Australasian Journal of Philosophy
Vol. 64, No. 4; December 1986

RESTRICTIVE CONSEQUENTIALISM

Philip Pettit and Geoffrey Brennan

1. Introduction

Restrictive consequentialism is a variant of the standard consequentialist doctrine. It is not an unknown variant, since it has been widely endorsed in the past, but it is infrequently explicated and less often defended. This paper offers both explication and defence.

Standard consequentialism is a theory of decision. It attempts to identify, for any set of alternative options, that which it is right that an agent should take or should have taken.[1] The theory is characterised by three propositions.

1. Every relevant state of the world, realised or not, has an evaluator-neutral value.
2. The right option in any decision is a function of the value to be realised in the world: as the function is usually understood, it is that which maximises objectively probable value, that which promotes the best objectively probable consequences.[2]
3. The function which determines what is the right decision is also the function which ought to be applied in decision-making: it serves at once to evaluate options, and to select them.

For each of these three assumptions, there is a non-standard variety of consequentialism in which that assumption is lifted. The first is lifted in the evaluator-relative sort of consequentialism recently explored by Amartya

[1] This role in assessing decisions should not be confused with the other functions that ethical doctrines serve in the assessment of agents, actions and the like. A bad agent may choose the right option. And the right option may result in the wrong action: that is, in an action which is inferior to the action that would have come of a different choice. We think that a great deal of confusion has come of failing to distinguish the concern with options from the concern with actions: the failure is marked by the ambiguity often attendant on the phrase 'act-consequentialism'. For some relevant distinctions see Anthony Quinton, *Utilitarian Ethics* Macmillan, London, 1973, p. 49.

[2] These phrases are meant to suggest that the value of the option is to be computed in a manner parallel to the computation of expected utility: see Ellery Eells, *Rational Decision and Causality*, Cambridge University Press, 1982, Chapters 1 and 3, for a survey of some approaches to that computation. The suggestion is standard: see J. J. C. Smart's comments on page 42 in Smart and Bernard Williams, *Utilitarianism: For and Against*, Cambridge University Press, 1973. The suggestion needs further explication however, since we do not mean to insinuate for example that a consequentialist's evaluative ordering must satisfy the conditions for being representable in a real-valued utility function: on such conditions, see for example H. A. John Green, *Consumer Theory*, rev.ed., Macmillan, London, 1976. In a forthcoming paper, Frank Jackson argues that the subjectively right option is that which maximises subjectively probable value. On related matters, see Hugh Mellor, 'Objective Decision-Making', *Social Theory and Practice* 9, (1983).

Sen.[3] The second is relaxed, more famously, in the sort of doctrine known as universalistic or rule-consequentialism: here the right option is that type of option which is such that if everyone chose it or took steps to choose it, then that would maximise objectively probable value.[4]

Restrictive consequentialism lifts the third standard assumption. It suggests that while it may be appropriate to evaluate options by the criterion of maximising probable value, it need not be sensible to select them on that basis. The idea is that the way to satisfy the criterion of evaluation may often be to restrict or forswear its application, relying rather on some other criterion of choice.

Opponents of consequentialism have traditionally alleged that the doctrine cannot go restrictive. The allegation is turned against consequentialism, for it is said that no theory can be plausible if it compels agents to ignore ingrained habits, spontaneous motives and principled commitments, forcing them always to choose on the basis of calculation over outcomes. F. H. Bradley put the objection nicely. 'So far as my lights go, this is to make possible, to justify, and even to encourage, an incessant practical casuistry; and that, it need scarcely be added, is the death of morality'.[5]

More sympathetic expositors of the consequentialist approach have usually taken a different view. They have suggested that with the choice of actions in particular, consequentialism need not require explicit application of the criterion of option evaluation. It may allow people to accelerate and avoid deliberation, taking their guidance from more homely maxims or motives. It will permit this if that generally seems to be the way of actually achieving the best probable consequences.

Sidgwick was explicit on the point. 'It is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are reasonably to be preferred on utilitarian principles'.[6]

Nor was Sidgwick alone. Austin writes in richer vein but to similar effect in The Province of Jurisprudence. 'Of all the pleasures bodily or mental, the pleasures of mutual love, cemented by mutual esteem, are the most enduring and varied. They therefore contribute largely to swell the sum of well-being, or they form an important item in the account of human

[3] See Amartya Sen 'Rights and Agency', *Philosophy and Public Affairs* 11, (1982), p. 30.

[4] See David Lyons, *The Forms and Limits of Utilitarianism*, Oxford University Press, 1965, Ch. 4. See too R. M. Adams 'Motive Utilitarianism', *Journal of Philosophy* 73, (1976), p. 480. Often universalistic consequentialism is characterised as an ethic of action rather than decision. Sometimes the phrase 'rule-consequentialism' is used loosely, like the phrase 'indirect consequentialism', to encompass the sort of approach that we describe as restrictive.

[5] F. H. Bradley, *Ethical Studies*, Oxford University Press, 1962, p. 109. See too Bernard Williams in Smart and Williams *op. cit.*, pp. 118-135.

[6] Henry Sidgwick, *The Methods of Ethics*, Dover, New York, 1966, p. 413. See too James Griffin 'Modern Utilitarianism', *Revue Internationale de Philosophie*, 36, (1982), p. 347; R. M. Hare *Moral Thinking*, Oxford University Press, 1981, pp. 35-40; David Lyons *op. cit.*, p. 149; and Smart in Smart and Williams *op. cit.*, pp. 42-57.

happiness. And, for that reason, the well-wisher of the general good, or the adherent of the principle of utility, must, in that character, consider them with much complacency. But, though he approves of love because it accords with his principle, he is far from maintaining that the general good ought to be the motive of the lover. It was never contended or conceived by a sound, orthodox utilitarian, that the lover should kiss his mistress with an eye to the common weal'.[7]

We think that the supporters of consequentialism are in the right on the issue dividing them from their opponents; but that with some recent exceptions, they have not done enough to argue their restrictive point of view; and that without exception, they have done too little to develop it.[8] Our paper is designed to show how the fault may begin to be remedied. It may also serve, we hope, to draw attention to a significant research programme for applied consequentialist ethics.

2. The Argument for Restrictive Consequentialism

The possibility of restrictive consequentialism is rooted in the fact, more or less ignored in our introduction, that the options which consequentialists are concerned to evaluate are not limited just to behavioural ones. In particular, they also include the psychological options of whether to encourage this or that trait, this or that motive, this or that policy, and so on.[9] Such non-behavioural options are significant, because the choice of a trait or motive or policy is likely to pre-empt certain decisions between act-options.

Psychological profiles which threaten such pre-emption are all predispositions in the following, stipulative sense: they are states whose manifestation in action means that the action is not chosen on a fully calculative or deliberative basis. These predispositions are to be distinguished from ordinary dispositions such as the belief that something is the case, or the desire that it should be so. One may encourage a certain belief or desire in oneself, and do so on conseqentialist grounds, without being thereby inhibited from calculating over the choice of any actions.[10]

If a consequentialist were concerned just with isolated decisions between act-options, then he would be inevitably calculating.[11] He would select what

[7] John Austin, *The Province of Jurisprudence Determined* (ed. H. L. A. Hart), London, 1954, pp. 107-108. We are indebted to Thomas Mautner for drawing our attention to this passage.

[8] The exceptions mentioned certainly include R. M. Adams *op. cit.*; R. M. Hare *op. cit.*; Derek Parfit *Reasons and Persons*, Oxford University Press, 1984; and J. J. C. Smart 'Benevolence is an overriding attitude', *Australasian Journal of Philosophy* 55, (1977). See to Brian Ellis 'Retrospective and Prospective Utilitarianism', *Nous* 15, (1981). An important background piece is R. Eugene Bales 'Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?', *American Philosophical Quarterly* 8, (1971).

[9] They also include more. See Joel J. Kupperman, *The Foundations of Morality*, Allen and Unwin, London, 1983, Ch. 7 for an insightful overview.

[10] The one sort of exception will be a belief that is inconsistent with the second standard consequentialist assumption. See footnote 31 below. We are not concerned in this paper with that sort of case.

[11] A choice between psychological predispositions might be represented as itself a choice or set of choices between act-options. If it were, then our point would have to be recast. The contrast between predisposition-options and act-options would have to be replaced by a contrast between those act-options which affect calculation over other act-options and those which do not.

he did in each case on the calculated ground that it maximised probable value. But as a matter of fact he must also be concerned with his own predispositions. He must ask himself whether he ought to preserve or promote this or that trait or motive or policy, as well as asking whether he ought to perform this or that action. That means that he has to be open to the possibility of deciding to restrict calculation in some areas of action. For to opt for a predisposition will be to accept that some actions—those that manifest the predisposition—will be uncalculatingly generated.

The reasoning, more formally stated, is this.

1. It is possible that the consequentialist agent will opt for (preserving and/or promoting) some predispositions as well as opting for (performing) various actions.
2. If he does opt for such predispositions, then he will not calculate over the choice of the particular actions which they generate.
3. Thus he will be a restrictive consequentialist; he will forswear calculation over some options; specifically, over those actions which manifest the predispositions.

An example will help to clarify the possibility. Suppose that a consequentialist agent finds that he is by temperament inclined to involve himself unselfconsciously in his activities. He will naturally be concerned, not just with what he ought to do in this or that situation, but also with whether he ought to preserve this predisposition. Suppose now that he applies the criterion of option evaluation to the choice of predisposition and decides in favour of keeping it. He must then accept that the actions which manifest that predisposition will not be selected by him on the calculated ground provided by that criterion. Having made a calculated decision in favour of the predisposition, he is bound to forswear calculation in regard to the actions which it generates.

The possibility that a consequentialist may have to go restrictive is of some importance. It means that he may have to choose in a manner which fails to guard against taking less than the best option. The restrictive consequentialist trusts himself in the field of moral action to the control of relatively unseeing predispositions: predispositions which, in the nature of things, are not fine-tuned to the requirements of the circumstances on hand. Inevitably then, he is going to act occasionally in a non-optimific manner. That is the cost he must pay for seeking out optimific predispositions as well as optimific actions.[12]

We have argued for the possibility that the consequentialist will have to go restrictive. We must now identify the conditions under which that possibility will materialise. In order to do this, we need to examine the strategies by which a consequentialist might hope to circumvent the need for

[12] The risk of taking less than the best option will not concern a satisficing consequentialist such as Michael Slote describes. See his *Common-Sense Morality and Consequentialism*, Routledge and Kegan Paul, London, 1985. For a critical notice which rejects Slote's relevant claims see Philip Pettit 'Slote on Consequentialism', *Philosophical Quarterly*, forthcoming.

restriction. If we know those strategies, then we can tell the conditions under which they will fail; and those conditions are precisely the circumstances where restrictive consequentialism becomes more than a mere possibility.

Consider a predisposition P whose presence promises the realisation of a benefit B. There are two ways in which the consequentialist agent might envisage taking B into account without ceding control of his actions to P. These are the strategies by which he might hope to promote the consequences associated with the predisposition without having to pay the cost of restricting his calculative sovereignty.

First of all, he might imagine that having recognised the value of B, there is nothing to stop him from ignoring the predisposition with which it is normally associated, concentrating on the issue of how he ought to act in order to promote B and, more generally, to maximise probable value. On such grounds he might hope to be able to focus on the selection of act-options alone. Why should he think of committing himself to the care of more or less coarse-grained predispositions, if all the benefits that such commitments might promote could be taken into account in a programme of calculating over every action?

If this recourse is to be blocked, so that restrictive consequentialism is a real possibility, then the benefit B must be such that it cannot be attained under the calculative choice of action. It must be calculatively elusive, as we will say. The lustre which unselfconscious involvement gives to behaviour is an example of a calculatively elusive consequence. It is a benefit which is reliably produced by the unselfconscious predisposition but which evaporates under a regime of sustained action-calculation.

Blocked in this way, the consequentialist who recoils from any restriction on calculation might consider a second course. He might envisage adopting P—or at least going through the motions of adopting it—but allowing it to issue in action only when calculative monitoring reveals that the action is indeed for the best. The idea behind this project is that the benefit B which P produces can be equally well produced by the monitored counterpart.

If this second recourse is also to be closed then the consequence B must have a further feature over and beyond being calculatively elusive. It must be unavailable, not just when each action is calculatively chosen, but even when the predisposition which normally produces that action is calculatively monitored. It must be vulnerable to the presence of calculation, even in a supervisory role. The benefit attendant on unselfconscious involvement is calculatively vulnerable in this sense, and not just calculatively elusive. It is destroyed as readily by calculative supervision of the involvement as it is by calculative choice of each action.

In summary: if the argument for the possibility of restrictive consequentialism is to have concrete significance, then the benefits which are thought to motivate a choice of certain calculation-inhibiting predispositions must have two distinctive features. They must be calculatively elusive and, more strongly, they must be calculatively vulnerable. If the benefits do not have these features, then the consequentialist cannot be persuaded to

relinquish calculative control of his actions; the risk of achieving less than the best will be an effective deterrent.

In the remainder of this paper we will be looking at examples of benefits which appear to meet these two requirements. In section three we consider a well recognised case but one which does not readily generalise. In section four we turn to a case that is also commonly recognised but which has the added attraction that it suggests a variety of parallels. Section five is devoted to a taxonomy of those parallels, a taxonomy which suggests that there are many different areas where the consequentialist may be required to go restrictive.

3. A Less Interesting Case of Restrictive Consequentialism

A person's tendency to throw himself unselfconsciously into his activities provides an example of a predisposition with elusive benefits. Examples similar to it abound. They are predispositions which raise the psychological returns to an agent of the actions which they select: the actions become more rewarding than they would have been had they been chosen in a calculating way. Examples are available wherever some form of enthusiasm or dedication or even obsessiveness has the effect of giving an agent greater pleasure in his achievements than he would have if he had been more calculating and detached.

This sort of example will provide a first case where the consequentialist may have to go restrictive, provided that the consequences are calculatively vulnerable as well as elusive. We suggested that they are vulnerable in the case of our unselfconsciousness example. We believe that they are calculatively vulnerable for all examples of this kind.

Suppose that one could monitor the operation of such predispositions, suspending them where necessary for the maximisation of probable value. It is not obvious that such monitoring would necessarily undermine the relevant effects of the dispositions. Nevertheless the effects are calculatively vulnerable. For what is true is that predispositions of the type in question are not such as can be suspended by calculative monitoring. Genuinely to instantiate predisposition is to be more or less incapable, in the particular instance, of inhibiting their operation. The dispositions have a pathological aspect.

This pathological character of the predispositions makes their relevant consequences calculatively vulnerable in a special sense. It means that effective monitoring would cancel the consequences, as vulnerability requires, but this is because the availability of such monitoring would involve the absence of the dispositions. It is not that if an agent could both have a predisposition and monitor it, then the monitoring would destroy the relevant effects. It is rather that if an agent was in a position to monitor and control the predisposition, that would undermine its realisation and thereby the realisation of its effects.

We have found a first case where a consequentialist may be rationally led to forswear calculation. It is not unfamiliar, since it is the sort of case that

has led various writers to argue the merits of motive utilitarianism.[13] The thought behind such a doctrine is that certain predispositions generate results which make them worth having; that the results are not available under a strategy of calculation, or even of calculative monitoring; and that the right thing for a consequentialist to do therefore may be to put himself in thrall to the dispositions. He will encourage the development of those characteristics in the knowledge that he will thereby lose calculative control over certain actions. And he will do this because he believes that that is the way for him to realise the consequentialist optimum.

This case is interesting, but it does not suggest any very general lessons. It shows that there are some predispositions which allow the argument for restrictive consequentialism to assume significance. But it does not indicate how that argument can be borne out more generally. The illustration is available only for dispositions with the highly distinctive character that they are difficult to monitor and suspend.

4. A More Interesting Case of Restrictive Consequentialism

A second case of predispositions with calculatively elusive and vulnerable consequences has been longer recognised in the literature. It is of greater interest, because the predispositions in question do not have to be difficult to suspend. This means that the case may allow of extrapolation to a variety of dispositions. We attempt to sketch some lines of extrapolation in the next section.

The case has been recognised both by utilitarian philosophers and by certain decision theorists. It arises with predispositions to take various calculative short-cuts that have the benefit of saving the agent time and trouble. In order to establish that this sort of case is one where the requirements for restrictive consequentialism are met we need to demonstrate that the time-saving consequence is calculatively elusive and calculatively vulnerable. The demonstration is not difficult.

The consequence is calculatively elusive, because it is not a result mediated by action. What saves time for the agent with a predisposition to take calculative short-cuts is not the action which his predisposition selects. Rather it is the exercise of that disposition itself. Where full calculation would take considerable time, the exercise of the predisposition is likely to consume little. There is no way that the benefit procured by this means could be obtained by someone who followed a calculative route, even if that route led to the same actions.

But is the time-saving consequence also calculatively vulnerable? It certainly is in those cases where the calculative short-cuts spring from habits which, like the predispositions discussed in the last section, are more or less pathological in character.[15] But it turns out to be a vulnerable consequence

[13] See for example Adams *op. cit.*, Hare *op. cit.*, pp. 36 ff, and Parfit *op. cit.*, section 2.
[14] See Smart *op. cit.*, p. 42 and H. A. Simon 'A Behavioral Model of Rational Choice', *Quarterly Journal of Economics* 69, (1955).
[15] See Hare *op. cit.*, p. 38.

too, even when the short-cut involves a calculative strategy which can be suspended at will.

The most obvious example of such a strategy is the disposition to satisfice: that is, to set a level of aspiration in advance of choosing, and in the absence of full knowledge of available options, and to adopt the first alternative that is expected to reach that level. This predisposition can be suspended at will by someone who instantiates it and yet the time-saving which it produces is calculatively vulnerable. That vulnerability can be established by a form of regress argument.[16]

Suppose that a satisficer is persuaded that he ought to monitor his disposition and allow it to control what he does only when it seems likely to maximise probable value. He will commit himself then to a sort of second order maximisation. Presented with a decision situation, he will ask whether at a first order level he ought to satisfice or to maximise: that is, to calculate the probable value of each action and choose the most valuable. He will decide the question by a higher order maximising procedure: he will calculate the probable value of each first order strategy and select the strategy with the higher value. But this monitoring—this higher order maximisation—will cancel out, or at least reduce, the sort of benefit which unmonitored satisficing would have procured. It will involve time costs of exactly the kind that satisficing was designed to avoid.

In the case mentioned in the last section, the consequences which justify the choice of a predisposition are calculatively vulnerable for a special reason: calculative monitoring is incompatible with the disposition itself and for that reason incompatible with its effects. Here in the satisficing case we have a disposition which has more straightforwardly vulnerable consequences. The disposition secures the benefit of saving time and any attempt to monitor it for the importance of that effect, even one which endorses the strategy, will ensure that the effect is not realised, or not realised in the same measure.

In view of its calculatively elusive and vulnerable consequences, the satisficing strategy—and indeed any time-saving maxim—will often attract the commitment of the consequentialist. He will make that commitment in any range or sphere of activity where it seems likely that time costs will be important relative to other considerations. An initial, schematic calculation will usher further calculation from the scene—subject perhaps to periodic review—and will put the agent on automatic pilot, submitting him to the more or less mechanical direction of the appropriate maxim.

The metaphor of the automatic pilot is appropriate in a further respect that is worth specifying. Just as an automatic pilot will be disengaged in emergencies, so the maxim is subject to escape clauses. Thus if the agent comes to learn in any instance that the best thing for him to do there is after all to break with the maxim, then he can have no ground for not doing so. Equally, if it becomes clear to the agent that the situation is out of the ordinary run, say because it involves some sort of emergency, then he must avoid

[16] See Philip Pettit 'Satisficing Consequentialism' *Proceedings of the Aristotelian Society*, Supp. Vol. 58, (1984), where this argument is presented.

an unthinking reliance on the general rule. The consequentialist satisficer is not wedded to his maxim; he espouses it only so far as it promises to deliver optimal results. Where that promise is withdrawn, he has no reason for remaining faithful.

5. The Possibility of Generalising the Second Case

In the satisficing case, the consequence which is liable to justify choice of the predisposition—the fact that it saves time—is deliberatively fragile. It cannot be pursued in direct calculation over actions, nor in calculative monitoring of the disposition to satisfice. Any such pursuit would be self-defeating. It would itself consume time, and would eliminate or reduce the benefit on offer.

This single feature of satisficing explains both why its potentially justifying consequence is calculatively elusive and why it is calculatively vulnerable. The observation is useful, for it suggests that we can expect to find cases where the argument for restrictive consequentialism applies wherever we can identify predispositions which have such deliberatively fragile results. The results are liable to justify selection of the predispositions and eschewal of calculation over the actions to which one is thereby predisposed.

We have a method of extrapolating our second case in prospect, for the notion of deliberatively fragile results is a familiar one.[17] We have only to find such deliberatively fragile results as attach to predispositions and we will have new cases where consequentialists should go restrictive. The dispositions to which we will particularly look are commitments to follow certain maxims; such commitments are the paradigm of non-pathological predispositions that can be suspended on particular occasions.

In view of the currency of the phrase in consequentialist circles, it is tempting to think of the maxims which we shall be identifying as rules of thumb. But care is needed, for this phrase is sometimes used to pick out merely presumptive maxims, not properly restrictive or pre-emptive ones.

A presumptive maxim tells an agent to make a presumption in favour of doing a certain act A under conditions C, when he is calculating what to do in such circumstances. It does not prohibit the agent from applying the consequentialist criterion of option evaluation when he is selecting the action. It just prescribes that he should be loath to trust evidence that suggests doing anything other than A.[18]

[17] For a virtuoso review see Jon Elster, *Sour Grapes*, Cambridge University Press, 1983, Ch. 2.

[18] It is not clear to us whether any consequentialists are counselling a presumptive rule approach in advocating selectional rules of thumb. If their only reason for prescribing rules of thumb is a belief that agents in the field are calculatively fallible, then this probably is what they are recommending. But in any case we do not think that the approach is significantly distinct from that of just applying the evaluation criterion in selecting inputs. It is not the sort of line for which we shall be arguing.

Bernard Williams foists something like the presumptive rule approach on his opponents. In Smart and Williams *op. cit.*, p. 127 he writes as follows of the rules that they recommend: 'it is important that I treat them as rules of thumb, which means not only that if I do discover that this is an exceptional case, then I treat it as an exception, but also—and importantly—that I keep a utilitarian eye open for signs that a case may be exceptional'.

A restrictive or pre-emptive maxim tells an agent to do A under conditions C, given at least that there is no (loosely specified sort of) emergency.[19] The rule pre-empts the agent's calculating in accordance with the criterion of option evaluation, for while the conditions C may be of various kinds, they cannot—on pain of making the rule redundant—include the condition that the agent calculates that A is for the best. The rule directs the agent to ascertain that circumstances C obtain, and that the escape condition is unrealised, but not to bother with any information beyond that: in particular, not to try to identify and weigh the pro's and con's of doing A in that particular instance of C.

But to return now to the main business, we want to identify various predispositions, consisting of commitments to pre-emptive maxims, which are distinguished by the fact that they have desirable but deliberatively fragile outcomes.

Our original case, exemplified by satisficing, is one where the deliberation involved in calculatively monitoring a predisposition, or in replacing it by straight calculation over action, itself undermines a benefit which the unmonitored disposition would have. That case suggests three analogues. In each of these a consequence of the deliberation, not the deliberation itself, destroys that benefit. In the first the destructive consequence is the agent's becoming aware of deliberating; in the second, other people's becoming aware of his deliberating; and in the third, his acting on the basis of the deliberation.

In the remainder of the section we will illustrate this range of cases. The illustration may be of interest in its own right. Mainly, however, it should serve to reveal further avenues for research, indicating the rich resources of restrictive consequentialism.

The original case

The original case is already well illustrated by the satisficing example. But we would like to offer one further instance, in particular an instance with a less technical aspect. The example is provided by the maxims which serve to produce virtue, at least on one particular conception of virtue.

On that conception, a characteristic feature of the virtues—or in any case of the virtues to which the conception applies—is that they require the eschewal of a certain kind of calculation. There may be various background qualifications to be entered but within the limits which these set, to possess one of the virtues is to be able to hearken to certain considerations, while remaining deaf to others. This filtering of attention may be principled,

[19] The restrictive rule approach is clearly what David Lyons ascribes to consequentialists when he writes *op. cit.*, p. 148: 'follow the rules, indeed, but not when you know or are quite certain that breaking one will have better effects on the whole than keeping to it'. We shall see at the end of section 6 that one sort of plausible restrictive rule does not even require this escape clause. While C cannot sensibly include the clause that the agent has not calculated and found that some action other than A is best, they may in some cases include a weaker calculative condition: for example, a condition to the effect that the agent has not calculated and found that A falls by more than a certain margin below the optimum available. Such a possibility is raised in Samuel Scheffler, *The Rejection of Consequentialism*, Oxford University Press, 1982, Ch. 2.

requiring renewed commitment, or ingrained, being a matter of established habit.

To be honourable, so the story then goes, is to find one's commitments obligating and motivating—albeit they may be overridden—without first having to see whether their fulfilment is for the best in some impersonal or global scheme. To be courageous is to have the ability to be galvanised by a venture on hand without the prior assurance that it is worth the danger to oneself involved in its pursuit. To be generous is to enjoy the capacity to respond to certain demands made by others without weighing up the cost to oneself in time or money. And to be a person of integrity, as Bernard Williams argues, is to have the gumption that enables one uncalculatingly to manifest one's deepest, even self-defining, attachments in the bulk of one's normal behaviour.[20]

We assume that some such virtues are indeed worth having. That is not unreasonable, since they all have the aspect of powers. They are capacities to be motivated by considerations which it is easy to lose sight of; capacities which make one proof against weakness of the will, the pale cast of thought, and other such ailments of practical reason. In some approaches indeed the virtues are capacities with a cognitive dimension: they are necessary even for a person to become attuned to the considerations on which they bear.[21]

By some accounts fidelity to appropriate maxims is sufficient to constitute virtue. By all accounts it is sufficient to cause virtue eventually to appear: this, through leading to the formation of corresponding habits. In either case we can say that virtues of the kind surveyed are consequences of a commitment to suitable maxims. And the question then is whether they are calculatively elusive and vulnerable consequences.

The reason that virtue can come as a consequence of committing oneself to appropriate maxims is that such a commitment enforces the calculative discipline associated with virtue. It allows a place in deliberation only to those considerations that should count with the virtuous agent. This practice of ratiocinative exclusion is what constitutes virtue or causes it to appear.

Given the reason why fidelity to suitable maxims may produce corresponding virtues, the production of those virtues is bound to be calculatively elusive and vulnerable. The agent who calculates consequentially over every action is certainly not going to exhibit or pick up the virtuous habits in question. And neither is the person who goes through the motions of virtue but stops to check every manifestation for its consequentialistic sense. If the consequentialist agent wants to develop virtuous patterns of thought, then he has no choice but to go restrictive. He must forswear the sort of practical reasoning that is classically associated with his ethic.

[20] See Bernard Williams *Moral Luck*, Cambridge University Press, 1981, Ch. 3. See too Michael Stocker 'Values and Purposes: The Limits of Teleology and the Ends of Friendship', *Journal of Philosophy* 78, (1981).

[21] John McDowell 'Virtue and Reason' *The Monist* 62, (1979).

The first derived case

So much for the original sort of case: the case where deliberation itself undermines a benefit that may serve to justify choice of a predisposition. We now turn to three derived cases, in each one of which a consequence of deliberation destroys the effect. In the first the consequence is the agent's awareness of deliberating, in the second other people's awareness of his doing so and in the third his acting on the basis of his deliberation.

The first of these further cases is of a familiar kind. It arises when the potentially justifying consequence of fidelity to some maxim is a state of the agent which requires a degree of unselfconsciousness. This requirement means that if the result enters the arena of deliberation, whether in calculation proper or in calculative monitoring, then the fact that deliberation generates selfconsiousness jeopardises the result. If you want the consequence on offer therefore, you had better bind yourself to the maxim that produces it and forswear further calculation.

The primary example of such a result is unselfconsciousness itself, since this is often prized for its own sake. Other instances are states in which unselfconsciousness is a component or for which it is a precondition. Spontaneity is one example, being relaxed another.[22]

Suppose you wish, among other things, to be in such a state. If you try to calculate over actions to determine whether the achievement of the state justifies choice of one rather than another, then inevitably you will become conscious of deliberating. Similarly if you try to monitor a maxim – say, 'Act first, ask questions after' – which on its own would produce the state. Either way you will destroy whatever chance you had of getting the desired result. You will condemn yourself to its absence.

It follows that if you reckon that the state is generally of great importance in a certain sphere of activity – perhaps only a limited one, such as a game provides – then you should be prepared to eschew all calculative control there. You should trust yourself to a maxim like that advising the postponement of questions. Only by going on automatic pilot in this way, can you achieve your desired result. The state will be realised by fidelity to such a maxim, but in a calculatively elusive and vulnerable manner. It cannot spring from action that is chosen with calculation and consciousness. And similarly it cannot issue from attachment to a maxim, if that attachment is subject to conscious supervision.

The second derived case

The second of our three derived cases is of potentially greater importance than the first. It arises when I can get another to believe that I am following a certain maxim in my behaviour towards him only if I do actually follow that maxim; where his believing this produces some desired effect; and where his believing that I calculated, or calculatively monitored, every move would not do so. In such a case I will secure the effect only if I follow the maxim. Thus the effect will be calculatively elusive and calculatively vulnerable.

[22] See J. J. C. Smart 'Benevolence is an over-riding attitude'.

A first, uncomplicated illustration is provided by the security which lovers or friends produce in one another by being guided, and being seen to be guided, by maxims of virtually unconditional fidelity. Adherence to such maxims is justified by this prized effect, since any retreat from it will undermine the effect, being inevitably detectable within a close relationship. This is so whether the retreat takes the form of intruding calculation or calculative monitoring. The point scarcely needs emphasis.[23]

The example lends itself to a general pattern of analysis, if it is conceptualised with the help of the idea of loyalty. The definition of loyalty presupposed in the analysis is stipulative; those who reject it may replace the term by another.

1. If you are loyal to family, friends or associates, then you will not weigh their claims on you in a standard consequentialist way. You will be committed to the fulfilment of such claims independently of whether fulfilling them turns out to be the optimific action; and this, though you may decide on occasion that other considerations override the commitment.
2. Because such a commitment will inevitably be visible to your contacts, it will be a source of security for them. It will be an assurance that they matter to you, since their claims provide you with an independent motivation to action: that is, a motivation which does not depend on the discovery that fulfilling the claims is impersonally optimific.
3. This effect of the commitment is bound to be calculatively elusive and vulnerable. It will be obvious to your beneficiaries if you calculate over your responses to them, or if you calculatively monitor those responses, and once they see this, they will see that they do not matter to you in a way that they would expect of someone loyal. You will not be thought loyal, merely conscientious.
4. The security of immediate contacts is likely to concern you and so, if you are a consequentialist, you should be prepared to forswear calculation and calculative monitoring in favour of the commitments—in effect, maxims—distinctive of loyalty.

The benefit of analysing our initial example in this way is that it suggests a further, more significant illustration of the second derived case. Instead of loyalty and the special claims which correspond to it, this illustration involves respect for persons and the rights acknowledged under a regimen of respect.

In parallel to our earlier analysis, the analogy suggests the following argument:

1. If you respect a person then you will regard certain of the claims he makes as privileged: that is, as claims which block goals whose realisation you would otherwise see as more important than the

[23] The case for loyalties is more fully explored in Philip Pettit 'Social Holism and Moral Theory', *Proceedings of the Aristotelian Society* 86, (1985-86). See too Philip Pettit and Robert Goodin 'The Possibility of Special Duties', *Canadian Journal of Philosophy*, forthcoming.

fulfilment of those claims. The claims will count as rights in the sense in which rights are distinguished by their trumping role.[24]

2. Such respect invests the beneficiary with dignity. Since your attitude will be obvious to him, and probably to people at large, it means that he can enjoy a certain sort of discretion over your behaviour; specifically, the discretion to stop you sacrificing his interest to the achievement of something that you regard as a greater objective good.[25]
3. This dignity is a calculatively elusive and vulnerable consequence. You cannot achieve it if you are known to calculate over your treatment of the individual, or to monitor that treatment, because the fact of deliberating in such a manner means that you do not yield any controlling discretion to him. And if you do actually calculate or calculatively monitor your responses to the individual, then that will inevitably become known.
4. The dignity of the people with whom you deal is likely to be important to you and so if you are a consequentialist you should be prepared to forswear calculation and calculative monitoring in favour of the commitments—in effect, maxims—distinctive of respect for persons.

This consequentialist argument for the importance of respect and of the rights recognised under the dispensation of respect is in significantly greater need of elaboration and defence than the corresponding argument for loyalty.[26] For present purposes however, it may stand in its austere form. It serves in tandem with the loyalty argument to demonstrate the potential significance of our second derived case. Clearly there is a real possibility that if a consequentialist is concerned about dignity then he will have to take rights seriously, restricting his own calculative impulses. That is a prospect sufficient on its own to make restrictive consequentialism worth exploring.[27]

The third derived case:

And now, finally, to the third derived case. This arises where deliberation undermines a potentially justifying consequence of adhering to a maxim, not by its nature, nor because of the awareness it generates, but because of the action to which it leads.

The sort of example which we propose in illustration of this possibility is an effect of maxim-fidelity which materialises only over a considerable period of time. Suppose that the effect is forthcoming from behaving in

[24] For this view of rights see Robert Nozick, *Anarchy, State, and Utopia*, Basic Books, New York, 1974, pp. 28-30 and Ronald Dworkin, *Taking Rights Seriously*, Duckworth, London, 1977, pp. 90-94. See to Philip Pettit, 'Rights, Constraints and Trumps', *Analysis*, forthcoming.

[25] See Joel Feinberg, *Rights, Justice, and the Bounds of Liberty*, Princeton University Press, 1980, Ch. 7.

[26] One of us has elaborated the case elsewhere. See Philip Pettit 'A Consequentialist Case for Rights' in Denis Galligan and Charles Sampford, eds, *Law, Rights and the Welfare State*, Croom Helm, London, forthcoming and 'The Consequentialist Can Recognise Rights', *Philosophical Quarterly*, forthcoming.

[27] For a different line see David Lyons 'Utility and Rights', *Nomos* 24, (1982).

accordance with the maxim over a certain period; that each act of compliance with the maxim is troublesome; and that no one act on its own makes the difference between the achievement and non-achievement of the effect. In that case deliberation would undermine the effect, because it would select non-compliance in each instance.

The maxim in question might be a trivial one, such as that of washing one's teeth after every meal. Sticking to a maxim like this over one's lifetime would ensure dental health; so at least we may assume. But the effect is both calculatively elusive and calculatively vulnerable. It would be undermined by the actions to which instance-by-instance deliberation must lead.

Suppose that I were to calculate after every meal whether to clean my teeth. On the negative side I would count the non-trivial cost of going to the trouble required. What would I count on the positive? The temptation is to say: the importance of the cleaning for my dental health. But the fact is that the significance of each individual cleaning for my oral well-being is negligible or next to negligible. No individual cleaning is sufficient to make the difference between having healthy teeth and not having healthy teeth; this is because dental health is a vaguely defined gestalt. Thus, other things being equal, calculation after every meal would always fail to elicit a walk to the bathroom; the result would be, bad teeth.[28]

Dental health then is a calculatively elusive consequence of adherence to a certain maxim. In order to achieve it, one must forswear calculation over individual actions. Better find a rule and stick to it. Wash after every meal, or wash every day, or wash when some regular chance event occurs. Do anything other than weigh the pro's and con's in every instance. Othewise you will never resist the opportunity to free ride on your future selves, and you will fail to ensure that you maintain your teeth.

We have laboured the fact that dental health is a calculatively elusive consequence of adherence to the maxim. The other question is whether it is also a consequence of a calculatively vulnerable sort.

Consider again the situation of deciding whether or not to wash my teeth now. If we agree that other things being equal, strict calculation of the pro's and con's would not galvanise me to action, then equally it must be admitted that a calculatively monitored maxim would fail to do so. The unmonitored maxim would certainly have the desired effect. But it would be robbed of all its power, once subjected to the question of whether adherence in this particular instance is strictly for the best. The question would be an immediate stimulus to free riding.

The dental health example is trivial but it stands proxy for many other illustrations of the third derived case where our argument for uncalculating consequentialism applies. We find similar examples wherever there is a good which has the following characteristics: it emerges over a period from

[28] For some discussion of intertemporal, intrapersonal free riding, see Richard Tuck 'Is there a free rider problem?' in Ross Harrison, ed., *Rational Action*, Cambridge University Press, 1979. If principle C10 in Derek Parfit *op. cit.*, p. 77 is accepted, then an analogue will block the temptation to free ride in this way. For more on free riding see Philip Pettit 'Free Riding and Foul Dealing', *Journal of Philosophy*, forthcoming.

independent actions; each of those actions is relatively burdensome; and none of the actions makes the difference between the appearance and non-appearance of the good. Under any such circumstances the person will be tempted at each moment to free ride on future efforts. Even where the good is a matter of the deepest self-interest therefore, the agent may be sure of achieving it only if he commits himself to an appropriate maxim.

In concluding discussion of this final case, it is worth noting that the case is exceptional in an important regard. It does not satisfy an otherwise general rule: viz. that the consequentialist will remain faithful to a maxim only if he does not independently come to believe that breaking it is actually for the best in the instance on hand. The consequentialist will abandon the sort of maxim that produces virtue, or unselfconsciousness, or loyalty, if that is known in a given instance to be genuinely for the best. In such an event—however unlikely—he will even violate the maxim that ensures respect and rights. But he will not defect from the maxim in our third derived case, even when he knows that doing so secures the optimal result.

It is not surprising that this case should be exceptional, for the damage done there by deliberation is due precisely to the fact that deliberation selects the option with the best probable results. If the restrictive consequentialist has found reason for submitting himself to some maxim in such a case, then knowing that following the maxim in a particular instance produces less than the best will not deprive him of that reason. It is precisely that sort of knowledge which motivate his adherence to the maxim in the first place.

Still, there is a paradox here. The consequentialist endorses the criterion of option evaluation associated with best probable consequences but in this sort of case he operates with an option selection criterion that is inconsistent in each instance, and is known to be inconsistent, with the achievement of the best probable consequences. How can that be?

The answer has to do with temporal perspective. The selection criterion for each option is inconsistent with the criterion of evaluation as applied to that particular option but not with the criterion of evaluation as applied to the series of options over which selection has to be made. In order to secure the best result overall therefore, the agent has to be sure that he does not pursue the best result in each case. Like Ulysses he must tie his hands. But he must do so to protect himself from his rational impulses, not from any irrational visitation.

6. Conclusion

We have tried to show that there is good reason why the consequentialist may want to go restrictive in certain areas of action and that those areas abound. The upshot is a fresh view of the possibilities of consequentialism, albeit one which squares with the traditional emphasis of defenders of the doctrine on the need for indirection. The view is worthy of further exploration, we believe, for we have done little more than provide some scattered illustrations of the need for consequentialists to go restrictive. It holds out

the prospect of a consequentialism which fits better than many other variants with common sense intuitions about morality.[29]

There are two objections which will certainly be brought against the line which we have argued and we would like to end the paper by mentioning these and indicating why we think that they do not succeed.

The first objection is that our approach is not of any significance, representing a trivial variation on the consequentialist themes. The suggestion is that a consequentialist who proclaims the importance of restrictive rules still fits the familiar picture. He differs from the old-style image of the consequentialist only on matters of detail, not in his essential attitudes.

In his essential evaluative attitudes it is true that our consequentialist fits the standard profile. But we maintain that the difference on questions of selection makes for a real discontinuity. There is nothing trivial about the change in outlook demanded by accepting the points that we have made.

Amartya Sen has observed that moral principles can be usefully distinguished on the basis of the sort of information whose use they exclude.[30] Utilitarian rules render information on the identity of those who benefit from an action irrelevant; rules drawn from natural rights make information on the overall outcome of respecting a right redundant; and so on. Different moral principles are different informational constraints.

This observation connects illuminatingly with our approach. At the level of evaluation the consequentialist endorses the sort of informational constraint which we would expect; all that is relevant to the assessment of an input is how the world promises to lie in its wake. But if our approach is adopted, then at the level of selection quite different informational constraints may prevail. The essence of the approach is to say that often the consequentialist has reason to restrict the sort of information-use by which his position is distinguished.

What we have argued then is that in the field of action the best thing for the consequentialist to do may be to forsake his established moral persona. This thesis can scarcely be dismissed as insignificant. There could hardly be a more demanding amendment of consequentialism than to require that it should be self-effacing in this manner.

The second objection to our approach is that it endorses self-deception and even deception of others. It will be said that you are recommended self-deception in being told that while you should strive to bring about your own virtue or unselfconsciousness, you should do so by a path that involves not thinking about those goals. Equally it will be held that you are prescribed dishonesty in being instructed to promote the security or dignity of your beneficiaries by making it seem that your choice of action is not dictated by a concern for that good.

[29] We do not claim that all such intuitions are accommodated. For example we believe that for all we have said, consequentialism lacks the resources to provide an appropriately strong prohibition on interpersonal free riding. See Pettit 'The Prisoner's Dilemma and Social Theory', *Politics* 20, 1985, pp. 9-10.

[30] See Amartya Sen 'Informational Analysis of Moral Principles' in Ross Harrison *op. cit.*

The charge of dishonesty is obviously misplaced. You may make it clear to your beneficiary that your choice of maxim, and ultimately therefore your choice of action, is motivated by a concern for his security or dignity. What you are required to do is to make it equally clear that acting on the maxim does not involve weighing the importance of that effect in the scale of pro's and con's. There is no deception of others involved here.

And neither in the other cases is there any deception of self. I do not have to hide from myself the fact that following such and such a maxim is designed to produce my virtue or unselfconsciousness or whatever. I can retain a keen awareness of that goal, provided that in my choice of particular actions I stick to the maxim and keep the goal out of deliberative play. What is required is not self-deception, only self-discipline.

In order to emphasise that point, consider the contrast between the sorts of goals which we have mentioned and a goal like that of believing something. A belief is a route-specific state, in the sense that you can coherently aspire to instantiate it, only if you aspire to achieve it along a particular path: specifically, by being rationally persuaded that it is true. There would be something self-deceived about cleaving to a maxim with the goal of inducing a particular belief in yourself. But there is no such self-deception involved in the oblique pursuit of the goals mentioned in our examples. Goals like virtue and unselfconsciousness are not route-specific. You can aspire to exhibit them without aspiring to achieve them along any particular path.[31]

The upshot is that the restrictive consequentialist can be both ingenuous and reflective. He can keep all his aspirations in the open, so long as he is capable of insulating them from his decision-making processes. What is required of him may not be easy to attain but it is not a dissonant or unattractive cast of mind. It may even be an essential part of moral wisdom.[32] [33]

Australian National University

Received May 1985
Revised January 1986

[31] The belief case shows that we should add a further category to the calculatively elusive and vulnerable consequences already distinguished. These are the sort of consequences which are vulnerable even to strategic planning: to pre-calculation, as we might say.

[32] Perhaps the part emphasised in the tradition of Zen Buddhism. See Elster *op. cit.* In this connection it is worth mentioning that one of the effects of adopting a restrictive form of consequentialism will be to make room for an emphasis on the non-mechanical, non-actuarial character of moral judgment.

[33] We are grateful for helpful comments provided when this paper was read at the Dept of Philosophy, Faculty of Arts, Australian National University and at the Dept of Philosophy, University of Adelaide. We are also grateful for comments received from Stanley Benn, Jerry Dworkin, Peter Forrest, Jerry Gaus, Loren Lomasky, Hugh Mellor, Graham Nerlich, Huw Price, Christina Slade and Kim Sterelny.