SE cases is intended to show. But (4) in (C12) simply asserts

16 C12 is similar to other principles invoked by Parfit, C10 and C11 (RP, 77-8), which need not be discussed here.

17 To be more precise here and throughout, in saying that the problem is to show that a given act is wrong because of its consequences I make an additional assumption that is not spelled out in the text. To make the cases under consideration more closely parallel the cases concerning political obligation that interest us we would have to posit some benefit from allegedly harmful actions. For instance, in regard to the torturers, we must assume that they gain personal gratification from their work. The problem is to show that the imperceptible effects of their torturing make their actions wrong in spite of this gain.

this, thereby begging the question.¹⁸ It is clear that Parfit needs a stronger argument to account for the wrongfulness of the torturers' actions.

In the final analysis Parfit's case rests upon a particular notion of an act's consequences. He appears to believe that each of the torturers acts wrongly because his acts are part of a set of acts that *together* have perceptible effects. We can refer to a view of this kind as a 'collective view of consequences,' and the consequences of such acts as their 'collective consequences.' It must be clear, however, that the notion here is the consequences of a collective act – i.e., a set of acts that together have consequences – rather than an assemblage of consequences (of separate acts) considered collectively. The main problem we encounter is determining the specific consequences of one particular act of a set of acts that together have harmful collective consequences, in order to be able to show that the particular act is wrong for this reason.

In response to early criticisms of *Reasons and Persons*, Parfit clearly embraces a collective consequence view, saying that he should have based his argument on (C7) rather than (C12):¹⁹

(C7) Even if an act harms no one, this act may be wrong because it is one of a set of acts that *together* harm other people. Similarly, even if some act benefits no one, it can be what someone ought to do, because it is one of a set of acts that together benefit other people. (RP, 70; his italics)

At first sight, (C7) has two great advantages. First, it does not have the grievous problems of (C12). In addition, its collective conception of consequences appears to explain our intuitive response to the harmless torturers, and to countless similar SE cases, including our obligation to obey the law.²⁰ However, though (C7) is plausible on an intuitive level, Parfit does not give adequate reasons why we should accept it, or show that it can be supported on consequentialist grounds.²¹ On

18 This important point is made by B. Gruzalski, 'Parfit's Impact on Utilitarianism,' *Ethics* 96 (1986) 779-82, to whose discussion I am indebted. It is to be noted that Parfit does not have an adequate response in his 'Comments,' *Ethics* 96 (1986) 846-9; see below, n. 21. Because I believe that Parfit's sorites argument (RP, 78-80) is weak, I do not discuss it; on this see Gruzalski, 779-80.

19 Parfit, 'Comments,' 847; for emphasis on sets of acts, see RP, 70-1, 77-8, 82-6.

20 For some similar cases, see RP, 84-5.

21 In his 1986 'Comments' Parfit cannot be said to *argue* for (C7). He writes: 'An act can be right or wrong because of its effects, even if the effects of this particular act are imperceptible. These are not the only relevant effects. The act may be right or wrong because it is one of a set of acts which *together* have perceptible effects.'

an intuitive level it is clear that if Haddock is one of a large number of people who cause severe air pollution, which has widespread harmful effects, then he is guilty of serious wrongdoing, even if his own contribution is imperceptible. But the moral arithmetic to demonstrate this is not easily produced. It is notable that while Parfit supplies detailed figures to support his analysis of SC cases, nothing similar is presented in support of (C7) or his view of SE cases.

V Consequentialist Analysis of SE Cases

In this section I examine a series of attempts to justify the collective conception of consequences expressed in (C7). I will examine what appears to be Parfit's most important argument for (C7), and then address the possibility that (C7) can be established on the basis of a particular consequentialist notion of the consequences of an act. Two different construals of consequences are found in the literature, which can be referred to as the 'marginal' and 'contributory' views. Parfit appears to subscribe to the former, and we can begin by looking at that.

Utilitarians generally hold that the consequences of act A are the differences that its performance makes in future states of affairs. Roughly, X is a consequence of A if, given the circumstances in which A is performed, (a) A is causally sufficient to bring X about, and (b) X would not have obtained had A not been performed. As numerous scholars note, the marginal view is consistent with the moral thrust of AU and is dominant in consequentialist literature.²² There is strong evidence that this is Parfit's view (or at least his preliminary view; see below). He criticizes something along the lines of the contributory view as the first of the five mistakes in moral mathematics (RP, 67-70), while the marginal view is central to his analysis of the Presidential election. As we have seen, Grey's vote can make the difference between A's

In my imagined case, what each torturer does makes no perceptible difference, but what they together do makes a great difference' (847; his italics). These sentences constitute his main support of (C7), but they strike me as mere assertion. (I discuss one possible argument for [C7] in RP, below.) Though numerous examples indicate that acts with imperceptible consequences are clearly wrong when performed along with large numbers of similar acts, Parfit has not shown that they are wrong on consequentialist grounds.

22 H. Silverstein, 'Simple and General Utilitarianism,' *Philosophical Review* 83 (1974) 339-63; 'Utilitarianism and Group Coordination,' *Nous* 13 (1979) 335-60; P. Horwich, 'On Calculating the Utility of Acts,' *Philosophical Studies* 25 (1974) 21-31; the terms 'marginal' and 'contributory' are taken from Regan, *Utilitarianism and Cooperation*, 13-17.

winning and losing. The consequences of Grey's vote are therefore immense, the difference that A's election means to all Americans – all inhabitants of the globe – over the length of his term. As this example shows, the marginal construal is necessary for the Pascalian logic of SC cases.

The mathematics of the marginal view can be illustrated with an example. Assume that a patch of grass, G, valued at 10 units of utility, will be completely destroyed if three people cross it simultaneously, but not damaged at all if fewer than three cross. If Tinker, Evers, and Chance cross, destroying G, the value of each of their actions is not $-3\frac{1}{3}$ but -10 . The failure of any one to cross would have preserved G, thereby saving 10 units of utility. There is a possible problem here: if each crossing is valued at -10 , the sum of the values of the three crossings is -30 , while G is valued at only 10. Thus linear relationships do not necessarily hold between the values of individual acts considered separately and collectively, as groups.²³ Now if we assume that a fourth person, Quinn, crosses along with the other three and that none of the four is able to influence the actions of any of the others, the actions of four individuals destroy G, doing the same 10 units of damage. But because the damage would result from the crossings of any three, each of the four does *no* damage; the consequences of each separate crossing are 0. In this case, the consequences considered individually are $4 \times 0 = 0$; taken collectively, they are of course -10 . Though the arithmetic in the two crossings may strike us as odd, to require that the consequences of acts considered individually be identical to their collective consequences is to commit the fallacy of composition.²⁴ The great advantage of the marginal conception of consequences is that it preserves the intuition, central to AU, that an action should be assessed according to the difference that it makes. If an agent does A, then he should be held responsible for and only for the ways in which A changes future states of affairs.

But accepting the marginal view does not help us with SE cases. By preserving central intuitive features of AU, the marginal view also preserves AU's problems. If Greene's tax payments would have an imperceptible effect upon the national treasury and by not paying he could perceptibly benefit his family and friends, then why is it wrong for him not to pay? If by removing the catalytic converter from his car Haddock would contribute imperceptibly to dangerous air pollution, it is

23 Citations in previous note; on 'linearity,' see Lyons, *Forms and Limits of Utilitarianism*, 63-91.

24 Silverstein, 'Utilitarianism and Group Coordination,' 341

not clear why his action is wrong. Replacing the catalytic converter would make things only imperceptibly better. Limiting the consequences of an act to the differences that *it alone* makes for future states of affairs, makes it difficult to explain why acts with imperceptible consequences are wrong because of their consequences.

Some consequentialist theorists have been led by PD problems to reassess their views of the consequences of actions. In this sense Parfit seems to fall into a long tradition of theorists. Perhaps the best known is R.F. Harrod, who argued that, in assessing certain acts, consequences should be generalized:

*There are certain acts which when performed on n similar occasions have consequences more than n times as great as those resulting from one performance.... It is in this class of cases that generalizing the act yields a different balance of advantage issuing from each individual act.*²⁵

Harrod indicates the mathematical problem here. The single act of removing the catalytic converter from one's car has imperceptible effects; multiplied many-fold, the consequences can be catastrophic. To deal with such cases, Harrod developed an incipient form of utilitarian generalization, which he identified as a form of Kantianism.²⁶ For proponents of AU, however, the problem remains.

25 R.F. Harrod, 'Utilitarianism Revised,' *Mind* 45 (1936), 148 (his italics)

26 Harrod, 'Utilitarianism Revised,' 148 ff. I do not discuss utilitarian generalization (UG) in this paper, as it is not considered by Parfit. Nor do I believe that UG is able to solve SE problems. Though it is not possible to discuss UG in detail here, a few points can be made. To begin with, I do not believe that the famous argument of Lyons (in *Forms and Limits of Utilitarianism*), which criticizes UG for being extensionally equivalent to AU, is valid. Lyons' argument has been criticized along various lines by Silverstein, 'Simple and General Utilitarianism'; Horwich, 'On Calculating the Utility of Acts'; Regan, *Utilitarianism and Cooperation*, Chap. 6.

However, although UG is not extensionally equivalent to AU, UG still fails because in certain cases AU is more consistent with our intuitions. The clearest cases here concern 'minimizing' conditions, actions (a) with imperceptible consequences, (b) which if generalized would produce great benefits, but (c) unless generally performed are likely to be greatly harmful to those who perform them. An example is publicly protesting against an odious, ruthless political regime, e.g., Nazi Germany. Unless very widely performed, such acts of protest are likely to lead to grievous consequences for individual protestors without perceptibly benefiting society. On this, see Gruzalski, 'The Defeat of Utilitarian Generalization,' *Ethics* 93 (1982) 22-38. (An additional, serious problem UG confronts is the problem of 'competing descriptions'; on this, see Gruzalski, 'Utilitarian Generalization, Competing Descriptions, and the Behavior of Others,' *Canadian Journal of Philosophy* 11 [1981] 487-504. But note the problems he encounters with different degrees of inconvenience, which are arbitrarily assumed to be equal, on 493-4.)

As we have seen, Parfit believes that SE cases should be dealt with by appealing to the collective consequences of acts, in accordance with (C7). It is not completely clear how he argues for (C7), though he does not appeal to utilitarian generalization. In Chapter III of *Reasons and Persons* Parfit introduces (C7) in the course of pointing out certain flaws in the marginal view of consequences, before moving on to SE cases. He presents a variety of cases in which the marginal view is inadequate and must be revised. What Parfit calls the second mistake in moral mathematics is ignoring the effects of sets of acts:

(The Second Mistake) If some act is right or wrong *because of its effects*, the only relevant effects are the effects of this particular act. (*RP*, 70; his italics)

Parfit's first kind of case involves effects that are overdetermined. Assume that X and Y simultaneously shoot and kill Q, and that either shot by itself would have been fatal. As Parfit notes, on the marginal view of consequences the absurd result is that, because Q still would have died if either X or Y had not fired, neither has harmed him. It is in response to this absurdity that Parfit advances (C7).²⁷ To repeat (C7)'s crucial sentence: 'Even if an act harms no one, this act may be wrong because it is one of a set of acts that *together* harm other people' (his italics). According to Parfit: 'On any plausible theory, even if each of us harms no one, we can be acting wrongly if we together harm other people' (*RP*, 70). In order to account for the absurdity of the double shooting, then, Parfit revises the marginal view of consequences in favor of a collective conception.

Though (C7) would clear up the absurdity in the double shooting, it is not the most economical approach. In the double shooting, X and Y perform acts that are:

- 1.1 causally homogeneous;
- 1.2 performed simultaneously; and
- 1.3 each act by itself is causally sufficient to bring about the result in question (R).

A case of this sort may provide a legitimate counter-example to the marginal view of consequences, but the problem could be dispelled by less drastic measures than introducing a collective conception. The absurdity in the double shooting arises from (1.3) the fact that each shot by itself would bring about R, and so both X and Y are absolved of responsibility for R by the coincidental fact that the other also shoots.

²⁷ Other, similar cases are discussed in *RP*, 70-1.

Though technical details cannot be discussed here, a more economical solution to the problem would be to regard the marginal construal of consequences as open-ended to a certain extent, and so to attach a qualifying clause to it in order to account for such cases. This is surely preferable to the drastic revision that (C7) represents. The crucial term in (C7) is *together*. In reference to the double shooting, the proper sense of (C7), with 'together' unpacked is:

(C7A) Even if an act harms no one, this act may be wrong because it is one of a set of acts *each of which is causally sufficient to harm other people*.

Because the two shootings are performed *together* only in a temporal sense, (C7A) addresses the absurdity more precisely than (C7) does. The case does not support a blanket conception of collective consequences, because the shootings are not performed together in a *causal* sense. Each act is causally sufficient to bring about R.

The notion of collective consequences in (C7) that the double shooting justifies – that in (C7A) – is misleading if applied to SE cases. Actions in SE cases are performed *together* in two senses. In addition to being performed simultaneously, each action must be accompanied by other similar actions to bring about the result in question; indeed, to have perceptible consequences at all. For instance, in the air pollution case a highly undesirable situation (R2) is brought about by a set of acts that are:

- 2.1 causally homogeneous;
- 2.2 performed (if not simultaneously) in reasonable temporal proximity; and
- 2.3 each act by itself is *not* causally sufficient to bring about (R2), but must be accompanied by a large number of similar acts.

The sense in which these acts are performed *together* is stronger than in the double shooting, encompassing both temporal and causal senses. Thus I believe that Parfit's rejection of the marginal view in collective cases is not supported by his analysis of overdetermined cases. To the extent that Parfit applies (C7) indifferently to both kinds of cases, he equivocates upon the word 'together' in (C7).²⁸

²⁸ The absurdity in the double shooting arises from the fact that each shooter, though having performed an action causally sufficient to bring about Q's death, is held blameless. In assessing moral responsibility for an action the question of causal sufficiency is a crucial factor. The situation is different in the collective cases

Though the application of a principle such as (C7) to SE cases cannot be justified along the lines Parfit presents, there is another possible reason for replacing the marginal view of consequences with a collective conception. In the SE cases we have discussed, our intuitions tell us that something like (C7) is true, that individuals act wrongly, though their acts have imperceptible consequences, when they fall into sets of acts that together cause severe harms. Each of Parfit's harmless torturers is clearly guilty of great wrong, though he harms each victim imperceptibly. The wrongs here emerge if we employ a collective conception of consequences.

Though it is not clear exactly what Parfit means by his collective conception of consequences, it is likely that this should be interpreted along the lines of the contributory conception, which is the other major view of consequences commonly discussed in the literature. According to the contributory view, an agent, Q, who performs act A, is assigned a proportion of the consequences of the set of causally homogeneous acts that are performed along with A. On this account, if Tinker, Evers, and Chance simultaneously cross grass G, thereby doing 10 units of damage, each receives credit for $-3\frac{1}{3}$ units. If they are joined by Quinn and the same 10 units of damage result, then each receives credit for $-2\frac{1}{2}$ units. The great advantage of this view is its ability clearly to trace the harms in SE cases back to the individuals who produce them.²⁹ From Parfit's point of view the advantage of the contributory conception is that it appears to support (C7), by showing how a given action

that interest us. Because each particular action in these cases is not causally sufficient to produce the given result, the question of moral blame is more problematic. Parfit's use of overdetermined cases to call into question the marginal view allows him to smuggle causal sufficiency and so absurdity into collective cases.

I should note that it is not clear the extent to which the reasoning in *RP* is affected by the equivocation, as it is not clear that Parfit's logic actually is:

- A. the double shooting establishes (C7); and so:
- B. (C7) can be used to clear up the SE cases.

However, Parfit presents no other argument to justify a concept of collective consequences in SE cases – other than the fact that such a view gives the desired results (see above, n. 21). It appears most likely that in *RP* Parfit is misled by the flawed reasoning discussed above, in Sec. IV. In his 1986 'Comments,' Parfit appears to rely directly on (C7), without adequate explanation; see above, n. 21.

²⁹ The contributory view has been supported for this reason; see P. Singer, 'Is Act-Utilitarianism Self Defeating?' *Philosophical Review* 81 (1972), 102-3.

is wrong, even if its effects are imperceptible, if it is one of a set of acts that together harm other people. For example, if the harmful effects of air pollution are assigned to the N individuals who cause them, each can be seen to be guilty of serious wrong. In a similar manner this approach establishes the serious wrong committed by each of the harmless torturers.

However, the contributory view has severe problems. As we have noted, Parfit criticizes something akin to this view, which he calls the 'share of the total' view, as the first of his five mistakes in moral mathematics (see *RP*, 67-70). Parfit and other theorists have devised a series of examples in which the share of the total view yields absurd results. For instance:

The Second Rescue Mission. The lives of a hundred people are in danger. These people can be saved if A and three other people join in a rescue mission. These four are the only people who can join in the mission; if any one fails to join all of the hundred people will die. If A fails to join, he can go elsewhere and save, single-handedly, fifty other lives. (*RP*, 68; slightly altered)

If A joins in the rescue mission, one hundred people will be saved. His one-fourth of the credit for this would be 25 lives. Because he can save 50 lives single-handedly if he does not join, he should not do so. But the absurd result is that fifty more people would die. It is to be noted that the marginal view gives the correct answer here. On this view, if A joins the rescue mission he – and each of the three others – gets credit for saving 100 lives by joining, and he gets credit for only 50 by failing to join. Additional counter-examples could be advanced.³⁰

Parfit's discussion of the share of the total view raises several problems. As far as computing the consequences of specific acts is concerned, it appears that the contributory and share of the total views are closely related, while these in turn appear to be closely related to the collective conception of consequences that Parfit embraces with (C7). But if this is true, then we must ask how Parfit can reconcile his acceptance of (C7) with his prior criticism and rejection of share of the total (*RP*, 66-70). There are three questions here, which Parfit does not address: (a) *exactly* what view of consequences is implied by (C7); (b) exactly how the contributory conception of consequences that (C7) appears to imply differs from the share of the total view, and so why the criticisms of share of the total do not apply to it; and (c) the kind of mathematical support that Parfit could present to substantiate the

³⁰ See Regan, *Utilitarianism and Co-operation*, 14-17, 231 n.6; Silverstein, 'Simple and General Utilitarianism.'

conception of consequences expressed in (C7), which he relies upon in *RP*, Chap. III.

In the absence of detailed discussion of these matters by Parfit, I will attempt to shore up (C7) by analyzing it in terms of the contributory view of consequences. Though Parfit's criticisms of the share of the total view appear to damage the contributory view, we can perhaps appeal to the special features of SE cases in order to deflect the criticisms in *these particular cases*. The counter-examples to the contributory view generally concern situations in which definite numbers of individuals must act in specific ways in order to hit (or not hit) specific thresholds.³¹ As noted in Section I, among the features that characterize SE cases (all PD cases) are (b) the need for large numbers of participants, and (c) that each participant's actions are independent of the actions of others. In these cases, it is unlikely that any one agent's actions will cause a specific threshold to be met (or not met), and so the contributory view could perhaps be sustained. As we have seen, various examples indicate that the contributory view appears to give correct answers in SE cases.

If the reasoning in the last paragraph is accepted, then consequentialists could perhaps be able to deal effectively with SE cases. However, even then the contributory view encounters severe problems. Let us assume that society X which contains 100,000 individuals is beset by air pollution, and that to alleviate the pollution entirely, 99,000 X-ites must comply with a series of anti-pollution laws. Thus the non-compliance of 1,000 X-ites will have slight or imperceptible effects. We can assume that the anti-pollution laws are burdensome (the compliance of each X-ite costs 2 units of utility), though the benefits that each would receive from clean air would greatly outweigh his compliance costs. Let us identify a threshold (T1) at which the failure of (say) 20,000 to comply makes pollution in X a significant health hazard; from this point on the costs of pollution begin significantly to outweigh the costs of general compliance with anti-pollution measures. Therefore, levels of non-compliance below but approaching T1 have greater utility than a situation of complete compliance. We can assume that the non-compliance of each 1,000 X-ites, from 0 to 99,000, increases pollution in increasingly large increments, though the effects of any single individual's non-compliance are imperceptible. Finally, we can assume

³¹ I do not rule out additional counter-examples, which would work in SE cases, though the qualifications I introduce allow one to avoid the counter-examples most often found in the literature (see the last note). However, the possibility of additional counter-examples need not concern us, as the contributory view will be seen to fail, for other reasons.

that if 90,000 X-ites fail to comply, the situation will be so grave as to imperil the survival of society and its members. Exactly how air pollution affects a given society is of course a factual question, which, again, probably varies between societies. But altering the example to account for more realistic conditions would probably not affect it greatly.

Under these circumstances, there is a clear range of cases in which a given X-ite, Haddock, would do wrong not to comply. If the pollution in X is at or above T1, then by not complying he would contribute to a harmful state of affairs. If he is one of 30,000 not in compliance, we can assume that pollution is a grievous problem (valued at -400,000 units of utility). In this case, Haddock gets credit for his share of the situation, or -13.33 units. If 40,000 do not comply and the dirty air is valued at -800,000 units, his contribution is -20 units. In these cases, the contributory view supports our intuition that it is wrong to contribute to a damaging state of affairs by performing one of the set of causally homogeneous acts that create it. Once again, the marginal view is unsatisfactory here. Because Haddock's own contribution to the pollution is imperceptible and he receives perceptible benefits (+2 units) by not complying, the marginal view cannot justify compliance.

But the contributory view comes into conflict with our intuitions at both ends of the scale. As we have constructed the example, 1,000 X-ites can fail to comply while doing very slight harm to the air, and enjoying the more significant benefits of non-compliance. If non-compliance is worth +2 units, and the harm to the air is only -500 units when 1,000 X-ites are not in compliance, the total utility here is +1,500. Haddock receives credit for 1/1,000 of this, or +1.5 units. In this case, then, even with the contributory view of consequences, AU recommends against compliance. This clearly clashes with our intuitions concerning fairness. Because clean air is a public good, which is purchased through the sacrifice of all compliers and which Haddock and the other 999 non-compliers receive whether or not they comply, why should they also receive the advantages of non-compliance, unless there is some morally relevant difference between them and the other X-ites?³² As we will see in the next paragraph, a rather different but equally serious problem crops up in a situation of general non-compliance.

32 As Regan (*Utilitarianism and Co-operation*) argues, this sort of case could be addressed by cooperative utilitarianism (CU); see also Narveson, 'Utilitarianism, Group Actions, and Coordination.' However, because CU depends upon the possibility of getting potential cooperators to engage in mutually advantageous cooperative efforts, it is not a practical possibility in many of the cases that interest us. In many important cases, especially those involving very large numbers of people, the behavior of others must be regarded in 'the same way as brute natural phenomena' (Regan, 207; see Chap. 12).

To make the pollution example more directly applicable to questions of political obligation, we can substitute general obedience to the law for general compliance with air-pollution measures, keeping all other factors (and figures) intact. Thus the rule of law rather than clean air is the resultant public good. If these changes are accepted, then, it seems that consequentialism is able to justify obeying the law in an important range of circumstances, but unable in others. Because the logic of the previous paragraph can be assumed to hold here as well, Haddock and 999 other X-ites would be justified in failing to obey the law, if they were the only non-compliers. The costs of their additional compliance in a situation of almost general compliance would outweigh its benefits. A more interesting difficulty is encountered at the other end of the scale. As we have noted, the non-compliance of 90,000 X-ites would place society and all of its members in jeopardy, creating in this case a Hobbesian war of all against all, with a utility of (say) -25,600,000. Under these conditions, the consequences of Haddock's failure to comply would be -258.58; i.e., by not complying he would do a great wrong. But our intuitions strongly suggest that one should not obey the law in a situation in which virtually no one else does. As Hobbes says: 'For he that should be modest and tractable, and performe all he promises, in such time and place, where no man else should do so, should but make himselfe a prey to others, and procure his own certain ruine....'³³ It is notable that the marginal view gives the correct answer here. Because by obeying the law Haddock would contribute imperceptibly to the public good but cause significant harm for himself, he should not obey.

I conclude, then, that whatever the merits of the contributory view in regard to a certain range of SE cases, it is not supportable at either the high compliance or low compliance ends of the scale. Recourse to the contributory consequence view does not justify (C7) or provide consequentialists with a satisfactory way of dealing with SE cases.

VI Conclusion

Throughout this paper we have explored a number of strategies with which consequentialists can attempt to defuse SE cases. We have examined the logic of SC cases, and then McMahon's unsuccessful attempt to apply it to SE cases. We have looked at Parfit's arguments

33 T. Hobbes, *Leviathan*, C.B. Macpherson, ed. (Harmondsworth: Penguin 1968), Chap. 15 (215)

concerning SE cases in *Reasons and Persons*, and then his apparent attempt to argue for (C7) in Chapter III. Finally, we explored the possibility that the collective conception of consequences that underlies (C7) can be established upon the contributory view of consequences. Our conclusion is that none of these attempts has proved successful, and so that Parfit's approach does not yield adequate answers to these important questions. Actions with imperceptible consequences may well be wrong because of their consequences, but strong consequentialist arguments able to demonstrate this remain to be developed.³⁴

Received: February, 1989

Revised: August, 1989

34 On the need to reinforce consequentialist theories with other moral notions, esp. a principle of fairness, see Lyons, *Forms and Limits of Utilitarianism*, Chap. 5; see above, n. 6.

I would like to acknowledge my gratitude to the editors of and anonymous referees for the *Canadian Journal of Philosophy* for valuable criticisms and suggestions. An earlier version of this paper was delivered at the 1989 meeting of the American Political Science Association, in Atlanta. I am grateful for a University of Virginia Summer grant, which greatly facilitated my research.