Consequentialism and Decision Procedures

Toby Ord
Balliol College
University of Oxford
June 2005

Thesis for the degree of Bachelor of Philosophy

Abstract:

Consequentialism is often charged with being self-defeating, for if a person attempts to apply it, she may quite predictably produce worse outcomes than if she applied some other moral theory. Many consequentialists have replied that this criticism rests on a false assumption, confusing consequentialism's criterion of the rightness of an act with its position on decision procedures. Consequentialism, on this view, does not dictate that we should be always calculating which of the available acts leads to the most good, but instead advises us to decide what to do in whichever manner it is that will lead to the best outcome. Whilst it is typically afforded only a small note in any text on consequentialism, this reply has deep implications for the practical application of consequentialism, perhaps entailing that a consequentialist should eschew calculation altogether.

In this thesis, I take this consequentialist reply and examine it more closely. I begin with some preliminaries, expounding the objection and showing how the decision procedure approach differs from that of distinguishing subjective and objective rightness. I then focus for the remainder of the thesis on providing a consistent consequentialist approach to the question 'How should I decide what to do?' (in contrast to the more familiar question 'What should I do?').

I first show that act-consequentialism does not have sufficient resources to answer this question and why rule-consequentialism fails to provide a promising alternative. Instead, I argue that we should turn to the theory of global consequentialism which explicitly judges decision procedures (along with everything else) in terms of their consequences. It thus allows us to say that I should follow the decision procedure that will lead to the best outcome.

I then examine the nature of decision procedures themselves as well as the role in which they are to be judged. I demonstrate that consequentialist authors have not paid sufficient attention to the many possibilities, and argue for an interpretation where the good of following a decision procedure is construed in terms of the benefits of accepting it, rather than those of complying with it or of executing it flawlessly.

I conclude that the consequentialist analysis of decision procedures can be made much more precise than it has been and that the question of 'How should I decide what to do?' deserves a much more central place in our understanding of consequentialism.

Contents

Chapter 1

How we should decide what to do

A common criticism of consequentialism is that it cannot be practically applied. Moreover, if we attempt to apply it in practice, this may predictably lead to worse outcomes than if we were to adhere to some other moral theory. In this chapter, I shall introduce consequentialism then provide a sketch of these criticisms and of the standard consequentialist reply. This reply, which says that we are to decide what to do according to the best available decision procedure, will then form the focus of subsequent chapters.

1.1 The act-consequentialist criterion of rightness

Act-consequentialism is, at heart, a very simple account of the connection between the rightness of acts and the goodness of their outcomes. It rose to prominence in the work of the early utilitarians[1] and has had an enormous impact on moral philosophy in the 19th and 20th centuries. During this time it has been developed for its own sake and has also served as a measuring stick to which rival moral theories are compared. Act-consequentialism is, roughly speaking, the doctrine that:

> An act is right *iff* it will lead to at least as much good as any alternative act.[2]

To this a consequentialist will add an *axiology*: an account of what it is that constitutes the good. For example, the earliest and most well known consequentialist theory, *hedonistic utilitarianism*, holds that the goodness of an outcome is the total balance of happiness over suffering. Other consequentialist theories claim that the good is constituted in some different

[1] Notably: Jeremy Bentham, John Stewart Mill, Henry Sidgwick and G. E. Moore.

[2] The definition here is formulated so as to explicitly take ties into account, however it is often convenient to speak as if there is but a single best act and, since nothing important shall turn on this, I take this convenience herein. In addition, the term 'iff' is to be read as 'if and only if', and the term 'right' is to be read as 'permissible', but since I am discussing consequentialism and ignoring ties, 'permissible' and 'obligatory' turn out to be true of the very same acts. Also, while I shall be concerned with the subjective sense of rightness in Chapters 2 and 5, I use it elsewhere only in the objective sense (this shall be explained further in Chapter 2).

way, such as by the fulfilment of rational desires or the possession of certain attributes from some objective list (*e.g.* happiness, education, being loved by someone). Consequentialist theories can move further still from the classical utilitarian versions, taking account of the distribution of personal good over the population or even ranking the outcomes by the number of broken promises they contain. In this essay, I shall not be concerned with particular axiologies. The topic at hand affects them equally and a resolution would be just as useful to all. Thus, it shall be convenient to take the term 'consequentialists' to refer to adherents of particular consequentialist theories complete with their own (unspecified) axiologies.

Even ignoring the question of axiology, one finds considerable debate about how the core consequentialist doctrine should be interpreted.[3] For example, some consequentialists take it that the criterion is concerned with the good that is *caused* by the act under consideration. Alternatively, we might interpret the criterion as being concerned with the future that comes about when the act is performed regardless of causal connections. On a third interpretation, the proper objects of evaluation are not mere futures, but entire possible worlds. This would let us take past events into account when judging actions, and allow concepts like desert or promise breaking to enter the axiology. On such points I shall try to remain neutral, as I hope my arguments shall apply in all interpretations.

I shall, however, take a position on a few contentious issues. One of these concerns the debate between the *actualists* and the *probabilists*. The actualists hold that consequentialism is concerned with actual outcomes, whatever they happen to be. The probabilists, on the other hand, hold that consequentialism is concerned with the possible outcomes in some proportion to the chance that they will occur. Typically, this takes the form of concern with the *expected good* that results from the action.[4] For simplicity's sake, I shall side with the actualists and take consequentialism as being concerned only with the goodness of the outcomes that actually occur. However, most of my claims could be easily modified so as to apply to the probabilist interpretation instead.

I have introduced the criterion as one that assesses the rightness of individual acts and does so on the basis of their consequences in the situation at hand. It is thus said to be an *act-consequentialist* interpretation. Opposed to this are various forms of indirect consequentialism, where the acts are assessed on the basis of something else. The classic example here is *rule-consequentialism* in which rules are assessed on the basis of their consequences and the right acts are those which are in accord with the best set of rules.

For now it shall be convenient to use the act-consequentialist interpretation, since this is the most common interpretation and the major target of the criticisms that I wish to discuss.

---

[3] See Carlson (1995) for a detailed overview.

[4] 'Expected good' is a technical term which needn't bear any relation to what the agent actually expects. Instead the expected good is the sum of the good in each possible outcome, weighted in direct proportion to the probability that this outcome will occur. Thus a situation with a 20% chance of 100 units of good and a 80% chance of 10 units of good would have an expected good of $(100 \times 0.2 + 10 \times 0.8)$ which equals 28 units of good.

When people choose to advocate act-consequentialism, they are typically motivated by a strong intuition which we shall call the *fundamental consequentialist intuition*:

> The only thing that matters morally is how good an outcome is.

Later, we shall see that act-consequentialism does not have sufficient resources to reply to its critics, but that a simple extension can mount such a reply. Moreover, it can do so without deviating from this fundamental consequentialist intuition. Until we return to the topic of different forms of consequentialism, we shall simply use 'consequentialism' to refer to 'act-consequentialism'.

1.2 Two major concerns for consequentialist theories

A common complaint about consequentialism is that it is *impractical*.[5] When deciding which act to perform, consequentialism only tells us that we should perform whichever one leads to the most good: there are many situations in which this is not immediately helpful. In order for an ethical theory to be 'practical' or to 'guide our actions', it would need to tell us which act is right and to tell us this under a description that allows us to recognise the right act. For example it would be acceptable if it told me that I should keep my promise, or that I should donate half of my income to charity.

In order for us to use the description that consequentialism provides ('the act which will lead to the most good'), we need to know which act that is. This involves knowing a considerable amount of information about the future and is frequently beyond our ability. A typical example is that of being in a position to save the life a young child who will later grow up to be a ruthless dictator. On a consequentialist account it is wrong to save this child's life even though there is no way that we could be expected to determine this for ourselves at the time. In light of such considerations, it appears that consequentialists must be in a state of perpetual ignorance concerning which of their acts are right.[6]

Suppose however, that we nevertheless attempt to produce as much good as we can. We consider the possible futures that our acts could lead to, estimate the relative likelihoods of each possibility and calculate the expected good of each available act to some degree of precision. We could thus go about doing good where we can by using this earthly approximation to the omniscient criterion of rightness. Unfortunately, such a strategy of

---

[5] See, for instance, Singer (1977), Smith (1988), Lenman (2000). Note that there are several different types of impracticality which consequentialism is alleged to possess. Here I have just sketched the most common criticism.

[6] We are, here, talking of rightness in the objective sense. The subjective sense will be discussed at length in the next chapter.

action is known to be self-defeating.[7] There are many occasions on which it would predictably lead to worse consequences than other ways of making our moral decisions, such as via the familiar dictates of what Sidgwick calls *common-sense morality*. For example, we often don't have the time needed to perform any of these consequentialist calculations and the best outcome can only be achieved by following moral rules, such as 'don't lie' or 'treat others as you would have them treat you'. In such cases the fundamental aim of consequentialism (that things go as well as possible) is worse achieved when we are aiming directly at it.

1.3 A consequentialist reply

Consequentialists typically reply to these concerns in terms of rules-of-thumb, or 'decision procedures'. They say first that when we act, we are not always guided by the consequentialist criterion of rightness. Instead, we are guided by other rules, principles or procedures which lead us from our present information to our actions. These methods allow us to act and produce good outcomes. These consequentialists thus treat the criterion of rightness like physicists treat the fundamental physical laws: these laws govern the facts of the world (which acts are right, or where the planets are located at a given time), but it may not be expedient for us to use them to calculate these answers. Instead, say the consequentialists, we should be guided by whichever rules or principles will lead to the best results.[8]

Thus, in the second chapter of his essay *Utilitarianism*, Mill states:

> 'To inform a traveller respecting the place of his ultimate destination, is not to forbid the use of landmarks and direction-posts on the way. The proposition that happiness is the end and aim of morality, does not mean that no road ought to be laid down to that goal, or that persons going thither should not be advised to take one direction rather than another.'[9]

We can see that this goes some way to countering the impracticality objection because it denies the assumption that we need always appeal to the consequentialist criterion in decision making. This reply also provides the basis for a defence against the objection that consequentialism is self-defeating. There are many ways for us to decide what to do, and they will lead to different outcomes. We shall call such a way of deciding what to do a *decision procedure*, be it a precise sequence of steps or an intuitive process which is difficult to articulate.[10] One of these decision procedures involves calculating the detailed consequences

---

[7] Derek Parfit (1984) distinguishes several different ways in which a theory can be self defeating. In his terms, the claim here is that consequentialism is sometimes *indirectly collectively self-defeating* and also *indirectly individually self-defeating*. This manner of being self-defeating is said to be indirect because problems only arise when the agent(s) aim at following the theory and, as we shall see, a theory need not require this.

[8] Bales (1971) provides a very good account of the need for this distinction and the confusion that has arisen in its absence.

[9] Mill (1861), Ch. 2, p. 276.

[10] Later on we shall spend considerable time exploring the nature of decision procedures, but for now we leave it as an intuitive concept.

of our alternative actions and choosing the act that leads to the best outcome. Let us call this procedure *naïve calculation*. Why should we follow this decision procedure when there are many others to choose from and this one can be seen to be detrimental in certain cases? It seems that we should instead follow the best decision procedure: whichever one it is that leads to the best outcome. Since consequentialists are not required to follow naïve calculation, the objection that it would be self defeating to do so becomes irrelevant.

Thus, as early as 1832, John Austin asserted:

> '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.'[11]

In *The Methods of Ethics*, Sidgwick makes the point explicitly:

> '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.'[12]

These quotations are by no means exhaustive. A great many consequentialist authors have explicitly made this appeal to our being obliged to follow the best decision procedure rather than *naïve calculation* and, to the best of my knowledge, no prominent consequentialists have claimed otherwise.[13]

One way to highlight the distinction being made here is to distinguish between two questions:

'What should I do?'[14]

'How should I decide what to do?'

The act-consequentialist criterion of rightness is a reply to the first of these questions, determining the act an agent should perform in any particular circumstance. The second of these questions, however, concerns the way in which we are to make our decisions and thus it does not judge acts themselves, but the ways in which an agent can make her decisions. It is typically assumed by deontologists that the answer to the second question is intimately connected to the first. For example, that the obligatory actions are those that do not break any of a certain set of rules and that we should decide what to do by testing our actions against these rules.

---

[11] Austin (1832), p. 108.

[12] Sidgwick (1907), p. 413.

[13] See, for instance: Mill (1861), Ch. 2; Sidgwick (1907), p. 413, 489–90; Moore (1903), p. 162–4; Smart (1961), § 7; Bales (1971); Hare (1981), Ch. 4; Parfit (1984) pp. 24–9, 31–43; Railton (1984), pp. 140–6, 152–3; Pettit and Brennan (1986), Hooker (2000). I shall address their individual approaches in Chapter 4.

[14] Here, and throughout the essay, I am using the term 'should' in its moral sense. That is, I use it synonymously with 'ought', expressing a moral obligation.

However, for consequentialists it seems much more promising to answer both of these questions through an appeal to the fundamental consequentialist intuition: I should do whichever act leads to the most good and I should decide what to do in whichever way leads to the most good. In other words, I should follow whichever decision procedure it is that leads to the most good. Perhaps this decision procedure involves inculcating certain virtues in myself and acting as they dictate, or perhaps it involves reflecting on whether the maxim behind a proposed act can be universally willed. Perhaps the best decision procedure is a small modification of common-sense morality, or a method that no moral theorist has yet discussed.

This essay is largely an attempt to flesh out this common consequentialist reply that we can overcome the alleged problems of impracticality and self-defeatingness by paying due consideration to the role of decision procedures. In doing so, I shall make considerable reference to the question 'How should I decide what to do?' which, due to its present imprecision, I shall call *the rough question*. While there is no doubt that this is a fundamental ethical question, it has largely been obscured due to its apparent similarity to the question of 'What should I do?'. This has been of particular detriment to consequentialism because, unlike in other ethical theories, the answers are so different. I shall thus be expanding considerably upon the brief statements that consequentialists have made on this issue, subjecting their views to new criticism and providing a much fuller account of the rough question and how consequentialists are to answer it.

In the next chapter, I will demonstrate just why it is that a decision procedure account is called for. The first part of this will be an exploration of naïve calculation and how it can predictably fail to produce good outcomes. I shall then introduce the distinction between the subjective and objective senses of rightness and show that, while related to the task at hand, this distinction cannot be used to answer the rough question.

In Chapter 3, I shall look at ways in which a consequentialist can coherently talk of rightness of decision procedures alongside rightness of acts. I will first argue that act-consequentialism does not have sufficient resources to judge the rightness of decision procedures as well as the rightness of acts. In effect, this means that the rough question is beyond the scope of act-consequentialism. I will then consider its well-known rival, rule-consequentialism, and conclude that this too fails to provide the necessary account. Instead, I shall advocate the theory that Pettit and Smith call *global-consequentialism*,[15] and show how this can incorporate an account of rightness for decision procedures, while upholding the fundamental consequentialist intuition.

In Chapter 4, I shall explore many different interpretations of the rough question. I begin with a survey of the views of prominent consequentialists. I then take the formulation of the rough question in terms of decision procedures ('Which decision procedure should I follow?') and analyse it in two stages. The first concerns the nature of decision procedures themselves. The second concerns the role in which they are to be assessed, or in other words, what it means to *follow* a given decision procedure. I conclude that the most plausible analysis of the

[15] Pettit and Smith (2000).

rough question involves the decision procedure that it is best for me to *accept*, rather than that which it is best for me to *comply with* or to *execute flawlessly*.

In Chapter 5, I shall address several objections to the position that I have presented. Foremost among these is the objection made by Brad Hooker and others that the decision procedure approach is inconsistent because it involves an obligation to follow a decision procedure even though it will lead us at some points to perform acts that we ought not do. I also address two other objections one of which concerns the existence of decision procedures whose benefits are merely accidental and the other of which concerns a potentially damaging regress.

Finally, I shall conclude that the question of how we are to decide what to do is more important to consequentialism than its historical treatment would suggest, and that the analysis herein provides a detailed account of the issue which has been lacking in the consequentialist literature.

Chapter 2

The need for a decision procedure approach

In this chapter, we shall achieve two aims. Firstly, we shall see exactly how it is that naïve calculation leads to predictably worse outcomes than other decision procedures, and thus why consequentialists require a means of assessing the ways in which we could make our decisions. We shall then briefly introduce the distinction between subjective and objective senses of rightness and show that, while it addresses similar concerns to the decision procedure approach, it cannot be used to suggest an alternative to naïve calculation or to answer the rough question.

2.1 The problems with naïve calculation

2.1.1 The 'paradox' of hedonism

We all, to greater and lesser extents, desire pleasure and seek it out. However, it is fairly well known that aiming to increase one's pleasure can sometimes predictably lead to acquiring less of that pleasure. Sidgwick termed this the *paradox of hedonism*:

> 'Here comes into view what we may call the fundamental paradox of Hedonism, that the impulse towards pleasure, if too predominant, defeats its own aim.'[16]

> '…it seems true that Happiness is likely to be better attained if the extent to which we set ourselves consciously to aim at it be carefully restricted.'[17]

The 'impulse towards pleasure' defeats its aim if it leads to the acquisition of less pleasure overall than would have been acquired without it. This phenomenon can be clearly seen in many cases and in varying strengths. In particular, we shall see that the stronger of these cases show that naïve calculation — as applied to maximising one's own pleasure — can be self defeating. That is, we can see that other decision procedures, which refrain from

[16] Sidgwick (1907), p. 48.

[17] Sidgwick (1907), p. 405.

calculating in certain cases, are superior. Unsurprisingly, this will have significant implications for naïve calculation as applied to consequentialism and maximising the objective good.

The paradox of hedonism typically manifests when the maximisation of pleasure requires one's mind to be used in a way that is incompatible with calculation. For example, the enjoyment gained from peaceful relaxation, meditation or quiet reflection is incompatible with active calculation. If we are always calculating how to achieve the most pleasure, then the pleasures inherent in such tasks are unavailable to us. Similarly, if we are to watch a film or read a novel then we mustn't be constantly thinking about whether we are enjoying it (or whether it would be best to stop and do something else) otherwise we will never get sufficiently immersed to enjoy it. The same is true for spending time with a friend or going to a party: constant questions about the quality of our enjoyment diminish it.[18]

There are also cases where we need to be in a state of mind that is incompatible with calculation, but where this state of mind is not itself pleasurable. For example, we may be driving a car in difficult circumstances or trying to remember our lines in a performance and if we were also calculating the relative benefits of other courses of action (stopping the car, improvising some lines) then we may quite predictably fail at the task at hand. Even if this does not lead to a loss of pleasure at the time, it may well prevent much pleasure later when we have to pay to repair the car or do not get cast in the next play. There are a great many cases like this, most notably that of falling in love. If we are always weighing up faults and virtues, calculating whether or not we should pursue a relationship, then it is unlikely that we shall ever fall deeply in love and experience all the pleasure that this brings.

A somewhat different case is the so-called 'tyranny of choice'. When going to the supermarket, we are frequently confronted with a multitude of choices for every trivial option. Which type of toothpaste should I buy? There are many types of toothpaste which each claim their own different advantage. If we are to calculate the best option from the available evidence in all such cases, then we will spend our lives paralysed by such trivial choices. Much better in such a situation would be to *satisfice*: to accept the first choice which was sufficiently good and then move on.

Related to this are those cases in which the time we have to choose is very small. If we are heading out to meet friends and see the tram pulling into the stop while we are still a block away then we have two options: to run for the tram or to keep walking and catch the next one. However, if we spend the time to calculate which is best then the tram will leave and we will deny ourselves the possibility of running to catch it.

Finally, consider the following case:

> *The race.* Thomas has enrolled in a long distance footrace and expects to do quite well. His enthusiasm helps him through the final stages and he has a very enjoyable run, coming third overall. As it happens, if he were to have carefully considered his

[18] See Pettit and Brennan (1986) for a detailed discussion of this and many other cases in which calculation is incompatible with pleasure or other things we might judge to be good.

> knowledge of his competitors he would have realised that they all had better form than he did and he would have become discouraged. This would have made him unable to keep running so strongly in the final stages and he would not have done very well.

If Thomas had been performing the calculations as to whether running in the race would maximise his pleasure, he would have had to consider his prospects of finishing well. However, this would have led to a poor performance and a decrease in pleasure. In the case of competitions, this effect is often present and quite predictable in nature. Having optimism untarnished by reasoned prediction can frequently lead to a better performance and, at the very least, a more pleasant experience.[19]

2.1.2 The 'paradox' of benevolence

While the cases above have been framed in terms of pleasure, the underlying principle can easily be seen to affect other quantities that people might try to maximise. We have seen how naïve calculation can lead to lost races, missed appointments, wasted time, car accidents, unappreciated novels and failed relationships. In short, it can lead to a reduction in well-being on any measure we might consider. Furthermore, since well-being is a part of the total good on any plausible axiology, we have also seen how naïve calculation can lead to a reduction in the good.

The above examples were all focused on people aiming at their own pleasure or well-being and consequently lowering it (thus lowering the total good as well). Unsurprisingly, there are also examples where aiming at increasing the total good via naïve calculation can predictably lead to a reduction in the total good. We shall refer to the existence of such examples as the 'paradox' of benevolence.[20]

A frequently cited example concerns the special relationships between lovers or friends. Since a great deal of what matters in our lives is generated through such relationships, this is very important indeed. Pettit and Brennan put it thus:

> '[An] 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.'[21]

---

[19] In other cases the reverse is true: we do best to dampen our expectations so that disappointment will be lessened and excitement raised.

[20] The examples of the previous section can also be easily modified into true examples of the paradox of benevolence, because people may try to increase the overall good by trying to increase their own well-being and, by undermining this goal, ultimately undermine the overall good.

[21] Pettit and Brennan (1986), p. 450. See also Smart (1961), pp. 44–5 and Railton (1984).

Pettit and Brennan, also give the example of acting virtuously.[22] On a certain conception of the virtues, they involve forgoing a calculative attitude to decision making. Thus, the generous person responds to the needs of others without weighing up the relative costs. So too for courage, honour and integrity. If such virtues are worth having (because the axiology holds them to be good in themselves or as conducive to other goods), then they are another example of a good whose presence is incompatible with the strategy of naïve calculation.

A major problem for benevolent calculation concerns the cost of deliberation. If someone is drowning in the river and we pause to calculate, he will likely die. The only way to achieve the best outcome in such a case is through some easily applied rule, such as those found in common-sense morality. While this example is extreme, there are a great many cases in which the costs of the time spent calculating predictably outweigh the benefits. This may be either when the costs of calculation are extreme, as in the case of the drowning man, or when the benefits of calculation are very low, as in the case of toothpaste selection.

Furthermore, natural biases will affect benevolent calculation even more than in the case of pleasure maximisation. For while we are biased towards our near future over our further future, we tend to be even more biased towards ourselves over others. A practitioner of naïve (benevolent) calculation is unfettered by the standard moral prohibitions and so there are many opportunities for his own biases to lead him to steal or commit other crimes. This problem would be even more pronounced if everyone, or the great majority of society, were naïve calculators. While it is not clear that the resulting situation would be worse than our present one,[23] it takes little imagination to suggest decision procedures which improve upon naïve calculation by curtailing its scope and thus avoiding cases where bias is prevalent.

2.1.3 The prospects of the naïve account

The discussions so far regarding naïve calculation rest on an assumption which is in fact untrue. Naïve calculation is not, as it stands, a decision procedure at all. There are several key points on which it offers no answer and these gaps prevent it from being a fully specified decision procedure. The act-consequentialist criterion tells us that an act is right if and only if it leads to at least as much good as any other available act. The idea behind naïve calculation is that we should thus go about calculating which act this is — but how should we go about calculating it? How should we go about determining the available acts? How long should we spend acquiring evidence? How long should we spend on the calculation? All of these questions would need to be answered if this were really to be a fully specified decision procedure. No doubt one could attempt such a project of developing a particular form of naïve calculation. Indeed, the details could be specified in many ways,[24] but each of these

[22] Pettit and Brennan (1986), pp. 447–8.

[23] Smart (1956), pp. 175–6, points out that the benefits of having the world's governments (and their nuclear weapons) run by people devoted to benevolence would likely overwhelm the disadvantages. To this I would also add the good that would be done by the resulting massive increase in aid to the world's poor.

[24] To be plausible it would presumably need to rely on the common sense of the individual using it at many

pulls us away form the original formulation. Instead of *the* strategy of naïve calculation it would just be *a* strategy of naïve calculation and the intuitive benefit from matching the decision procedure with the criterion of rightness diminishes. The choices of how long to calculate and which possibilities exclude fracture naïve calculation into a cluster of related decision procedures, none of which has the lustre of the original.

When we consider the arbitrariness involved in choosing one particular version of the strategy, it becomes all the more clear how unlikely it is that one of these calculatory decision procedures would in fact be the best way of deciding what it is that we should do. There are a multitude of decision procedures available to us, some of which involve constant calculation, some of which combine calculation with other approaches and some of which eschew calculation altogether. Each of these decision procedures will lead us to a different possible future, with a different quantity of good. Why, out of all of these possibilities would we expect that a form of naïve calculation will give us the best results? With so many options, the burden of proof is on the proponents of a particular form of naïve calculation to show why their decision procedure is best. However, the examples of the paradoxes of hedonism and benevolence illustrate exactly why such a proof will not be found: naïve calculation is often predictably worse than other, more commonsense ways of deciding what to do.

What, then, are the prospects for consequentialists who wish to defend the naïve account of how one should decide what to do? They would be forced to accept that the way in which we should decide what to do leads to predictably worse outcomes than other options, undermining the fundamental consequentialist intuition. But why should we accept such a reduction in the good when we could use other methods for our decisions? A common argument in favour of consequentialism is that naïve calculation reflects a certain aspect of good moral decision making, but this does not seem to be sufficient justification. For one thing, naïve calculation is advocated because it allows us to maximise the good — something that is predictably failing to happen in the cases in question.

It is also said that consequentialism allows us to be flexible to the situation at hand in cases where we must adapt in order to promote the good: to do otherwise, it is claimed, would be some form of perverse rule worship. However, it appears that consequentialists who endorse naïve calculation would be guilty of this same mistake; not 'rule worship' perhaps, but 'decision procedure worship'. Their system would be adaptive at a basic level (with the decisions changing to reflect the circumstances), but inflexible at a higher level with the method of making decisions being constant regardless of circumstance. It is this lack of higher level adaptivity which puts the naïve account so much at odds with the aims of consequentialism.

The potential for flexibility at both levels is not a mirage, but is attainable if we answer the rough question in much the way that we answer the question 'What should I do?'. Thus, just as we should do the act that leads to the best outcome, so we should use the decision procedure that leads to the best outcome. This answer requires considerable development which it will receive in subsequent chapters.

---

junctures.

2.2 Objective and subjective rightness

Recall the (act-)consequentialist criterion of rightness:

> An act is right *iff* it will lead to at least as much good as any alternative act.

This criterion defines the rightness of an act in terms of the outcome to which it leads. Thus, the motives and beliefs of the agent involved do not enter into the determination of what is right, beyond their direct effect upon the outcome (we might be able to run faster when sufficiently motivated and so forth). This independence of rightness from motives and beliefs is somewhat at odds with many people's intuitions. To illustrate this, consider the following example (which will recur later in this essay):

> *Test X.* Disaster has struck: Alex's medical results have come back and he has scored a positive result on Test X. As is well understood, there are only two conditions which can lead to a positive result on Test X: condition *C* and condition *R*. Of these, condition *R* is very rare, accounting for only one case in ten, while the rest suffer from the much more common condition *C*. Both conditions are fatal within one week unless the patient takes the drug appropriate to that condition (conveniently named 'drug *C*' and 'drug *R*'). These drugs are cheap and readily available, but if the patient takes the wrong drug (or both drugs) then there will be no improvement and death will surely result. Unfortunately, there is no way of telling whether the patient has condition *C* or condition *R* and thus no way of telling which drug will save them and which will leave them to die. Alex's doctor, Felicity, knows all of these relevant facts and so whenever one of her patients tests positively for test X, she gives them drug *C*. At least that way, she reasons, almost all of them will be saved.

Associated with Alex's predicament is the following table of outcomes and their values:

	Condition *C*	Condition *R*
Prescribe *C*	1000	0
Prescribe *R*	0	1000

Table 1.

For the (actualist) consequentialist view of rightness, only one of these two columns is relevant. If Alex has condition *C*, then we look to the first column and see that it is much better if Felicity prescribes drug *C* and thus that this would be right. If he has condition *R*, then the second column tells us that only prescribing drug *R* will save his life and thus be the right act.

Now let us continue the example:

> *Alex's doom.* Following her method, Felicity gives Alex drug *C*, but unfortunately he was actually suffering from condition *R* and consequently dies within the week.

According to the criterion of rightness, Felicity's act of prescribing drug *C* was wrong. She should have prescribed drug *R* instead. This is quite counterintuitive. It shows that while

consequentialism may guide our actions to a degree, showing Felicity how to do the right thing in a high proportion of cases, it does not show her how to do the right thing all of the time. Consequentialism allows her to perform the wrong act through no fault of her own and this appears to go against our common conception of rightness.

With this in mind, Sidgwick introduced the notions of objective and subjective rightness. He stated that an act is *objectively right* if it is right according to the criterion of rightness above, whilst an act is *subjectively right* if the agent who performs it believes it to be objectively right. The idea then, is that the everyday conception of rightness divides into two. Our intuitions on rightness are sometimes linked to what will really happen ('I guess my decision was wrong after all…') and sometimes linked to what was believed to be the case ('It would have been wrong for Felicity to have prescribed drug *R*…'). This splitting of the conception of rightness has been maintained by many later authors[25] and is thought to help reconcile the harsh and ultimately unknowable objective rightness criterion of consequentialism with an account that considers the agent's limited knowledge of the world. However, it has since been argued that while there may indeed be a subjective and objective forms of rightness, Sidgwick's definitions of these will not do. For example, Frank Jackson has pointed out a particular type of case in which Sidgwick's criterion of subjective rightness gives unacceptable answers:[26]

> *The skin complaint.* Cindy has arrived at the office of her local doctor, Matt, looking for help with a serious skin complaint. This skin complaint is caused by one of two underlying conditions: condition *A* or condition *B*. There are three drugs available to treat those suffering from this skin condition: drug *A*, drug *B* and drug *S*. Drugs *A* and *B* are very risky, for if the drug corresponding to the underlying complaint is taken then the patient will be completely cured, but if the wrong drug is taken then they will be killed. Drug *S*, on the other hand, is much safer: it will simply provide partial relief to either condition. There is no way to tell which underlying condition is present in any particular case and the prevalence in society is roughly equal.

Associated with Cindy's predicament is the following table of outcomes and their values:

	Condition *A*	Condition *B*
Prescribe *A*	5	−1000
Prescribe *B*	−1000	5
Prescribe *S*	3	3

Table 2.

We would be justifiably shocked if Matt were to prescribe either drug *A* or drug *B*, as the additional benefit that these might bring Cindy is completely overwhelmed by the high (subjective) risk of death. Instead, we would hope that he would prescribe drug *S*. However, on Sidgwick's account, prescribing drug *S* turns out to be subjectively wrong. This is because

[25] See, for example, Parfit (1984), pp. 24–40, Gibbard (1990) pp. 42–3, Zimmerman (1996), pp. 10–20. While these authors hold a distinction between subjective and objective senses of rightness, they do not define it in precisely the way that Sidgwick does.

[26] Jackson (1991), p. 463.

Matt knows that if Cindy has condition *A* then prescribing drug *A* is objectively right, and if she has condition *B* then prescribing drug *B* is objectively right. Thus, he knows that regardless of the underlying circumstances, prescribing drug *S* will be objectively wrong. Since Sidgwick's account makes the belief that an act is right a necessary condition for the act being subjectively right, he must conclude that Matt's action is not subjectively right. However, this conflicts with many people's intuitions on the subjective sense of rightness. Surely due consideration of the scales of the benefits and burdens, as well as the chances that they are to occur makes it subjectively right (in an intuitive sense) for Matt to prescribe drug *S*.[27]

These intuitions can be accommodated by a change in the criterion of subjective rightness. Instead of an act being subjectively right if it is believed to be objectively right, we can say:

> An act is *subjectively right* iff it has the greatest 'subjective expected goodness' of any available act.[28]

The beliefs of the agent are taken into account here through the concept of subjective expected good. This is similar to objective expected good (defined earlier), but where the agent's degree of belief (or *credence*) in each outcome is used in the place of an objective probability. Thus, if Matt's credence that Cindy suffers from condition *A* is 50% and his credence that she suffers from condition *B* is also 50%, then the subjective expected good of prescribing *A* is $(5 \times 0.5 - 1000 \times 0.5)$ which equals −497.5. This is the same as the subjective expected good of prescribing *B*. However, the subjective expected good of prescribing *S* will be $(3 \times 0.5 + 3 \times 0.5)$ which equals 3. Since 3 is higher than −497.5, this account holds that it is subjectively right for Matt to prescribe *S*.

A further modification we might make to the account of subjective rightness is to consider the degrees of belief that the agent has reason to assign to the propositions, rather than the degrees of belief that the agent *actually* ascribes to the relevant propositions. Thus, on this modification, an agent is held to account for their epistemic commitments: they can do what maximises their subjective expected good and yet, due to their failure to form the appropriate beliefs given their information, their act may still be subjectively wrong. This revised account of subjective rightness is now the majority position amongst those who recognise the subjective/objective split.[29]

The distinction between subjective and objective rightness thus serves a similar role to that of the decision procedure approach which will be the main focus of this essay. We have seen that because of the subjective probabilities, it was subjectively right for Felicity to prescribe *C*

---

[27] Even if we modified Sidgwick's account (as is often done) so that an act is subjectively right iff the agent has reason to believe that it is objectively right, this problem remains. We shall thus pass over such accounts here.

[28] There is a further question here, of which axiology should be used: the true axiology, the axiology the agent believes is correct or the axiology that the agent has reason to believe is correct. I shall not be able to explore this point here.

[29] See Parfit (1984), p. 25 and Gibbard (1990), pp. 42–43.

to Alex and for Matt to prescribe *S* to Cindy, despite both these acts being objectively wrong. Similarly, we could say that the best decision procedure for Felicity to follow involves her prescribing *C* to those that test positive to Test X and the best decision procedure for Matt to follow involves him prescribing *S* to those that suffer from the skin complaint. While their acts are objectively wrong, they are consequences of following the best decision procedures.

However, there is a fundamental difference between these two ways of making consequentialism more practical. The appeal to subjective rightness provides an alternative method of assessing acts based on the agent's beliefs. In contrast, the decision procedure approach assesses the ways in which an agent might make his decisions, and does not modify the existing manner in which the resulting acts are assessed. Because the decision procedures are assessed in light of all relevant facts, this approach provides an account of the *objective* rightness of decision procedures, rather than an account of subjective rightness of acts.

When consequentialists suggest that the failure of naïve calculation implies that we should use whichever decision procedure leads to the best outcome, they are making a claim about the objective rightness of decision procedures — a claim about how we should decide what to do. It may turn out that this best decision procedure leads to the agent performing a subjectively right act, but there is no necessary connection since the calculatory costs of weighing in all the relevant beliefs may be too great. We shall therefore consider only objective rightness of acts and decision procedures in the remainder of this essay.

Chapter 3

Rightness for acts, rightness for decision procedures

The rough question asks 'How should I decide what to do?'. I wish to explore the answer given by many act-consequentialists: that one should decide what to do in whichever way it is that leads to the best outcome. I have defined a decision procedure as a way of deciding what to do, so we can paraphrase this answer as saying that one should decide what to do according to the decision procedure that leads to the best outcome, or that the right decision procedure for someone to follow is the one that leads to the best outcome. Defenders of this answer thus have terms of moral approbation ('should', 'right') applied to decision procedures as well as to acts. Since the act-consequentialist criterion of rightness is defined in a way that only makes reference to assessing individual acts, it would appear that it must be extended in order to meet these demands.

In this chapter I shall first argue that act-consequentialism (as it is typically understood) does need to be extended. I shall then introduce rule-consequentialism and show that, while it can perhaps provide the two forms of assessment that we are looking for, it does so at to great a cost to what I have been calling the fundamental consequentialist intuition. I thus introduce global consequentialism and show that it appears to offer exactly what we need: it elegantly extends act-consequentialism to deal with the rightness of decision procedures and does so without offending the underlying consequentialist intuitions.

3.1 Act-consequentialism and decision procedures

Recall the act-consequentialist criterion of rightness of acts:

> An act is right *iff* it will lead to at least as much good as any alternative act.

What, if anything, does this principle tell us about which decision procedure a person should follow or which decision procedure is the right one to follow? On the surface, it appears to talk only of the rightness of acts and to say nothing at all about the rightness of decision procedures. However, there are a couple of reasons to believe that this might be mistaken.

When Henry decides to stand up, he has done something even before he moves a muscle. He has made a decision, and this would appear to be a kind of act. It is not a physical act like standing up, or going for a walk, or even breaking a promise. Instead, it is a *mental act*, like imagining a perfect circle or counting to ten in your head. If someone accepts this, then they can say that since decisions are acts, act-consequentialism assesses them. Henry's decision to stand up would thus be right if and only if it (the *decision* to stand up) led to a better outcome than any other act he could have performed.

Those who hold this view might go further and try to use act-consequentialism to assess the way in which the decision was made. There are several ways in which Henry could have made his decision: he might have accepted his first impulse, he might have determined that standing up was essential to getting himself a drink, or he might have calculated the benefits of standing up over remaining seated and found that the former won out. When we are talking of physical acts, we may consider that different ways of doing something are different acts. For example, if Michael can hit the ball hard or hit it softly, then we could say that he has (at least) two different acts available to him. Similarly, we might say that the different ways in which Henry could decide to stand up constitute different mental acts available to him. Thus we could say that Henry was right to decide to stand up by accepting his first impulse if and only if this was the best available alternative. If we accept all of this, we could say that act-consequentialism thus tells us the right way to decide what to do.

There are many problems with this approach. There are difficulties involved in an account of mental acts and more difficulties involved in individuating different ways of deciding to do a certain thing as different acts. There are also difficulties in having the act-consequentialist criterion comparing physical acts with decisions to do these acts, and there are potential regresses concerning deciding to decide in a certain way.[30] However, even if these can all be worked out and act-consequentialism is in fact committed to assessing different ways of deciding what to do, this still falls short of our aims.

The proposed act-consequentialist assessment of decision-making is restricted to the consequences of this one decision. Many factors would go into determining the best way for Henry to make all of his decisions — the best single decision procedure to follow throughout his life — or his decisions over a period of his life, or in one domain of his life. However, on the interpretation above, act-consequentialism just determines the best way for Henry to decide this one thing. Let us assume that it is indeed best for him to stand up. What is there to recommend one way of deciding this over another? Consider the purpose-built decision procedure: 'If seated, stand up immediately'. If Henry decides to stand up by using this decision procedure, he will indeed stand up and will have wasted very little time in the decision making process. Indeed, if speed is of any importance in the situation, it seems that deciding to stand up using this decision procedure is superior to deciding to stand up according to a more generally useful (and thus more complex) decision procedure. In such cases, we can always find a very good way of deciding what to do by creating a decision procedure for the purpose of making the agent choose this particular action. However, this

[30] See, for instance, Smith (1991).

appears to have little relevance to study of morality. What we were originally interested in was the best way of making decisions in general (something that requires some flexibility), not the best way to make one particular decision.

For example, consider again the case of *Test X*, in which Felicity does not know whether to prescribe Alex drug *C* or drug *R*. Since on this occasion it is much better for Felicity to prescribe drug *R*, this act-consequentialist account of decision procedures would recommend that she decided by following a decision procedure such as 'always prescribe drug *R*'. Later, when she meets a patient with the underlying condition *C*, she should decide what to do by following either naïve calculation or something like 'always prescribe drug *C*' (which will get the decision made faster with no chance of error). The intuitive account of decision procedures when applied to such a situation seems plausible precisely because it judges the effects of the agent's decision procedure over some longer period of time in which it will have to deal with the many different circumstances the agent will face. In contrast, this act-consequentialist account of assessing decision procedures on single instances does not answer any plausible interpretation of the rough question is insufficient for our needs.

A separate account of how it is that act-consequentialism must judge decision procedures comes from Gerald Lang.[31] He begins with a restatement of the act-consequentialist criterion:

> (3) An agent ought to act in such a way as to produce at least as much good as that which would be produced by any other act available to her.

He then introduces two further premises:

> (4) An agent would not be able to produce acts which produce at least as much good as any other act available to her unless she is guided by some or other decision-making procedure, D*.

> (5) If an agent ought to secure X, and if doing Y is the only means of securing X, then the agent ought to do Y.

From these he draws the conclusion:

> (7) An agent ought to be guided by D*.

The first thing to say regarding this argument is that premise (5) may well be unpalatable to many act-consequentialist, as it means that from a claim that an agent ought to do *A* and *B* we can derive that the agent ought do *B*. This derivation is denied by *actualist*[32] act-consequentialists and thus the inclusion of premise (5) begs the question against this sizable

[31] Lang (2004). His argument is originally aimed at act-utilitarianism, not act-consequentialism, and so I have simply substituted occurrences of the term 'utility' with 'good'.

[32] Used here in contrast to *possibilist*. See Jackson and Pargetter (1986).

school of consequentialists. However, let us grant Lang premise (5) and see what other troubles the argument runs into.

It seems to me that we then run into one of two problems: either the conclusion cannot be reached due to the falsity of (4), or that the conclusion must be interpreted in a manner that is much weaker than it first appears. Premise (4), as stated, is somewhat ambiguous. Does it refer to a single decision procedure being used to make all the acts or does it allow for different decision procedures for each choice of act? Let us first assume the former and restate the premise for clarity:

> (4a) An agent would not be able to produce a sequence of acts, each of which is optimific, unless she is guided in all of them by some or other single decision-making procedure, D*.

From here Lang might hope to get to the (more precise) conclusion:

> (7a) An agent ought to be guided in her sequence of acts by some or other single decision-making procedure, D*.

However, this cannot be achieved, since (4a) is false. Lang states that (4) 'trades on little more than the innocuous claim that, in order to act, we need to employ some or other decision-making procedure'[33], and while this claim may well be innocuous, (4a) requires much more. It requires that to act rightly on a whole sequence of acts, a single decision procedure must be used, but it is not clear why we should think so. Perhaps the first few right acts are performed and then the agent switches to a new decision procedure and performs the last ones. If this were possible (and surely it is), then (4a) is false. Lang's claim would only be innocuous if we were instead to interpret (4) as:

> (4b) An agent would not be able to produce the optimific act in any particular situation unless she is guided by some or other decision-making procedure, D*.

This may well be true, and it can be used in conjunction with (5) to reach a conclusion:

> (7b) In any particular situation, an agent ought to be guided by some or other decision-making procedure, D*.

However, in (4b) and (7b), the decision procedure does not need to be the same across all choices the agent makes. In a given situation, there are many decision procedures that would guide the agent and we could grant that she indeed ought be guided in her decision by one of them. However, at the next decision, there is another list of decision procedures that would appropriately guide the agent and she need not use the same one again. Thus, the conclusion (7b) would merely be saying that each time an agent chooses she ought to be guided by one of the decision procedures that will lead to the right choice in that case. We thus have another argument which suggests that act-consequentialism is committed to judging decision procedures to be right in particular situations (not a very useful thing in itself), but does not judge their merits over a series of decisions. In order to find an account that does this, we

[33] Lang (2004), p. 224.

shall thus look further afield.

3.2 Rule-consequentialism

There are many ways in which rule-consequentialism can be formulated. A simple definition is given by Smart:

> '[Rule-consequentialism] is the view that the rightness or wrongness of an action is to be judged by the goodness or the badness of the consequences of a rule that everyone should perform the action in like circumstances.'[34]

It is said that act-consequentialism is an instance of *direct consequentialism* because the rightness of an act is determined directly by its consequences as compared to those of the alternative acts.[35] In contrast, Smart's formulation of rule-consequentialism is an instance of *indirect consequentialism* because the rightness of an act is determined not by the consequences of the act itself, but by the consequences of a rule that everyone should perform the act in like circumstances. By allowing this layer of indirection, rule-consequentialism allows for a theory with consequentialist underpinnings that appears to generate more intuitive moral claims while avoiding all of the concerns regarding naïve calculation. However, it also introduces several well-known problems.

The major criticism is that rule-consequentialism must either collapse into act-consequentialism or be incoherent. The first part of the argument can take one of several forms, but the basic idea is that the best rule will be to do what leads to the most good. Opponents of rule-consequentialism then argue that if it can somehow avoid such a collapse, it is at odds with the fundamental consequentialist intuition of promoting the good, since it sometimes tells us that it is right to follow a certain rule when doing so leads to a decrease in the good. More sophisticated formulations of rule-consequentialism, such as Brad Hooker's have therefore been introduced to attempt to address these (and other) criticisms:

> 'An act is wrong if and only if it is prohibited by a code of rules the acceptance of which by the overwhelming majority of people in each new generation would have the greatest expected value.'[36]

Hooker argues that rule-consequentialism does not collapse into act-consequentialism for a number of reasons. Firstly, he makes the point that rule-consequentialism is a theory which refers to rules that can form part of someone's actual decision processes. Thus, if we

---

[34] Smart (1961), p. 9.

[35] It is sometimes said (see Pettit and Smith (2000), and Kagan (2000)) that act-consequentialism is an indirect theory, assessing rules and other things in terms of the acts that they produce. I have instead been assuming that act-consequentialism assesses only acts and does so via its criterion of rightness. Nothing shall turn on this as it is merely a terminological issue.

[36] Hooker (2004), § 6. There is also a somewhat more complicated account in Hooker (2000), which refers to a specific axiology and includes rules for tie breaking.

internalised the rule 'always do what leads to the most good', we would not always do what leads to the most good, but would presumably follow a strategy of naïve calculation. Given the well-known problems with this strategy, he concludes that there would be other sets of rules whose internalisation would be less costly (by fitting more closely with common-sense morality) and for which the consequences of the vast majority of people following them would be better than that of following naïve calculation.

Rule-consequentialism, on Hooker's account, assesses two types of things: sets of rules, and acts.[37] The set of rules to be preferred is that set whose internalisation by most people in each new generation would lead to at least as good an outcome as the internalisation of any other set. Acts are then to be considered right if and only if they accord with this set of rules. While Hooker does not explicitly say that the preferred set of rules is the *right* set of rules or the *best* set of rules, it is clearly a set of rules which is accorded a particularly important position in his moral theory and is judged to be superior to all other sets of rules. It is also a set of rules with a particularly close relationship to obligation, for we *ought* to act in accordance with this set of rules and (to rephrase this) it is the set of rules that we ought to comply with. Given this close moral approval for this set of rules, I shall follow Pettit and Smith in saying that rule-consequentialism does in fact offer a criterion of rightness for both rules and acts, where the right set of rules is that set whose internalisation leads to the best outcome and the right acts are those that comply with the right rules.[38]

Rule-consequentialism thus offers an account of how we are to decide what to do; one that is based on consequentialist principles and yet does not simply advocate the strategy of naïve calculation. Moreover, it explains this with two compatible forms of rightness: we should decide what to do by attempting to follow the right set of rules and (assuming we succeed in doing so) the acts we perform will be judged as right.[39] We thus have an account which plausibly fulfils the aims of those act-consequentialists who wish to answer the rough question. However, I do not think that it quite succeeds.

Firstly, there is a question of whether a set of rules is a sufficiently general approach to decision making. There are many ways in which we could decide what to do: many formulae we could invoke, many ways we could calculate and many ways we could rely upon certain intuitions. If deciding according to a set of rules could not include all such ways of deciding, then we would have a problem, for it may be that deciding according to one of these ways that is not captured by rules will lead to a better outcome than deciding in any rule prescribed manner. We would thus have a good reason to abandon the rule based approach. Indeed, even if the best way to decide in the real world happens to be rule-based, this problem still remains at a theoretical level, for rule-consequentialism would then be relying on a supposition that is not *a priori* supportable, but which just happens to be correct.

---

[37] On other forms of rule-consequentialism it is individual rules, not sets of rules, which are the subject of moral approval.

[38] Pettit and Smith (2000). See also Kagan (2000).

[39] This notion of 'attempting to follow' a rule is imprecise, but will do for now. We will investigate this more closely in Chapter 4.

What are we to count as a rule? Hooker provides no precise definition, but appears to take a very general formulation. For example, he counts acting in accordance with naïve calculation as acting in accordance with a rule. Presumably the rule could be formulated as 'Always calculate expected utilities to such-and-such a degree of precision and then act upon them'. If a rule is allowed to refer to the performing of calculations then it seems that it would be sufficiently general to accommodate all decision procedures, and I shall assume this to be the case for the remainder of the section. Even if it is somehow shown there are decision procedures which cannot be accommodated, this is no major obstacle as one could always abandon rule-consequentialism for some form of *decision-procedure-consequentialism*, where all references to the internalisation of sets of rules are replaced with internalisation of decision procedures.

Another potential difference between the standard consequentialist reply and rule-consequentialism lies in the way that sets of rules are to be judged. For rule-consequentialists a set of rules is judged on the basis of how good it is for it to be internalised by everyone or by the overwhelming majority of the population. Rule-consequentialists are motivated towards this position by its similarity to commonsense morality (or Kantian ethics) in prescribing a universal set of rules which we must all attempt to follow. There are also strong similarities to a legal framework in which one set of laws governs all.

However, we may well question why rule-consequentialism judges rules on the assumption that almost everyone will use them. Instead it seems that the obvious consequentialist answer to 'How should I decide what to do?' is that I should decide what to do in whatever way would be best, *given that everyone else does whatever it is that they actually will do*. We know that our decisions to change the way in which we decide things will not mysteriously compel others to decide in the same way, so why should we judge the set of rules (or decision procedure) as if they would? The rule-consequentialists do have their reasons for their position and I am not trying to argue conclusively against it in this space. However, act-consequentialists who are merely looking for a way of assessing decision procedures are very unlikely to be swayed from their approach of assessing the consequences given what will actually happen.

Let us therefore consider a modified form of rule-consequentialism on which rules are assessed on an individual basis. Thus, we relativise the criterion of rightness for a set of rules so that it assesses whether a set of rules is right *for a given agent*. We then determine whether an act by an agent is wrong according to compliance with the rules that are right for that agent. This is a rather major modification of the theory (and one that would not typically be supported by the advocates of rule-consequentialism), but it serves the present purpose. Does this modified theory provide us with a good way of providing moral assessment of acts and decision procedures? I believe that it does not, and that the classical arguments for rule-consequentialism's incoherence show us why. On this modified view, rule-consequentialism is committed to the claim that it is right for an agent to perform a given act iff it is not forbidden by the set of rules which is right for her to internalise. However, this means that the rightness of an act will often come apart from its leading to the best outcome.

Suppose, for instance, that the best set of rules for Jim to follow include a rule stating that he should always tell the truth. There might be various exceptions to this built in to cover

disastrous consequences, but we can assume that to lie in a fairly typical situation when this would only lead to a small increase in the good will be forbidden. Consider such a situation in which the best outcome can only be achieved by Jim lying and where the benefits of this are rather small. If we share the fundamental consequentialist intuition that what matters is the goodness of the outcome, then it seems that — contrary to rule-consequentialism — it is (objectively) right for Jim to lie. Whilst this modified form of rule-consequentialism allows consequentialists to assess the way we make decisions, it does so at the cost of violating one of the major motivations for consequentialism itself. Hooker is correct in holding that this does not show the various versions of rule-consequentialism to be incoherent, but it is a sufficiently large flaw to make act-consequentialists who are looking for a criterion of rightness for decision procedures look elsewhere — especially given the existence of a much more favourable alternative.

3.3 Global consequentialism

A very different account of ascribing rightness to acts and to decision procedures is offered by *global consequentialism*. This is a form of consequentialism which focuses on outcomes being as good as possible and does not privilege a certain evaluand such as that of acts. Derek Parfit explicitly uses a form of global consequentialism in *Reasons and Persons*, and the concept is further developed and defended by Philip Pettit and Michael Smith.[40] They state that for any evaluand *x*:

> An *x* is right *iff* there is no alternative *x* which leads to a better outcome than it does.

Given that there is a unique right *x* (as we have been assuming for the sake of brevity), we can phrase this succinctly as:

> The right *x* is the *x* which would lead to the best outcome.

Pettit and Smith contrast global consequentialism with various *local* forms, which assess one type of evaluand in terms of its outcome and either do not assess other evaluands or assesses them in some way which does not involve their outcomes. Global consequentialism can be said to be a form of multi-dimensional direct consequentialism since any evaluand can be assessed directly.[41] It is perhaps easiest to think of it as a schema for producing specific criteria of rightness. Thus, looking at the evaluands: *act*, *disposition*, *climate*, *eye-colour*, we would get:

> The right act is the act which would lead to the best outcome.

> The right disposition is the disposition which would lead to the best outcome.

---

[40] Parfit (1984), pp. 24–8, Pettit and Smith (2000). See also, Kagan (2000) and Feldman (1975).

[41] This is how Kagan (2000) expresses it. He refers to his form of global consequentialism by the alternative label of *everywhere direct consequentialism* and abbreviates this to *direct consequentialism* (thus using this term in a slightly non-standard manner).

> The right climate is the climate which would lead to the best outcome.
>
> The right eye-colour is the eye-colour which would lead to the best outcome.

Note, however, that in their current forms these are slightly ambiguous. What does it mean for a disposition to lead to an outcome? Is it the possession of the disposition by a certain person, or by everyone? Is it the disposition that will lead to the most good on this occasion or are we assuming it to be possessed for the agent's whole life? Whilst the existence of such ambiguities is not explicitly dealt with by Pettit and Smith, it is clear that they can be removed by specifying the evaluand more precisely, both in terms of the relevant object (such as the act or disposition) and the role it is being assessed in (such as being performed by everyone or being possessed by an individual). For example:

> The right act for an agent to perform now is the act whose being performed now by that agent would lead to the best outcome.
>
> The right disposition for everyone to possess at all times is the disposition whose being possessed at all times by everyone would lead to the best outcome.
>
> The right climate for Oxford to have is the climate of whose being had by Oxford would lead to the best outcome.
>
> The right eye-colour for Carol to admire is the eye-colour whose being admired by Carol would lead to the best outcome.

As we can see, the peculiarities of natural language mask the underlying logical form, but we can still make it out if we look closely. Indeed, at least in the above cases, we simply have an object-type *x* being assessed in connection with a subject (or subject-type) *y* and a verb (or verb phrase) *z*:

> The right *x* for *y* to *z* is the *x* whose being *z*-ed by *y* would lead to the best outcome.

When it comes to ways of deciding what to do, we could start by saying:

> The right way of deciding what to do is the way of deciding what to do that would lead to the best outcome.

Or, in the language of decision procedures:

> The right decision procedure is the decision procedure that would lead to the best outcome.

We can then assess the decision procedure in a certain role:[42]

> The right decision procedure for an agent to accept is the decision procedure whose being accepted by that agent would lead to the best outcome.

---

[42] The choice of role will occupy a significant portion of the next chapter.

> The right decision procedure for an agent to comply with is the decision procedure whose being complied with by that agent would lead to the best outcome.
>
> The right decision procedure for an agent to flawlessly execute is the decision procedure whose being flawlessly executed by that agent would lead to the best outcome.

While rule-consequentialism is set up so as to have one privileged formulation of rightness for sets of rules and a single derivative criterion of rightness for acts, global consequentialism can support many different formulations at once. Since the three formulations of rightness for decision procedures above all fit the schema, they are all valid criteria for rightness of decision procedures. For a global consequentialist, the question of whether or not a decision procedure is right depends intimately upon what one is to do with it. Thus, if we want to know which decision procedure it is right for an agent to accept from now on, we see (as above) that this is the decision procedure whose being accepted from now on leads to the best outcome. Different decision procedures would be right for different purposes, as are different climates or different eye-colours.

Global consequentialism thus avoids the need to carefully design the criterion of rightness. Hooker's version of the rule-consequentialist criterion of rightness has many intricate points (rules are to be internalised, it is to be the overwhelming majority of people, we are to count the costs of internalisation…) and he must defend all of them. In contrast, the global consequentialist criterion is very simple, intuitive and general. The intricacies are not built into it, but come within the questions we ask. The lack of specificity means that we cannot just ask which rules are right, but could, for instance, ask which rules it would be right for the overwhelming majority to internalise or which rules it would be right for everyone to internalise. Whichever (sufficiently precise) question we ask, global consequentialism gives us an answer and it is up to us to choose our questions carefully. This gives global consequentialism a very different flavour to that of act-consequentialism or rule-consequentialism and we shall see this in some detail in the next chapter.

With its ability to assess acts and decision procedures (along with everything else), global consequentialism seems to be the perfect basis for consequentialists to answer the rough question. Unlike even the most favourable interpretations of rule-consequentialism, it allows us to assess both decision procedures and acts based purely upon the goodness of their consequences. It thus does not offend the fundamental consequentialist intuition, and indeed appears to be a very obvious way of using this intuition to produce a theory of rightness. Compared to global consequentialism, act-consequentialism can be considered to be a strict *sub-theory*, for it assesses a sub-domain of evaluands (the right acts for an agent to perform at a given time) and judges these in the same way. The historical focus on act-consequentialism would appear to stem from a belief that acts are the only sources of good or that acts are the only sources of good over which we have control.[43]

---

[43] I do not mean to claim that all those moral theorists who are typically termed act-consequentialists really subscribed only to the act-consequentialist criterion of rightness, eschewing all assessment of rules or principles or decision procedures. However, even if they had global consequentialist leanings, these were

This position is indeed worth some attention. For example, consider the decision procedure whose adoption by an agent will lead to the most good. While act-consequentialism does not call this decision procedure 'right' it does respect the goodness that will come of the decision procedure's adoption. Acts which lead to the agent adopting the decision procedure will lead to a great deal of goodness and, barring conflicting ways of acquiring even more good, these acts will be judged right. Thus, while the decision procedure is not itself considered to be right, there is a high chance that the acts which lead to its adoption are.

However, act-consequentialism, on this formulation, is lacking in moral expressiveness. 'How should I decide what to do?' is a reasonable moral question and one of considerable importance, yet act-consequentialism cannot answer it.[44] Act-consequentialism can tell us that deciding in a certain way leads to the best outcome, but so too can deontological theories. This alone is an important feature of the decision procedure, but need not be a recommendation for its use. On the other hand, global consequentialism can make such moral recommendations. It tells us that we should not follow the strategy of naïve calculation, but should instead follow whichever decision procedure would lead to the most good.

While global consequentialism is clearly very promising when it comes to providing an account of the rough question, it has been the target of several recent objections. At the end of their paper, Pettit and Smith put forward the challenge: 'If consequentialism is a sound strategy of evaluation in any one case, then it is hard to see why it should not represent a sound strategy in every case.'[45] Bart Streumer responds to this by pointing out a potential inconsistency between global consequentialism and the principle that 'ought' implies 'can'.[46] When act-consequentialism judges an act to be right, it implies that the agent in question ought to perform it.[47] However, it seems that the rightness in global consequentialism is not tied to 'can' in the appropriate way. For example, global consequentialism appears to specify the right time for me to become invisible even though I cannot become invisible. In addition, the best climate for Oxford to have may be a tropical climate, but it is not clear that Oxford *can* have a tropical climate: it would be *physically* possible, but is this the appropriate kind of possibility?

Such fears can be averted by appealing to a set of alternatives. The act-consequentialist criterion determines the best of the available acts for the agent to perform, where this set consists of all the acts that the agent *can* perform. We shall adopt this same approach for global consequentialism. From now on, when we say that:

---

not expressed anywhere near as explicitly as the theory has been in recent times.

[44] We have seen that act-consequentialism may be able to answer it in terms of deciding *on this occasion*, but it cannot answer the rough question in its more general sense.

[45] Pettit and Smith (2000), p. 132.

[46] Streumer (2003).

[47] As before, I am assuming that in these cases there is a unique right act. If there are two acts that lead to optimific outcomes, then they are both right, but neither is obligatory. We can, as before, assume that this is not the case to avoid unnecessary complexity in the analysis.

> The right *x* for *y* to *z* is the *x* whose being *z*-ed by *y* would lead to the best outcome.

We shall understand the *x*'s to be limited to those *x*'s that *y* can *z*. These shall make up the alternative set of *x*'s. Thus the right climate for Oxford to have is that particular climate which can be had by Oxford and whose being had by Oxford would lead to a better outcome than that of any other climate that Oxford can have. If Oxford can't have a tropical climate, then this prevents a tropical climate being the right climate for Oxford to have and prevents other climates for Oxford being judged against a tropical climate. Similarly, if there is no time at which I can become invisible, then there is no right time for me to become invisible.

There is another concern with global consequentialism, which has not been discussed in the literature. Consider the following application of global consequentialism: the right person for Julian to kill is the person whose being killed by Julian leads to the best outcome. Let us suppose that of all the people Julian can kill, the best outcome will be achieved if he kills Helen. We then say that the right person for Julian to kill is Helen. This is a correct application of the global consequentialist criterion, and shows that its use of 'best' to derive 'right' can lead to apparently troubling claims. However, it has not been claimed that it is right for Julian to kill Helen, merely that the right person for Julian to kill is Helen. To get that further claim, it would be necessary to show that it is right for Julian to kill someone. We can resolve this by claiming that:

> If it is right for *y* to *z* an *x* and the right *x* for *y* to *z* is *w*, then it is right for *y* to *z w*.[48]

This makes it explicit as to when we can move from the standard (and implicitly conditional) claims of global consequentialism (the right *x* for *y* to *z* is *w*) to the more traditional form (it is right for *y* to *z w*). We do so just when that implicit condition is discharged: when it is right for *y* to *z* an *x*. We thus take the forest of guarded claims like 'the right person for Julian to kill is Helen' and only produce the direct claims like 'it is right for Julian to kill Helen' when they are appropriate.

A final objection to global consequentialism concerns the possibility of conflict between the various ascriptions of rightness. For example, Hooker suggests that there might be such a conflict between the decision procedure that it is right for an agent to follow and the acts that it is right for her to perform.[49] In particular, he suggests that following the right decision procedure will sometimes involve the agent performing an act that is wrong. This does not appear to be a *logical* inconsistency, but it certainly hints at a most unintuitive aspect of global consequentialism.

---

[48] Using this requires some analysis of when it is right for *y* to *z* an *x*, but this does not appear difficult. We shall use the intuitive condition that it is right for *y* to *z* an *x* iff there is some *x* whose being *z*-ed by *y* is better than *y* not *z*-ing any *x* at all. For example, it is right for Julian to kill a person iff there is some person whose being killed by Julian is better than Julian not killing any person at all.

[49] Hooker (2004), § 5.

However, before we can properly formulate and address this complaint, we will need to explore the nature of decision procedures themselves and the roles in which they can be assessed. We shall thus move on to these matters now, tentatively accepting a global consequentialist framework. Then, in Chapter 5, we shall discuss this *inconsistency objection* at length and explain how it can be resolved.

Chapter 4

Interpreting the rough question

I have been suggesting that we should interpret the rough question 'How should I decide what to do?' as 'Which is the right decision procedure for me to follow?'. In the previous chapter we have seen that for global consequentialists this has the same answer as the question 'Which decision procedure would it be optimific for me to follow?'. However, there is still considerable ambiguity here on two fronts. Firstly, there is the question of just how we are to understand the concept of a decision procedure. It is to be a way of deciding what to do — that much is clear — but what else can we say? Then there are a variety of roles that decision procedures could play in answering the rough question. For example, are we to ask for the decision procedure which it would be right to comply with, the decision procedure which it would be right to accept, or something else?

We begin resolving these ambiguities by examining the received views on the decision procedure account and seeing how far they can take us.

4.1 Prior approaches

4.1.1 Aims and motives

Whilst there is widespread agreement amongst consequentialists that agents should not make all of their decisions via naïve calculation, their alternative analyses of the issue are disparate and often lacking in specifics. One influential approach is to consider which *aims* or *motives* are best for the agent to use in deciding what to do. Thus, Sidgwick tells us:

> 'Finally, the doctrine that Universal Happiness is the ultimate *standard* must not be understood to imply that Universal Benevolence is the only right or always best *motive* of action. For, as we have before observed, 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.'[50]

Similar sentiments are expressed by Parfit, who uses the term 'motives' to refer to both desires and dispositions.[51] Parfit points out that it is quite possible that having the set of motives that compel one to always try to achieve the most good will predictably bring about less good than having a different, causally possible, set of motives. That is, being disposed to choose one's actions in some other way, such as according to the dictates of common-sense morality, may be better than being disposed to try to bring about the most good. Sidgwick and Parfit thus both analyse the paradox of benevolence in terms of features of the agent's psychology such as aims, desires and dispositions.

Note that unlike my earlier presentation of the paradox, this version needn't have any connection to naïve calculation. Suppose that Michelle is what Parfit calls a *pure-do-gooder*: she is motivated so as to always do what makes the outcome best.[52] However, Michelle is not an advocate of naïve calculation. Instead, she just follows her instincts, because she believes them to be divinely inspired. With the appropriate details filled in, Michelle could be someone who would achieve better outcomes if she ceased to be a pure-do-gooder and could thus be illustrative of the paradox of benevolence considered in terms of aims and motives. She could not, however, be used as an example of the paradox considered in terms of the agent's deliberative strategy.

It would thus appear that an agent's decision procedure is not determined solely by her aims and motives, and that we must look further afield for a full account of the right way of making decisions. Sidgwick and Parfit might reply in two ways. One the one hand, they could hold that aims and motives indeed do not completely determine decision procedures, but that they were not trying to provide such an account. Instead they were just explaining which aims or motives an agent should possess and, while the question of decision procedures might also be quite an important one, it is not what they had in mind.

Alternatively, they could hold that an agent's aims and motives *do* completely determine the way in which she makes decisions, but only when we have a complete specification of these aims and motives. Thus, as well as having a disposition to try to do the most good, an agent may be disposed to do so via naïve calculation or by following her instincts or by some other method. Such an approach may be tenable, although it would involve quite a broad conception of aims and motives, including both our typical 'high level' aims and motives (such as aiming to promote the good or being disposed to help others) and 'low level' aims and motives (such as aiming to calculate according to a certain procedure or being disposed to deliberate in a certain way).

---

[50] Sidgwick (1907), p. 413. The emphasis is his.

[51] Parfit (1984), pp. 25–8, 31–40.

[52] Parfit (1984), pp. 27–9.

4.1.2 Habits and predispositions

Smart's account of the paradox of benevolence *does* explicitly point to the futility of naive calculation, but his response to this is rather different to that which I have been advocating. He points to the fact that we do indeed 'habituate ourselves to behave in accordance with certain rules, such as to keep promises'[53], and explains how these 'rules-of-thumb' free us from perpetual calculation. However, he regards these rules-of-thumb, not as any kind of decision making procedure, but as some lesser cognitive process: a mere habit or disposition. Thus Smart writes:

> 'He acts in accordance with rules, in short, when there is no time to think, and since he does not think, the actions which he does habitually are not the outcome of moral thinking. When he has to think what to do, then there is a question of deliberation or choice, and it is precisely for such situations that the utilitarian criterion is intended.'[54]

Smart thus separates our cognitive, action-producing, powers so that on the one hand there are decisions or choices, which proceed by deliberation, and on the other hand there are instinctive, habitual responses via rules-of-thumb. For good consequentialists, both can be combined and used at the appropriate times. Indeed agents should even use correct consequentialist deliberation to see that they should inculcate certain habits within themselves.

Pettit and Brennan offer a very similar distinction in their paper *Restrictive Consequentialism*. Where Smart talks of habit, they talk of mental states of a certain type which they name *predispositions*:

> 'They are states whose manifestation in action means that the action is not chosen on a fully calculative or deliberative basis.'[55]

As in Smart's account, the alternative to naïve calculation is considered to be something that can 'pre-empt certain decisions', rather than forming a natural part of decision making.[56] However, on this account the predispositions aren't claimed to be completely distinct from deliberation, but are just not *fully* deliberative.

I agree with Smart, Pettit and Brennan that the selection of actions via unthinking rules-of-thumb or predispositions can be very useful. However, it seems that this position does not go far enough. For example, consider Railton's example where an all-knowing all-powerful demon makes things go very badly for anyone who does not make decisions as a Kantian

---

[53] Smart (1961), p. 42.

[54] Smart (1961), pp. 42–3.

[55] Pettit and Brennan (1986), p. 440.

[56] Pettit and Brennan (1986), p. 440.

would.[57] In such a case it would be best for all of us to begin using a Kantian decision procedure: trying to determine the maxims under which we act and whether or not they are universalisable.[58] This would clearly involve actual deliberation and more than just habitual responses. We would thus be in a situation where it would be optimific for us to *decide* in a way other than naïve calculation. Indeed, we need not go so far as demons. Smart, Pettit and Brennan admit that rules-of-thumb or predispositions can allow us to produce better actions overall than perpetual deliberation for such reasons as time efficiency and personal biases. Surely though, such reasons also bear upon situations that cannot be addressed merely by a non-deliberative instinctive rule. A certain form of deliberation may well do better than both naïve calculation and any rule-of-thumb: we should not rule this out *a priori*.

4.1.3 Rules and principles within decision-making

When Mill confronts the objection that consequentialism cannot be practically applied, he states:

> '[M]ankind must by this time have acquired positive beliefs as to the effects of some actions on their happiness; and the beliefs which have thus come down are the rules of morality for the multitude, and for the philosopher until he has succeeded in finding better.… The corollaries from the principle of utility, like the precepts of every practical art, admit of indefinite improvement, and, in a progressive state of the human mind, their improvement is perpetually going on. But to consider the rules of morality as improvable, is one thing; to pass over the intermediate generalisations entirely, and endeavour to test each individual action directly by the first principle, is another.'[59]

While we may well question Mill's claim that we have our rules of common-sense morality *because* they are believed to be conducive to universal happiness, there is no questioning that they indeed *are* conducive to universal happiness. Since such rules give us a way of acting which promotes the good (in accordance with the 'first principle'), Mill advocates their use and names them 'secondary principles'. In this way, the secondary principles are merely derivative of the first principle, but (contrary to Smart) are not claimed to be mere habits of mind. There is not much firm evidence for precisely how they are to be used, except that Mill likens their use to that of landmarks which guide us on a journey.[60] From this and their being named as 'principles', it would appear that they are to be appealed to in deliberation

---

[57] Railton (1984), p. 155. Note that while Railton uses this to argue forcefully for our using the optimific decision procedure (or 'mode of decision making'), he says very little about what a decision procedure is or in which role we are to judge it (compliance, acceptance etc.).

[58] Perhaps this would be best achieved by attempting to convince everyone that consequentialism is actually false (making consequentialism a *self-effacing* theory), perhaps not. It does not matter to the present discussion.

[59] Mill (1861), Ch. 2, pp. 275–6.

[60] Mill (1861), Ch. 2, p. 276.

and not restricted to mere automatic action responses.

G. E. Moore also claims that we should make decisions according to a set of rules which are roughly modelled on common-sense morality.[61] He explains why following these rules will generally lead to the best outcome and then that we can on no particular occasion be justified in believing that we are in one of the exceptional cases in which we should break the rule.[62] This extreme pessimism regarding our knowledge of exceptional cases leads to a very strong view on which the 'principles by which the individual should decide' are to consist purely in the rules of common-sense morality, and leave no room for any appeal to the goodness of outcomes. However, this does appear to be an empirical assumption on Moore's part (and a dubious one at that) so we should leave open the analytical possibility of the best principles involving some form of calculation.

R. M. Hare provides a very sophisticated account of moral decision making via principles.[63] He divides these principles into two levels: the *intuitive level* and the *critical level*:

> 'Intuitive principles are for use in practical moral thinking, especially under conditions of stress. They have to be general enough to be 'of ready application in the emergency', but are not to be confused with rules of thumb (whose breach excites no compunction). Critical principles are what would be arrived at by leisured moral thought in completely adequate knowledge of the facts, as the right answer in a specific case.'[64]

Hare thus fleshes out Mill's distinction between his first and the secondary principles. Like Mill, Hare points out that people actually do use intuitive principles, such as 'Do not commit adultery' or 'Protect your children' and that, while they are not of primary theoretical importance, the use of such principles leads to a significant amount of good. The critical principles are then used to select new intuitive principles to indoctrinate in others, to help decide between conflicts of intuitive principles and to modify our own intuitive principles when we become aware of a systematic failing in them. This explanation in terms of the two levels of moral thinking fits well with our pre-theoretic moral reasoning and also gives a fairly explicit account of the rightness (or perhaps just the *purpose*) of sets of intuitive principles:

> '[T]he object being to have those intuitive principles whose cultivation and general acceptance will lead to actions in accord with the best critical principles in most situations that are actually encountered.'[65]

This does, however, seem a little confused. If we adopt consequentialist critical principles (as Hare believes that we should) then we must take care to note that the intuitive principles whose acceptance would lead to the best possible outcome are not necessarily those whose

---

[61] Moore (1903), pp. 152–67.

[62] Moore (1903), pp. 162–4.

[63] Hare (1981), Ch. 3; Hare (1989), Ch. 4 & Ch. 13.

[64] Hare (1989), p. 221.

[65] Hare (1989), p. 221.

acceptance would accord with the critical principles in most actual situations. In particular suppose that my acts are all be wrong by the act-consequentialist criterion, but are each only slightly suboptimal. This may well lead to more good than if only one of them was wrong, but it was catastrophically wrong. It thus seems best to amend Hare's claim so that intuitive principles which lead to the most good are to be preferred.[66]

Hare stresses the importance of judging the intuitive principles based on the good that comes from their being *accepted*, rather than their being *complied with*.[67] There are several significant advantages that come from such an approach. Firstly, as stated above, this is taken to involve some form of deep commitment to the principles as opposed to their being treated as a mere calculational convenience. This helps ensure that the agent does indeed keep to the rules in question and take them very seriously. Secondly, it allows the complexity of the principles to be taken into account, for there is a limit on how complex a set we can come to accept. Moreover, rules which are quite complex or, alternatively, quite demanding of the agent are more difficult for the agent to correctly apply, so these defects will be taken into account when the rules are being assessed. Finally, there may be benefits of acceptance which do not manifest in the acts that are produced. Typical examples involve the expression of a person's principles and feelings in her countenance. It may be that someone's benevolence to her friends or her willingness to retaliate are seen clearly in her expressions and difficult to fake, leading to additional benefits from the acceptance of these principles even without any associated acts taking place.

The most striking feature that Hare introduces to his account of consequentialist decision making is the split into two levels. He can thus account for the necessity of rules which provide for fast, reliable, unbiased decision making, and also for more considered decision making when time allows or when the situation is particularly unusual. A particular benefit of this is that it allows for the agent to reshape their intuitive principles in light of information regarding their usefulness. Let us take the agent's entire decision procedure to consist of the intuitive level, the critical level and the rules that determine how they related to each other. We can then see that Hare's system allows for this overall decision procedure to demand that we modify it in light of new information. We shall call such a decision procedure *self-modifying* (although of course it is the agent who performs the modifications at the decision procedure's request).

However, there is also a problem with Hare's two level approach. While it allows a flexible choice of principles at the lower level, the upper level is fixed. It is to be a form of consequentialist criterion of rightness against which intuitive principles and exceptional cases are judged. However, it seems that it is perfectly possible that we might to do better in consequentialist terms if we adhered to Kantian principles at the higher level and hence subscribed to various intuitive principles which were justified by this Kantianism. Railton's demon would provide an example where this might indeed lead to more good. Indeed, we

---

[66] I am not sure how this could be expressed more generally so as to take into account non-consequentialist critical principles as Hare's version does.

[67] This distinction will be drawn out later in this chapter.

might do better with some set of intuitive principles and no recourse to a higher level at all. Neither of these cases are likely, but this is an empirical matter and it does not seem right to build the theory on the empirical assumption that there are no such demons or similar effects.[68] Instead, we do best to conceive of Hare's two level approach as a very plausible example of a form that the best decision procedure may take.

4.2 The nature of decision procedures

Let us now attempt to put together a precise account of the nature of decision procedures. As I have (loosely) defined it, a decision procedure is a 'way of deciding what to do'. A decision involves an agent who has a certain epistemic state and a set of available acts.[69] What we are interested in then, is the way in which the agent is to use his epistemic state to produce an act via some mental process.

As Smart has pointed out, there are conscious and unconscious mental processes with which we can produce acts. In the present essay, I wish to focus on mental processes that can be genuinely termed 'decisions' and whose outcomes can be appropriately termed 'right' and 'wrong'. I thus do not wish to consider, for example, that instinctual process whereby one's hand is quickly withdrawn when it touches something very hot. In practice it can be quite difficult to tell how an action was actually arrived at. For example, at times it is easy to tell that I am reasoning quite explicitly with a certain principle to determine how I should act, but at other times — when the case at hand is very clear-cut — it is difficult to tell if there is a very short period of similar reasoning or if I have decided by unconscious 'instinct'. However, I shall stick to a loose conception of decision in which some form of mental deliberation occurs (for example, relevant facts are weighed against each other, or principles are applied) and accept that in some situations it may be difficult to whether or not this has occurred and thus whether or not a decision has been made. Thus, unlike Smart, I shall generally consider the applications of very simple rules-of-thumb to count as decisions and (as explained above), all decisions whether short or long shall be open to assessment on consequentialist grounds.

As we have seen in the previous section, the way in which we decide to do something depends on more than just our broad aims and motives. There is no doubt that these affect the decision making that does happen, but they do not completely determine it. Instead, one can have a particular set of aims and motives and yet be capable of deciding what to do in various different ways. For example, I might aim to keep my promise and to help my friend,

---

[68] Hare's reasons for building in this inflexible assumption stem from his irrealist meta-ethics. In my formulation, there are really three places for principles: the intuitive level, the critical level and then the principles which are actually true of the world. For Hare (and many others) there is no room for such a third level at which the consequentialist criterion could reside and it is thus built into the critical level.

[69] One could instead consider a set of *options* which may include more than just acts. Depending on one's view of acts, the class of options may be more general. However, the question at hand involves deciding what to *do* and we shall consider all things that the agent can decide do as acts. By this, I only mean to make a terminological and not a metaphysical assumption.

yet I could still decide what to do by calculation, or by trusting to instincts, or by appealing to various principles. These are different methods, or procedures, by which we can move from our information to action.

Such procedures typically involve an appeal to a set of principles, the performance of a certain calculation, or the weighing of consequences, but they need not be limited to such things. Indeed, these procedures are algorithmic in nature and can involve common algorithmic structures, such as those involved when we perform long division. For example, decision procedures could include conditional branches, manifested in patterns of thinking like so: 'Does condition *x* hold? If it does, then carry on the decision making like so otherwise, carry on like this…' They could also involve repetition of the form 'Keep doing *x* until condition *y* is satisfied.'

However, there are also features of our decision procedures which are not found in the algorithm for long division. For one thing, we can perform certain intuitive steps. Even when we are trying to calculate which of two courses of action will lead to the best outcome, we do not perform any explicit multiplication of the probabilities and values, but typically just think about the major possibilities for a while and then come up with an intuitive judgement. In effect something like a calculation of expected goodness is achieved, but there is no conscious 'carrying the one' taking place. These intuitive judgements are much faster than the explicit ones would be and it is hard to imagine how a decision procedure could be practical without making significant use of them.

We also perform other mental tasks which are not found in mathematical algorithms or culinary recipes. For example, we can imagine the amount of suffering that a certain person would feel if we performed an action and compare it with some other imagined suffering. We can also try to change our emotional state in order to empathise with someone or to calm ourselves down. Indeed, our decision making processes can call for us to perform any mental activity that we are capable of.

Finally, as Hare stressed, we need not just follow certain procedures mechanically, but can attach great intuitive moral weight to them. For example, a utilitarian might have an (overrideable) rule against lying and also feel a strong sense of repugnance when they consider telling a lie. Since this additional compunction will make it easier for the utilitarian to keep to the rule when he is suffering from weakness of the will, it improves the decision procedure.

An important issue that remains is the granularity of decision procedures. For example, when I ride my bicycle, I do so via a sequence of fine-tuned movements of my body. These movements are determined by my mind's processing of information from my environment, such as what I can see and my internal sense of balance. Moreover, the bicycle, my environment and me, together form a continuous dynamical system. The movements of the bicycle affect my mind, which causes new movements in my limbs, which move the bicycle and so on. In some sense, my actions are being determined by a decision procedure. It is not a conscious one and not a discrete one, but my mind is certainly transforming information into actions.

Similarly, suppose that you are batting in a game of cricket and are ready to hit the incoming ball. You have a certain procedure for hitting the ball: if it bounces at your feet, then just block; if it comes high on your left hand side, then sweep the bat across. In this case the situation is less dynamic, but there is still a question of how fine grained the decision procedure can be. It seems clear that a decision procedure can distinguish a block from a sweep, but what about different kinds of blocking strokes? How fine-grained an action can be considered? Is there a decision procedure that tells you to hit it in such-and-such a way — a way that happens to be the very best movement that you could make with your body?

In my view, the bicycle case requires so little conscious attention that it is probably best thought of not as a decision procedure at all, on the grounds that were given at the beginning of the section. There is some kind of information processing going on though, and even if it doesn't seem to be involved in the question 'How should I decide what to do?', there are other questions such as 'How should I process my information into action?' in which it might be quite relevant.[70]

The second example involves the question of how fine-grained individual actions can be in general and, specifically, how fine-grained they can be in decision procedures. In fact, there are two related features at play here: how fine-grained the actions can be and how fine-grained the input information can be. For example, we might have a decision procedure saying that if the ball is *precisely* 20° or more to the side, then sweep and otherwise block. It is not clear what to say about such 'decision procedures'. Even if we reject the purported bicycle riding decision procedure on the grounds of being too instinctual to count as decision making, that does not imply that we must reject this very fine-grained information and action.

I am unsure as to which level of granularity is of primary interest when we ask 'How should I decide what to do?'. On the one hand, when a decision procedure is proposed, it is done in quite a course-grained manner. This is partly for the ease of communication, but there is also an intuition that a decision procedure is sufficiently coarse-grained that several different people could all be said to follow the very same decision procedure. On the fine-grained account, this is unlikely to be possible since the slight differences in the ways that we interpret it will mean that we are actually following different (fine-grained) decision procedures.

On the other hand, consider a tiny improvement to my current decision procedure (understood in a coarse-grained way). For example, it might be that I could weigh in other peoples feelings very slightly more than I currently do. In order to reflect the fact that I should decide in this new way in preference to the old one, we would seem to require a fine-grained account of decision procedures. I shall thus remain uncommitted on the issue of granularity, accepting that opinions will differ on which account is the more natural here.

[70] A focus on this question would fit well with a fine-grained theory of action in which all bodily movements, however precise, are individual actions.

4.3 The role of decision procedures

Having established some important aspects of the nature of decision procedures, we can now turn to the role that they are to play in framing the rough question. When we assess things using global consequentialism we must assess them in some specific role. We will start from the question 'What decision procedure should I follow?' and investigate several senses of the term 'follow' which each lead to a different interpretation of the rough question. As we shall see, these are all valid global consequentialist questions and may have their own subtly different answers: the right decision procedure to follow in one sense may not be the right decision procedure to follow in a second sense. This is, however, no need for all of these questions to be equally central to a moral theory. As we make these distinctions we shall thus assess how well they each fit our intuitive idea of the rough question.

4.3.1 Compliance

One obvious approach is that of compliance. It is with respect to compliance that the early forms of rule-consequentialism were phrased, and it is a natural approach. An agent can be said to comply with a set of rules if her actions are consistent with them. Thus, an agent who lies at any time has failed to comply with the rule 'Never lie', while an agent who never lies (even if this is merely by never having an opportunity to do so) does comply with this rule. Let us then consider the question:

Q_1: 'Which decision procedure should I comply with?'

There are multiple ways in which this notion of compliance can be understood when it is applied to decision procedures rather than to rules. For the purposes of this essay, I shall say that an agent complies with a decision procedure if and only if she performs the action(s) that correct application of the decision procedure would dictate. Note that there is an alternate notion that could claim the name 'compliance' here: that of following every step of the decision procedure perfectly (and thus arriving at the correct act). This second concept is a stronger one and I shall discuss it in detail under the name 'flawless execution', but for the present purposes I shall take compliance to be defined in the manner above.

A significant problem with the compliance view is that it completely ignores the algorithmic nature of decision procedures. It judges a decision procedure on what it rules out, but not upon how it rules it out. Thus, decision procedures that involve time consuming calculations or are very prone to errors in their execution are not penalised for these features. In this way, many of the problems with naïve calculation that were discussed earlier are not going to be taken into account by this formulation of the rough question. It is for this exact reason that compliance based accounts of rule-consequentialism were charged with collapse into a single rule of doing what one judges to lead to the best outcome (or the expected best outcome). The right decision procedure to comply with is thus probably going to be one that makes

heavy use of complex rules to specify the actions to be performed.[71]

It is also worth noting that the compliance approach cannot distinguish functionally identical decision procedures: those that lead us to the same acts, even if by different methods. For any decision procedure we can always construct a modified version that dictates the same actions, but which does so in a slightly different way. There will thus always be a collection of decision procedures that are right for me to comply with.

We have seen that the compliance view does not take sufficient regard of the complexity of decision procedures, but there *is* an important manner in which complexity is relevant on the compliance view. On the global consequentialist criterion of rightness that we are using, it cannot be right for me to comply with a decision procedure unless I *can* comply with it. There are two different ways in which a decision procedure can be complex. Let us say that it is *internally complex* if it directs the user through a difficult sequence of mental steps along the way to producing an action and it is *externally complex* if it produces a sequence of actions that could *only* be produced by an internally complex decision procedure. Whilst all externally complex decision procedures are internally complex, the reverse is not true because the internal complexity might just end up producing a simple sequence of actions which could in turn be produced by a simpler decision procedure. The compliance view ignores internal complexity, since it is not concerned with how the actions are chosen. There is no difficulty complying with a decision procedure that leads to simple actions but via highly complex routes, for the agent can arrive at the actions by other simpler routes and still be compliant.

However, suppose that a certain decision procedure is externally complex, and thus that its resulting actions can only be produced by internally complex decision procedures. The agent will be able to follow some such decision procedures, but for every agent there will be a point at which a decision procedure leads to a sequence of actions which is so complex that the agent cannot comply with it. There will thus be decision procedures with which we cannot comply because of their high external complexity.

The compliance view thus does take complexity into account to some degree, but it is not enough to make a practical difference. It does not penalise a decision procedure for being overly complex so long as there is some decision procedure that is free of these flaws and still leads to the same results. Moreover, it does not penalise a decision procedure at all for being slow to use or susceptible to error of calculation or bias. Since it pays so little regard to the actual structure of decision procedures, the compliance view cannot be considered to be a satisfactory interpretation of the rough question.

---

[71] Note that rule-consequentialism was also charged with a very similar type of collapse which we should take care to distinguished here. On a compliance account, rule-consequentialism could be said to collapse not into the rule of naïve calculation, but to the rule 'Do what will make the outcome best'. This cannot happen in our study of the compliance account of decision procedures because 'Do what will make the outcome best' is not (on its own) the type of thing that could guide our action and thus not a decision procedure at all.

4.3.2 Flawless execution

As discussed above, we could also assess decision procedures in terms of the good that would result when we follow all their steps perfectly. We could thus ask:

> Q_2: 'Which decision procedure should I execute flawlessly?'

Unlike compliance, this condition allows us to account for the length of time that it would take to follow the decision procedure perfectly. For example, if someone is drowning in the river, it may be that it is much better to jump in and save him, than to do nothing, but better still to call out for assistance. Suppose that decision procedure D_1 recommends jumping in to help and gives this guidance very quickly, while D_2 recommends calling out for assistance, but requires ten seconds of thought to reach that conclusion. In this case it is much better if we flawlessly execute D_1 than if we flawlessly execute D_2, even though complying with D_2 may be better than complying with D_1. We can thus see that the decision procedure singled out by this version of the rough question would be a compromise between guiding me to good acts and guiding me quickly. Indeed, one could imagine cases where being guided more slowly is better or in which the good comes from some other consequence of having worked through the procedure — perhaps our axiology values my deciding in a certain way. Importantly, the level of abstraction is no longer so high as to obscure all the problems with naïve calculation.

It is not always clear what the agent must do in order to flawlessly execute a decision procedure. While some steps (such as those involved in performing long division) are precisely delineated, many steps involve some form of judgement or interpretation. For example, a decision procedure might involve a duty to aid others when you judge that you can do so at no significant cost to yourself. Alternatively it might only allow killing someone when you judge that they are trying to kill you. Even naïve calculation involves judgements about the goodness of the different possibilities. How are we to interpret flawless execution of a decision procedure such as naïve calculation? Even if the calculating is all done perfectly to the appropriate precision, there are still many ways that one could have judged the goodness of the possibilities. Should we assess the procedure on:

> (1) the best judgements that could be made,
>
> (2) the worst judgements that could be made,
>
> (3) the actual judgements that would be made,
>
> (4) 'reasonable' judgments.

Option (1) lends too much flexibility to the decision procedure. For example, the decision procedure 'Do whatever you judge is right' seems to be intuitively quite poor, yet if the best judgements are made it will lead to quick and optimific action every time. Option (2) would greatly penalise those decision procedures which create many opportunities for intuitive judgments, and this also seems to be incorrect. Option (3) is quite plausible although it takes

us very close to the acceptance account (discussed in the next section), for it means that only the non-judgment steps of the procedure are being idealised. Option (4) is also plausible. It would need some more fleshing out, but could introduce certain constraints on the judgments, such as freedom from personal bias, freedom from cognitive malfunction and proper use of available information.

Note that this question about how to understand flawless execution of judgments affects the compliance view too. The many ways in which we could make a judgment lead to a set of potential actions. We could say that performing the best of these counts as compliance or that performing any of these counts as compliance. Alternatively we could draw upon a notion of idealisation as in (4), saying that performing any of the acts created by 'reasonable' judgments counts as compliance.

The biggest problem with the flawless execution approach is that it does not take the difficulty of executing a certain decision procedure into account. Unless a decision procedure is both very simple and undemanding — such as a handful of simple prohibitions — it is unlikely to be flawlessly followed in practice. Sooner or later, the agent is going to make a mistake, succumb to weakness of the will or just decide that the decision procedure is too demanding in a certain case. The question then is why we would care that a given procedure would be the right one to execute flawlessly if it won't actually be executed flawlessly.

Suppose that the best decision procedure that Harold can flawlessly execute is D_3, but that the good from his actual execution of D_3 will be lower than that which would have come from his actual execution of some other decision procedure, D_4. Suppose further, that we knew that executing D_4 would actually be better for Harold than executing D_3 — perhaps Harold is known to be too weak of will or prone to small mistakes in complex deliberation. In such a case, we should certainly not advise Harold to execute D_3 and we would hope that he in fact chooses to execute D_4 instead. The fact that Harold should flawlessly execute D_3 just does not seem to be of great relevance.[72]

There is another, quite similar, problem with the flawless execution account. Suppose that the best decision procedure for Adam to execute flawlessly is D_5 and that, unlike Harold, he will actually execute it flawlessly. This does not mean that this is the best decision procedure for Adam to accept, for it may be that by trying to execute D_6 (and making occasional mistakes) he could actually lead to better outcomes than he could by flawlessly executing D_5. This case seems to be quite realistic, for few people, if any, flawlessly execute a decision procedure in the actual world. Perhaps the only decision procedures that we could flawlessly execute are very simple and very undemanding. It may well be better to attempt to execute more complex procedures even though we somewhat fail in our aims. For these reasons, it seems that the flawless execution interpretation of the rough question is also unsatisfactory.

[72] Note that the decision procedure which Harold should *comply with* will also be D_3.

4.3.3 Acceptance

A third approach is to move to an analysis of the rough question in terms of acceptance:

> Q_3: 'Which decision procedure should I accept?'

Unlike the previous two accounts, this interpretation gives appropriate weight to the difficulty of following a given procedure, as well as to the time taken. If a decision procedure is too demanding or complex then it will not actually be flawlessly executed and the benefit of accepting it will be lower. Thus, Q_3 picks out D_4 as the decision procedure that Harold should accept.

There is some potential confusion concerning what it means to accept a decision procedure. On the one hand Q_3 might refer to the decision procedure that it would be best for me to currently accept. On the other, it might refer to the decision procedure that it would be best for me to come to accept. These are two different senses of the word 'accept': the first is that I hold a certain attitude to the decision procedure and the second is that I will begin a process of achieving that attitude (the process of acceptance).[73] Both thoughts can be expressed by the question 'Which decision procedure should I accept?', but in Q_3 I wish to convey the former. Thus, Q_3 is to be considered as equivalent to the more cumbersome: 'Towards which decision procedure should I hold an attitude of acceptance?'.

We can then clearly state the second interpretation on its own:

> Q_4: 'Which decision procedure should I try to inculcate in myself?'

One problem with Q_4 is that it is so future oriented. Suppose that I am changing my way of life radically in a few weeks. I am a doctor now and will be retiring. It may be that the best decision procedure for me to presently accept is the one that I actually presently accept: a decision procedure that is very suited to the demands of a doctor's high-stakes decision making. However, it may be that there is another decision procedure that would be much better for me to follow once I have retired and which is the best decision procedure for me to start inculcating in myself now, since it would take quite a while for me to change over. In this case, Q_3 will recommend my present decision procedure and Q_4 my future decision procedure. Both recommendations have their place, but the second one does not seem to fit so closely with the question 'How should I decide what to do?'.

This disconnection can be further brought out by another 'demon' example. Suppose that there is an all powerful demon who will cause massive suffering unless I inculcate a certain bizarre decision procedure in myself. Suppose further that it will take me at least a year to fully inculcate this decision procedure in myself, but regardless of whether I do, the demon will kill me in a week's time, before there is any appreciable change to my actions from the inculcation. It would certainly be right for me to inculcate this decision procedure in myself, for it would avoid the great suffering and would not change my actions very much. However,

---

[73] The same ambiguity is present in the term 'internalise' and for these purposes I treat them as synonymous.

consider the question 'How should I decide what to do?'. The fact that I should inculcate the demon's bizarre decision procedure seems to have no bearing on this. I have one week left to live and should decide what to do by the method that will lead to the most good in that time. It thus appears that Q_4 cannot be an acceptable interpretation of the rough question.

In contrast, Q_3 suffers from none of the problems that I have so far discussed. Since it does not simply assume compliance, any factors that a decision procedure has which are conducive to such compliance will be properly taken into account in the assessment. Thus decision procedures that achieve less good because they are calculatively difficult or highly demanding or do not excite moral compunction will be appropriately penalised for this. It also takes into account the other problems that naïve calculation was found to have. A decision procedure that it is good to accept will allow the agent to cease deliberating when it would interfere with her concentration or her relaxation. It will allow her to satisfice when the stakes are low and to act quickly in emergencies. Finally, the acceptance account takes into consideration the non-act benefits of accepting a decision procedure. For instance, it might make me happier or calmer to accept one decision procedure rather than another and the goodness arising from this is taken into account. Since Q_3 is sensitive to all such considerations that we have searched for in our investigation of decision procedures, it is the best of these interpretations of the rough question and we shall use this acceptance account from here on.

Chapter 5

Problems addressed

In this chapter, I shall present and address three important objections to the decision procedure account.

5.1 The inconsistency objection revisited

Now that we have seen several ways in which the rough question could be formulated, let us return to the objection that the rightness of acts and of decision procedures can conflict. Hooker presents the objection as follows:

> 'Suppose, on the whole and in the long run, the best decision procedure for you to accept is one that leads you to do act *x* now. But suppose also that in fact the act with the best consequences in this situation is not *x* but *y*. So global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure. That seems paradoxical.'[74]

For example, the best decision procedure, *D*, may involve always telling the truth, which requires the agent to tell the truth about the Duchess even though lying about the Duchess would lead to a better outcome. In such a case:

(1) the agent ought to accept *D*,

(2) accepting *D* leads to the agent doing *x*,[75]

(3) the agent ought not do *x*.

We might object that this just goes to show that the agent ought to accept some other decision procedure, *D*'. In the example above, perhaps *D*' involves dropping the constraint

[74] Hooker (2004), § 5. See also Crisp (1992), Lang (2004).

[75] An alternate formulation treats the second claim as '(2b) *D* demands that the agent does *x*'. This would be a slightly different objection, but the same reply serves for both.

on lying or just dropping it with regards to lies about the Duchess. We may well be able to make such a move in this particular case, but for plausible examples of *D*, it is much harder. Indeed, the only way to avoid there being cases where (1) – (3) are true is for accepting *D* to always lead to the agent doing the right act. However, it does not appear that there are any decision procedures with this quality.

Thus it seems that adherents to global consequentialism must simply accept this unintuitive situation. It may 'seem paradoxical', but there is no logical contradiction present. Indeed, when examined closely, even the air of paradox may vanish. The only reason that 'global consequentialism tells you to use the best possible decision procedure but also not to do the act picked out by this decision procedure' is that even the best possible decision procedure for you to accept is not perfect. We are human and have limitations. Accepting sensible decision procedures can lead to our performing actions which bring about very good outcomes, but even the best decision procedures for us to accept will lead to the performance of non-optimal actions every now and then.[76] It is these sub-optimal actions that global consequentialism condemns and, when looked at in this manner, it is difficult to see what is wrong with this condemnation.

However, this inconsistency objection appears more damaging when phrased not in terms of *acceptance*, but in terms of *compliance*. I have already argued that the rough questions is better interpreted in terms of acceptance, but even so, global consequentialism assesses all evaluands and must give a consistent account of compliance too. If it failed to do so, then we would have to abandon global consequentialism, or at the very least make some serious modifications. Therefore, consider the following three claims:

(1a) the agent ought to comply with *D*,

(2a) compliance with *D* necessarily involves the agent doing *x*,

(3a) the agent ought not do *x*.

These claims may appear more troubling because the causal connection between accepting *D* and doing *x* has been replaced by a connection of logical necessity between complying with *D* and doing *x*. While (1a) – (3a) are not themselves inconsistent, they would become inconsistent when conjoined with an intuitive deontic principle:

$O(P) \wedge \Box(P \rightarrow Q) \rightarrow O(Q)$

roughly: If one ought to *P*, and (necessarily) *P* implies *Q*, then one ought to *Q*.

Our hypothetical agent ought to comply with *D*, compliance with *D* necessarily implies performing *x* and yet it is not the case that the agent ought to perform *x*. This contradicts the deontic principle, so one of these four claims must be dropped. However, as we shall see, there is reason to believe that it is the deontic principle which should go. If we substitute

[76] Either because it guides us to these actions or because we will not actually follow its guidance.

'$S \wedge T$' for 'P' and substitute 'S' for 'Q' then we get a special case of the deontic principle:

$O(S \wedge T) \wedge \Box(S \wedge T \rightarrow S) \rightarrow O(S)$

roughly: If one ought to S and T, and (necessarily) S and T implies S, then one ought to S.

Since it is indeed a necessary truth that S and T implies S, we can derive that:

$O(S \wedge T) \rightarrow O(S)$

roughly: If one ought to S and T, then one ought to S.

This new deontic principle is thus a consequence of the original one. However, it is well known to be problematic. Suppose that S is inviting Charles to dinner and that T is settling the dispute with him. Suppose further that the best thing for you to do is to invite Charles to dinner and then settle the dispute with him, the second best is to not invite him at all and the third best is to invite Charles but argue all night. Finally, suppose that if you invite Charles to dinner you will end up arguing with him all night. In this famous type of case, you ought to invite Charles to dinner and settle the dispute (that is you ought to S and T). However, now consider the question 'Should you invite Charles to dinner?' (or 'Ought you to S?').

Opinion is divided on this case.[77] *Actualists* hold that you ought not invite Charles to dinner, for this will actually lead to the worst of the possibilities. *Possibilists* deny this, holding instead that you ought to invite him, because this is what you do in the best of your possible futures. Thus, actualists deny the new deontic principle and consequently deny the original deontic principle too. I shall not enter the actualism/possibilism debate here, but merely point out that there are good independent reasons for rejecting this deontic principle which threatened global consequentialism.

5.2 The accidental benefits objection

Recall the example of *Test X* where a positive result means that one of two underlying conditions are present, each of which will lead to certain death unless the correct drug is administered. In 90% of cases, this is drug C and in 10% of cases it is drug R. This would appear to be a classic example in which the rightness of decision procedures comes to the fore: the right decision procedure for the doctor to accept would be one that guides her to prescribe drug C whenever a patient tests positive. In the long run, decision procedures that have this feature will produce better outcomes. Since drug C is the one that naïve calculation recommends, this consideration could be seen as providing support for the use of naïve calculation in the domain of prescribing drugs.

However, consider the following additional information. Suppose that the Felicity gets thirty patients who test positive for Test X over her career and that, true to form, three of them

[77] See Jackson and Pargetter (1986) for an exposition of the problem and defence of actualism.

suffer from condition *R*. Suppose that these people come in on the 18th of July 2003, the 14th of September 2011 and the 9th of December 2014. Now consider decision procedures that involve the following: if someone tests positive for Test X, look at the calendar: if it is the 18th of July, the 14th of September or the 9th of December, give them drug *R*, otherwise give them drug *C*. Assuming that no-one else came in on those days, this decision procedure would save all thirty people.

We could quite rightly maintain that such *ad hoc* modifications cannot be made to all decision procedures, since the decision procedures may become too complex for us to remember them and follow them properly. However, in cases like Felicity's where the stakes are high and the modifications are quite easy to commit to memory, we must admit that it is better for her to accept this decision procedure rather than one which always leads to drug *C* being prescribed. Even if she had previously been following the more broad decision procedure: 'use naïve calculation to prescribe drugs to patients', an exception for the three dates would lead to it creating significantly more good.

This pattern of reasoning can be applied outside the medical domain too. Adding a few judicious 'stay home on the 15th of March' clauses to an otherwise normal decision procedure could lead to considerably better outcomes at very little cost in complexity. Consider too the decision procedure: 'buy a lottery ticket with the numbers 18, 7, 79, 9, 12, 80 and donate the winnings to charity'. There is such a decision procedure (with the appropriate numbers) that would allow each of us to do a tremendous amount of good. Indeed, it would probably do more good than many people will actually achieve over their lives.

However, all of these decision procedures are united in being astoundingly *ad hoc*. We could never know that it would be best to follow them and they are not what we typically have in mind when we ask 'How should I decide what to do?'. The problem is that decision procedures can make the agent choose an action based on information which is *accidentally* correlated with the best act. There is no particular reason for 18th of July, the 14th of September and the 9th of December to be correlated with the goodness of prescribing drug *R*, it just turned out that way.

In contrast, the connection between observing that the patient has a hot forehead and the goodness of giving him medication for a fever is much less contingent. We could say that testing a patient's forehead and giving fever medication if it is hot, is a decision procedure that *reliably* leads to good outcomes, while checking the calendar for the three specified dates and prescribing drug *R* if and only they come up does not. It leads to good in the actual world, but would be a terrible procedure in many nearby possible worlds which are consistent with the agent's epistemic state. This issue of reliability across possible worlds is dealt with extensively in epistemology and it is perhaps possible to construct an account of rightness for decision procedures which takes account of their goodness across nearby worlds. However, I shall not have space to pursue such an approach here.

What is easier, and arguably more promising, is to give an internalist account of the rightness of decision procedures. We could say, for instance, that the right decision procedure to

accept is that which leads to the greatest subjective expectation of good.[78] This would assess a decision procedure in terms of the agent's mental state only, so that unknowable facts regarding the lottery numbers could not affect the rightness of decision procedures. Whereas we had so far been concentrating on the *objective* rightness of decision procedures, this would be an account of their *subjective* rightness. Just as these two senses of rightness are applied in judging acts, so too they could be applied in judging decision procedures. Roughly speaking, the objectively right decision procedure to accept would be the one whose acceptance leads to the greatest actual good, while the subjectively right decision procedure to accept would be the one whose acceptance leads to the greatest subjective expected good.[79]

The subjective and objective senses of rightness as applied to decision procedures both have their place. When agents are considering which decision procedure to accept, they do not have all the relevant facts and thus cannot tell what the objectively right decision procedure would be. However, they do have access to their degrees of belief about these facts and so can choose to follow the subjectively right decision procedure. The subjective and objective senses of rightness thus have a similar roles regarding both choices of acts and choices of decision procedures.

Let us consider again the original objection that led to the paradox of benevolence. If an agent decides what to do via naïve calculation then she predictably does worse than if she decides in other specifiable ways. To this, the consequentialist reply was that we should not use naïve calculation, but should decide what to do in whichever way it is that leads to the best outcome. This is an appeal to the objectively right decision procedure and thus it has been this sense of rightness that I have focused on. However, when it comes to fully developing the theory of rightness of decision procedures, we must take account of both senses.

5.3 A decision procedure regress

The trouble that we have in determining the right decision procedure to follow suggests the following regress.[80] I want to act rightly and I look to consequentialism to find out what the right act is. However, consequentialism merely tells me that the right act is that one which makes the outcome best. This may be true, but I can't see which act that is. How then should I decide what to do? In other words, what is the right decision procedure? Consequentialism tells me that the right decision procedure to follow is the one that makes the outcome best. Once again, this may be true, but it doesn't tell me just which one that is. In which way should I decide which decision procedure it is best to follow? Global consequentialism tells

[78] The subjective expectation of good for an agent's acceptance of D is the sum of the good of each outcome that accepting D could lead to, weighted by the agent's degree of belief that this outcome would eventuate if D were accepted.

[79] As before, this condition could be modified to involve the degrees of belief that the agent has reason to assign, rather than those that the agent actually assigns.

[80] Note that this regress is not one of those mentioned by Bales (1971). Something like it is, however, mentioned in Lang (2004) and Smith (1991).

me that I should decide this via the best available 'meta decision procedure'. This may be true, but I do not know which meta decision procedure this is…

This regress suggests that appeals to consequentialism do not tell us anything about how we are to act or to decide what to do. Even the decision procedure account that I have expounded at length appears to be just another step along this infinite, and ultimately unenlightening, sequence. To some degree this is true, but we must be careful not to exaggerate consequentialism's impotence. Consequentialism, on its own, does not give any concrete advice at any of these levels, but this is because it is so utterly exposed to the facts. From the philosophical armchair, we commonly try to imagine ourselves in ignorance of the facts so that we might try to find *a priori* truths. In this state of ignorance, consequentialism will indeed give us no advice in a form that enables us to immediately recognise the right option. That is, all of its claims will be more like 'Do that which promotes the good' than like 'Do not lie'. For consequentialists, this is just the way that morality is and to say any more in ignorance of the facts would be impossible. Railton puts this particularly well:

> 'A further objection is that the lack of any direct link between objective consequentialism and a particular mode of decision making leaves the view too vague to provide adequate guidance in practice. On the contrary, objective consequentialism sets a definite and distinctive criterion of right action, and it becomes an empirical question (though not an easy one) which modes of decision making should be employed and when. It would be a mistake for an objective consequentialist to attempt to tighten the connection between his criterion of rightness and any particular mode of decision making: someone who recommends a particular mode of decision making regardless of consequences would not be a hard-nosed non-evasive objective consequentialist, but a self-contradictory one.'[81]

The apparent regress was created because: (a) we can perform several different sequences of acts; (b) some of these sequences are to be preferred over others; and (c) there is a cost involved in choosing which sequence to perform. These criteria are met in many other areas, and the regress problem is thus not at all unique to consequentialism.

A particularly concrete example concerns robotics. When engineers develop robots to traverse rocky areas, they program them with path-finding algorithms which convert information about their environment into a particular route and a sequence of actions that they will use to travel along that route. Since the robots are designed to be operating remotely the aim is to travel to the destination using as little energy as possible. One might think that it would always be best to use the algorithms that find the shortest or most energy efficient routes, but this is not the case. If one tries this, then the robots do find more energy efficient routes, but they use more energy to find these routes than they save in following them. Like us, the robots do best if they compromise, spending enough time and energy on decision making to produce adequate actions, but not so much that they would have been better off with a more crude approximation.

---

[81] Railton (1984), p. 156.

When such a compromise between decision making and action is required, there is a question as to how one is to determine how much effort to spend on the decision making and what form it will take. This meta–decision making has its own costs and the same problem occurs at this higher level. However, we can now clearly see that there is nothing mysterious about it. We are not tempted to throw out robotics just because it is not energy optimising for the robots to determine their own energy optimising routes. Instead, we say that it is unfortunate, but a mere fact of the matter, showing that decision making under resource limits is a difficult thing to do perfectly. The problem affects consequentialism just because consequentialism sets a criterion of success which penalises lengthy deliberation. However, such criteria are set all the time. For example when we try to get a complex dinner ready on time, we have these same trade-offs between time spent doing and time spent thinking about the best way to organise the doing.

These trade-offs also occur in other moral theories — or at least, they should do. If moral theories are blind to the facts and prescribe one ultimate way of making decisions, then they are open to disaster. Suppose that one of Railton's demons causes great misery if anyone fails to make their decisions according to naïve calculation and that this was well known. If a moral theory does not have the flexibility to prescribe that we follow naïve calculation in such a case, it would knowingly embrace disaster. Alternatively, if the theory prescribes decision procedures based on their merits given the facts at hand, then it must face the same problems as the consequentialists.

Now that it has been established that consequentialists are not alone in suffering from this problem, we should stress that it is, in any event, far from crippling. Whilst Railton understates the difficulty of actually determining which decision procedures we should use, we can do *something* about it. Even if the (objectively) right decision procedure for me to accept is thoroughly *ad hoc*, we do have considerable information on the relative merits of decision procedures. For example, we have good reason to suspect that a typical lottery-style or calendar-style decision procedure will do poorly. We should also be able to see that decision procedures in the style of common-sense morality could be significantly improved by increasing the duties of charity to the very poor.

Finally, we can see that there is considerable merit in adopting a split level, self-modifying decision procedure like that advocated by Hare, for this approach would allow us to modify the intuitive principles in light of new evidence or argument. Whilst this involves using a form of naïve calculation at the higher level, this is less alarming than it may first appear, because its application is limited by the split level structure. Its use is restricted to those times when we are not rushed and since it is only used to choose principles rather than actions, there less scope for our being tempted into personal bias. We can thus derive some considerable benefit from applying consequentialism to decision procedures, even if we may not be able to derive the optimal benefit.

Chapter 6

Conclusions

In this thesis, I have presented a detailed account of how we should decide what to do. According to consequentialism, we should not decide according to naïve calculation, but should (as it is commonly suggested) follow whichever decision procedure it is that would lead to the best outcome. Where most consequentialists have considered this brief answer to be sufficient, I have investigated it in detail, demonstrating that a satisfactory account would have to address a number of issues.

One of these key issues is an explicit acceptance of global consequentialism. In Chapter 3 we saw that act-consequentialism, on its own, does not have sufficient resources to provide an account of how we should decide what to do and that rule-consequentialism's account has several features which make it distinctly unattractive to someone motivated by the intuition that what matters is just the goodness of outcomes. Instead, global consequentialism offers the resources needed to provide an account of rightness for any evaluand and thus to provide such an account for decision procedures. It a very natural account which is likely to be agreeable to many act-consequentialists.

We then saw that there is considerable ambiguity when it comes to understanding 'the right decision procedure to follow'. While consequentialist authors have been united in their agreement that following naïve calculation would not be optimal, they have not put forward a unified account of our alternative ways of choosing our acts. Indeed their separate accounts — motives, rules-of-thumb, predispositions and principles — were each put forward with little discussion about why the process of choosing acts should be analysed in this particular way. In Chapter 4 we examined these disparate analyses and formed a broad understanding of decision procedures as conscious algorithmic manipulations of information and intuition to form action. In the same chapter, we also saw that there are many ways in which we can understand what it means to 'follow' a decision procedure. Of these, the acceptance account was found to give the most natural interpretation of the rough question.

Finally, we explored several objections to this consequentialist analysis of how we should decide what to do. The inconsistency objection was found to reveal a counterintuitive property of rightness as applied to decision procedures and to acts, but to lead to no

contradictions. The accidental acts objection then revealed that the objectively right decision procedures would often be *ad hoc* and unknowable. While this result was initially counterintuitive, it became less so when we considered an account of the subjective rightness of decision procedures. Just as the understanding of subjective and objective rightness for acts explains how the (objectively) right act can be unknowable, so too for the rightness of decision procedures.

A regress was then presented which pointed to a gap in the account of how someone with calculational limitations could come to determine the right decision procedure. However, this was then seen to be a more general problem with all accounts of optimal decision making in moral and prudential theories, rather than a distinctive problem for consequentialism. Moreover, while we may not be able to see the optimal way for a limited agent to determine the best decision procedure, we can see how they would find *better* decision procedures and so allow themselves to increase the amount of good that they produce.

This consequentialist account of how we should decide what to do is of considerable importance. Not only does it provide a reply to the claims that consequentialism is self-defeating, but it also provides an important way to view consequentialism itself. Traditionally, consequentialism is introduced via its criterion of rightness for acts. For many students of moral philosophy, consequentialism is merely seen as the theory that the right act is the one which leads to the best outcome. The problem of how we are to decide is rarely considered and naïve calculation is typically assumed without question. At best a text on consequentialism may have a brief paragraph explaining the consequentialist reply that we should follow the best decision procedure.

One could, instead, introduce (global) consequentialism via its response to the rough question. Thus consequentialism would be explained, first and foremost, as the theory that the right way to decide what to do is whichever it is that leads to the best consequences. If introduced in this manner, the intuitive connection between consequentialism and naïve calculation is cut and the potential for consequentialism to suggest that we follow something similar to common-sense morality is brought to the fore. The potential for the best decision procedure to come apart from the consequentialist criterion of rightness is not just an abstract possibility, but is central to the theory. Focusing on this from the beginning may thus provide a much more fruitful way of presenting and understanding consequentialism.

References

Adams, R. M. 1976. 'Motive utilitarianism', *Journal of Philosophy* 73:467–81.

Austin, J. 1832. *The province of jurisprudence determined*, H. L. A. Hart (ed.), (London: Weidenfield), 1954.

Bales, R. E. 1971. 'Act-utilitarianism: account of right-making characteristics or decision making procedure?', *American Philosophical Quarterly*, 8:257–265.

Brandt, R. 1959. *Ethical Theory*, (Englewood Cliffs: Prentice-Hall).

Carlson, E. 1995. *Consequentialism reconsidered*, (Dordrecht: Kluwer).

Crisp, R. 1992. 'Utilitarianism and the life of virtue', *The Philosophical Quarterly* 42:139–60.

Feldman, F. 1975. 'World utilitarianism', in Keith Lehrer (ed.) *Analysis and Metaphysics*, (Dordrecht, Reidel), 255–71.

Feldman, F. 'Actual utility, the objection from impracticality, and the move to expected utility', Unpublished.

Gibbard, A. 1990. *Wise choices, apt feelings*, (Oxford: Oxford University Press).

Hare, R. M. 1981. *Moral thinking: its levels, method and point*, (Oxford: Clarendon Press).

Hooker, B. 2000. *Ideal code, real world: a rule-consequentialist theory of morality*, (Oxford: Oxford University Press).

Hooker, B. 'Rule-consequentialism', in Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)*,
URL = <http://plato.stanford.edu/archives/spr2004/entries/consequentialism-rule/>.

Jackson, F. and Pargetter, R. 1986. 'Oughts, options, and actualism', *The Philosophical Review*, 95:223–255.

Jackson, F. 1991. 'Decision-theoretic consequentialism and the nearest and dearest objection', *Ethics* 101:461–82.

Kagan, S. 2000. 'Evaluative focal points', in B. Hooker, E. Mason and D. E. Miller (eds.) *Morality, Rules and Consequences: A Critical Reader*, (Edinburgh: Edinburgh University Press), pp. 134–155.

Lang, G. 2004. 'A dilemma for objective act-utilitarianism', *Politics, Philosophy and Economics*, 3:221–39.

Lenman, J. 2000. 'Consequentialism and cluelessness', *Philosophy and Public Affairs*, 29:342–70.

Mill, J. S. 1861. 'Utilitarianism', in M. Warnock (ed.) *Utilitarianism*, (Fontana: Glasgow), 1962.

Moore, G. E. 1903. *Principia Ethica*, (Cambridge, Cambridge University Press).

Parfit, D. 1984. *Reasons and Persons*, (Oxford: Clarendon Press).

Pettit, P. and Brennan, G. 1986. 'Restrictive consequentialism', *Australasian Journal of Philosophy* 64:438–55.

Pettit, P. and Smith, M. 2000. 'Global consequentialism' in B. Hooker, E. Mason and D. E. Miller (eds.), *Morality, Rules and Consequences: A Critical Reader*, (Edinburgh: Edinburgh University Press), pp. 121–133.

Railton, P. 1984. 'Alienation, consequentialism, and the demands of morality', *Philosophy and Public Affairs* 13:134–71. Reprinted in

Sidgwick, H. 1907. *The Methods of Ethics*, 7th edition, (London: Macmillan).

Smart, J. J. C. 1956. 'Extreme and restricted utilitarianism', *Philosophical Quarterly* 6:344–54.

Smart, J. J. C. 1961. 'Outline of a system of utilitarian ethics', in Smart and B. A. O. Williams, *Utilitarianism: for and against*, (Cambridge: Cambridge University Press), 1973, pp. 3–74.

Smith, H. 1988. 'Making moral decisions', *Noûs* 22:89–108.

Smith H. 1991. 'Deciding how to decide: is there a regress problem?' in Michael Bacharach and Susan Hurley (eds.), *Foundations of Decision Theory*, (Oxford: Basil Blackwell), pp. 194–219.

Streumer, B. 2003. 'Can consequentialism cover everything', *Utilitas* 15:237–47.

Zimmerman, M. 1996. The concept of moral obligation, (Cambridge: Cambridge University Press).