

Chapter 8

The Value of Making and Keeping Promises

Michael Smith

Abstract

Consequentialists hold, quite in general, that what we are obliged to do are those things, among the alternatives available to us, that uniquely maximize value and minimize disvalue. But surely this cannot be right, some people say, because we may have an obligation to (say) keep a promise we made to someone whether or not doing so happens to maximize value and minimize disvalue. Fully understanding how we are able to put ourselves under an obligation to another person by making him a promise is, they think, thus the first step in a move away from a consequentialist analysis of obligation in general. But is this right? The issue turns on whether gaining a full understanding of how we are able to put ourselves under an obligation to another person by promising is a matter of coming to appreciate the special values produced by making promises. I will argue not just that it is, but that the opponents of this view are themselves committed to it by their own lights.

What is the value of making and keeping promises? In what follows I will describe and evaluate several different answers to this question. Equipped with one sort of answer in particular, my hope is that we might inch our way toward a better understanding of both promissory obligation and the long-standing dispute in moral philosophy between consequentialists and nonconsequentialists.

The essay divides into five sections. I begin by explaining the connection between understanding the nature of promissory obligation, on the one hand and the debate in moral philosophy between consequentialists and nonconsequentialists on the other (section I). Since consequentialist theories of obligation will occupy center stage in what follows, I next go on to

offer some preliminary clarification of what it is for a theory to be consequentialist (section II). Equipped with this account of what it is for a theory to be consequentialist, I then describe and evaluate two accounts of the value of making and keeping promises and the consequentialist accounts of promissory obligation to which they give rise (section III). This leads to a further clarification of what it is for a theory to be consequentialist (section IV) and a description and evaluation of one last consequentialist account of promissory obligation (section V).

To anticipate, the moral of the essay is that consequentialists have the upper hand in the debate over the nature of promissory obligation. The argument given for this conclusion is, however, sufficiently general and abstract that it will require us to take no stand at all on what the value associated with promissory obligation is. The conclusion, in other words, is that promissory obligation should be understood in consequentialist terms no matter what value it is associated with.

I. CONSEQUENTIALISM, NORMATIVE REASONS, AND PROMISSORY OBLIGATIONS

As I understand it, the dispute in moral philosophy between consequentialists and nonconsequentialists is a dispute about the nature of permissibility and impermissibility. According to consequentialists, facts about permissibility and impermissibility reduce to facts about what people can do—that is, facts about the options that people have in the circumstances of action in which they find themselves—together with facts about the values of the outcomes of their doing the various things that they can do. The opponents of consequentialism, by contrast, deny that a reduction of this kind is possible.

Myself I think of this as a debate within serious metaphysics (Smith 2005)—the issue, as I understand it, is whether facts of one kind consist in facts of another—but that may just reflect my own realist leanings. Since realists have their own metaphysically less serious way of understanding talk of “facts” about permissibility and impermissibility, and hence their own way of understanding talk of the “reduction” of one set of facts to another, they too can presumably take sides in the dispute between consequentialists and nonconsequentialists, as I have characterized it. Despite the realist-sounding language, I therefore take it that this characterization of consequentialism in terms of its reductionist ambition is neutral on the issue of realism versus idealism.

So understood, consequentialism is a broad church that allows for much disagreement, as consequentialists can and do differ among themselves about the precise nature of the reduction. But, at least when understood in

this broad church way, it has always seemed to me that some sort of consequentialism just has to be correct. It has to be correct because consequentialism, so understood, dovetails so readily with an extremely plausible and widely held account of the nature of normative reasons, an account that Joseph Raz dubs the "classical" account (Raz 1999, p. 23) and a plausible conceptual truth about moral obligation. Let me explain the connection between these two ideas.

According to Raz,

one approach to the explanation of agency, with origins in the writings of Plato and Aristotle, takes acting for a reason to be the distinctive and central case of human agency. ... Reason is then explained in part by invoking value: valuable aspects of the world constitute reasons. (Raz 1999, p. 22)

According to the classical account, what it is for an agent to have a normative reason to act in a certain way is thus for there to be some value realized by his acting in that way.

John Mackie further explains what it is for something to have value, according to the classical account of normative reasons, as follows.

Values themselves have been seen as at once prescriptive and objective. ... Sidgwick [for example] argues that ... what ought to be "must in another sense have objective existence: it must be an object of knowledge and as such the same for all minds"; but he says that the affirmations of this science "are also precepts," and he speaks of pleasure itself as "an end absolutely prescribed by reason." (Mackie 1977, pp. 23–24)

Sidgwick's idea is thus that something has value just in case—and here I paraphrase his idea of something's being "an end absolutely prescribed by reason"—there is a requirement of reason that we desire or approve or aim at it.

Though he doesn't trace the idea back to its classical origins, Thomas M. Scanlon holds a very similar conception of values.

To value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it. Exactly what these reasons are, and what actions and attitudes they support, will be different in different cases. They generally include, as a common core, reasons for admiring the thing and for respecting it. (Scanlon 1998: 95)

Thus, according to Scanlon, to value something is, *inter alia*, to take oneself to have reasons for desiring or approving or aiming at it, from which it presumably follows that for that thing to be valuable is, *inter alia*, for there to be reasons for desiring or approving or aiming at it. On the assumption that reason requires us to be sensitive to such reasons as there for holding certain positive attitudes, Scanlon's idea and Sidgwick's thus amount to much the same thing.

If what Sidgwick and Scanlon say about the nature of value is along the right lines, then it should come as no surprise that, according to the classical account, what an agent has a normative reason to do is fixed by the values associated with his actions. For it follows from the fact that an agent's actions realize something of value that, to put it in the way I paraphrased Sidgwick, there is a requirement of reason that he desires or approves or aims at that thing. What better candidate could there be for what agents have normative reason to do than what they are required by reason to desire or approve or aim at? The dispositional theory of value and normative reasons that I myself have developed elsewhere is, of course, simply a further elaboration and refinement of this classical account (Smith 1994, 2005).

I said earlier that consequentialism dovetails well with a plausible account of normative reasons. It should now be clear why I said this. For when we add to the classical account of what it is for someone to have a normative reason the observation that is a contradiction in terms to suppose that an agent could have an obligation to act in a certain way without his having a normative reason to act in that way, it seems hard to resist the conclusion that facts about agents' obligations themselves consist in the very facts about the values of the outcomes of their actions that their normative reasons consist in. Facts about agents' obligations thus reduce to facts about their normative reasons, which, in their turn, reduce to facts about the values of the outcomes of the actions they can perform. And with the reduction of facts about obligation to facts about value in hand, the reduction of facts about permissibility and impermissibility to facts about value simply follows suit.

Notwithstanding the evident plausibility of consequentialism, when argued for along such lines, it is striking that so many philosophers who accept the classical account of normative reasons—Raz himself is an example (Raz 1986)—reject consequentialism. Moreover, several philosophers—Scanlon is a notable example (Scanlon 1998)—argue that promissory obligation in particular resists capture in consequentialist terms. Here, accordingly, we find the rationale for asking what value there is in making and keeping promises. For, to focus for the moment on the value of keeping promises, if an agent is obliged to keep the promises he makes, then, on pain of contradiction, that commits us to the conclusion that there is a normative reason for him to do so, and the fact that there is a normative reason for an agent to keep promises he makes commits us to the conclusion that there is some value realized by his keeping the promises he makes. But if this is right, then what is that value? And how does the resulting consequentialist account of promissory obligation, grounded in that value, compare to the accounts on offer that supposedly resist being fit into a consequentialist mould? A similar set of questions arises about the permissibility of making promises.

Before answering these questions it is important that we get clearer about what it is for a theory to be consequentialist. Earlier I said that consequentialists disagree among themselves about the exact nature of the reduction of facts about permissibility and impermissibility to facts about what agents can do and facts about the values of the outcomes of the things that they can do. Though some of these disagreements can safely be put to one side, others will need to be addressed. They will need to be addressed because, as we will see, some theorists think that these disagreements bear on the plausibility of consequentialism itself. It is important to see that this is a mistake, as only so will we get clear about the sorts of values that might be realized by making and keeping promises.

II. CONSEQUENTIALISM CLARIFIED

As I said, the major source of disagreement among consequentialists concerns the precise nature of the reduction of permissibility to value. Let me begin by describing some different aspects of this reduction and some of the different views that consequentialists might take.

(i) Maximizing Views versus Satisficing Views

Some consequentialists think that an act is permitted only if it maximizes value. Others disagree. They think that an act is permitted so long as it produces enough in the way of value; that maximization isn't required. This is the well-known debate within the consequentialist camp between *maximizers* and *satisficers* (Slote and Pettit 1984). For ease of exposition, I will assume that maximization is required in what follows. An act is permitted, according to the maximizers, just in case it is one of the acts among an agent's options that maximizes value, and it is obligatory just in case it is uniquely permissible. But nothing I say will turn on this assumption. My arguments could all quite easily be recast if it turns out that satisficing is all that is required.

(ii) Monism versus Pluralism

Another disagreement among consequentialists concerns the number of values that are in play when we rank outcomes. Some consequentialists think that there is just one value—most famously, pleasure—while others think that there are many different values, all of which have to be weighed against each another in determining an overall ranking of outcomes. This is the well-known debate within the consequentialist camp between *monists* (Mill 1861) and *pluralists* (Moore 1903). For ease of exposition I will only

ever talk about one value at a time. Though this may make it sound like I embrace some form of monism, what I have to say is in fact noncommittal on the issue of monism versus pluralism. Everything I say is meant to be consistent with the truth of pluralism.

(iii) Actual Outcome Views versus Expected Outcome Views

Another disagreement that arises within the consequentialist camp that can safely be put to one side is that between those who hold the *actual outcome* view (Railton 1984) and those who hold the *expected outcome* view (Jackson 1991). Consequentialists who hold the actual outcome view think that facts about the permissibility of an agent's actions are a function of the value of the actual outcomes of that agent's options, whereas those who hold the expected outcome view think that they are a function of the expected value of the outcomes of the agent's options. For ease of exposition, I will simply assume that facts about the permissibility of an agent's actions are a function of the value of the actual outcomes of his options in what follows. But, again, nothing I say will turn on this assumption. The arguments I give could all be recast in terms of the expected outcome view, if it turns out that that is the correct view to hold.

(iv) Commensurability Views versus Incommensurability Views

Another disagreement that can arise within the consequentialist camp is a disagreement among pluralists. Pluralists hold that outcomes differ from each other in various respects. They differ not just in how much of any particular value is instantiated, but also in which values are instantiated. Pluralists can therefore disagree among themselves about whether, notwithstanding these differences between outcomes, it is possible to provide a single ranking of outcomes from best to worst. Those pluralist consequentialists who think that outcomes can be ranked from best to worst hold that the different values that are instantiated in the outcomes are *commensurable* with each other. Those who think that they cannot hold that at least some values are *incommensurable*: that when different values are at stake there is sometimes no saying which outcome is better and which is worse.

This disagreement is a little more tricky, as some who oppose consequentialism do so precisely on the grounds that, though all normative reasons are grounded in facts about values, some values are incommensurable (Raz 1986). But it is difficult to see why the incommensurability of certain values is supposed to be a barrier to the truth of some sort of consequentialism. At most, what this shows is that facts about the permissibility and impermissibility of certain of an agent's actions are themselves relativized to options, all

of which instantiate values that are commensurable. Certain of an agent's actions will therefore be permissible relative to one range of options, and others will be permissible relative to a different range of options, and there will be no fact of the matter about whether his actions are permissible or impermissible simpliciter: that is, relative to all of his options. Since this is still a reductionist doctrine—facts about the permissibility and impermissibility of the agent's actions still reduce to facts about the values of his options—it still qualifies as a form of consequentialism, at least as consequentialism was defined at the outset. The incommensurability of certain values thus seems to me to be no barrier to the truth of consequentialism.

(v) Global Consequentialism versus Local Consequentialism

One claim to which I am committed by my characterization of consequentialism, and which does affect the cogency of the arguments that follow, is that consequentialists cannot avail themselves of a rule-consequentialist account of promissory obligation. This is because, as I have characterized it, consequentialism is the view that an act is obligatory only if it is an act, among the agent's options, that maximizes value. This is an *act-consequentialist* account of what it is for an action to be obligatory. I must therefore deny the *rule-consequentialist's* suggestion that an act is obligatory only if it is permitted by the rules the acceptance of which is value maximizing, never mind about whether the act itself is value-maximizing (Hooker 2000).

Of course, rule-consequentialism of this kind still falls within an even broader consequentialist church than the one that I have identified, a church that shares a reductionist ambition much like the one I have spelled out. For, according to this kind of rule-consequentialism, facts about an agent's obligations still reduce, *inter alia*, to facts about the values of outcomes. It is just that the values in question are not the values of the outcomes of the various *acts* that the agent had available to him, but are rather the values of the outcomes of accepting the various alternative *rules* that might be accepted. But I think that this even broader consequentialist church is one to which we will be attracted only if we forget the argument for consequentialism that was outlined earlier. Rule-consequentialism of the kind just described implausibly breaks the connection between an agent's having an obligation to act in a certain way and his having a normative reason to act in that way.

In order to see that this is so, remember that, according to the standard view of what it is for an agent to have a normative reason to act in a certain way, his having such a reason amounts to there being something valuable realized by his acting in that way. To paraphrase Sidgwick, the agent must be required by reason to desire, or approve, or aim at whatever it is that is realized by his action. Bearing this in mind, the problem with the kind of

rule-consequentialism just described should be clear. For there is no reason at all to suppose that there is any value realized by an agent's action when he acts in accordance with a rule the acceptance of which is itself value-maximizing. True enough, his acceptance of a rule requiring him to act in that way must be value-maximizing. He must therefore be required by reason to desire, or approve, or aim at the outcome of his acceptance of the rule. But his acting in accordance with that rule need not be value-maximizing. It may even be value-minimizing. (In this connection, it is useful to remember what J. J. C. Smart has to say about rule-worship [Smart and Williams 1973, p. 10].) The upshot is thus that, at least according to the kind of rule-consequentialism we have been considering up until now, an agent may have a normative reason *not to do* what he has an obligation *to do*. To my mind, this is a decisive objection to this kind of rule-consequentialism, as it shows that it entails a contradiction.

Assuming that I am right about this, it might be thought that a certain question becomes rather urgent. For aren't those who advance the sort of rule-consequentialism we have been considering right that consequentialism provides us with the resources to answer not just the question "Which are the acts among an agent's options that are value-maximizing?" but also the question "Which are the rules among those that we might accept that are value-maximizing?" And, if they are right about this, then aren't they also right that we need to decide which of these evaluands—acts or rules—is the appropriate evaluand in terms of which to account for facts about obligation?

The answer is that they are of course right that consequentialism provides us with the resources to answer questions about the value of both acts and rules. Indeed, consequentialism provides us with the resources to answer questions about the value of anything and everything that can be evaluated. But the correct conclusion to draw from this is not that one or another of these evaluands must therefore be privileged when it comes to understanding the nature of obligation. The correct conclusion to draw is rather that there are many different claims about obligation that we can make, one corresponding to each of the evaluands. The correct form of consequentialism is thus not some form of *local* consequentialism, according to which we must choose a privileged evaluand and then assess all others in terms of it, but rather *global* consequentialism (Pettit and Smith 2000).

According to global consequentialism, *which acts we ought to perform* is fixed by a comparison of the values of the outcomes of those acts as opposed to those we might have performed instead (this is *act-consequentialism*); *which rules we ought to accept* is fixed by a comparison of the values of accepting those rules as opposed to those we might have accepted instead (this is *rule-consequentialism*, properly conceived, as it is simply a view about which

rules we ought to accept, not a view about which acts we ought to perform); *which motives we ought to have* is fixed by a comparison of the values of having those motives as opposed to those we might have had instead (this is *motive-consequentialism*); and so we could go on. These various forms of consequentialism—act, rule, motive, and so on—are thus not alternatives to each other, but are instead all elements of a global consequentialist outlook. This will be important in what follows.

(vi) Welfarism versus Nonwelfarism

Another debate within the consequentialist camp with which I will need to engage in what follows is the debate between *welfarists* and *nonwelfarists* (Sumner 1996). Welfarists hold that the only intrinsic value is well-being; that other putative intrinsic values, in so far as they are values at all, are either a part of well-being or a condition of well-being. Nonwelfarists deny that well-being is the only intrinsic value.

Perhaps the best known version of welfarism is hedonism, the view that the only intrinsic value is pleasure and that the only intrinsic disvalue is pain, and I will begin by assuming that welfarism of this kind is correct. But there are other versions of welfarism as well. Indeed, anyone who takes the plausible view that whether or not a life goes well or badly depends on more than just the amount of pleasure or pain enjoyed by the person whose life it is—those who hold that, for example, how much one achieves can affect how well one's life goes, independently of how much pleasure one gets from those achievements (Griffin 1988)—holds an importantly different version of welfarism. And nonwelfarism too comes in many different varieties. We will consider two such varieties in due course.

(vii) Neutral Values versus Relative Values

With welfarism clearly within our sights, we are in a position to identify a final debate within consequentialism with which I will need to engage, a debate that will loom large in what follows. This is the debate about whether values, be they welfarist or nonwelfarist, are by their nature *neutral*, or whether some values are *relative*.

As I understand it, to say that values are neutral is to make a claim about how the intrinsic value-making features associated with that value are properly to be characterized. Values are neutral when, in order to properly characterize the associated intrinsic value-making features, no mention of a specific agent or evaluator need be made. Values are relative, by contrast, when the proper characterization of the associated intrinsic value-making features requires the mention of an agent or evaluator. The intrinsic

value-making features associated with the value of welfare provides a neat illustration of this distinction, an illustration that allows us to bring out the significance of this debate within the consequentialist camp.

Consider *hedonistic ethical egoism*. As I understand it, hedonistic ethical egoism is a consequentialist theory that holds that the only intrinsic value is welfare and that also holds that the intrinsic value-making feature associated with welfare is relative. This is because, according to hedonistic ethical egoism, the intrinsic value-making feature associated with the actions that I ought to perform is *my* pleasure, that associated with the actions you ought to perform is *your* pleasure, that associated with the actions someone else ought to perform is *his* pleasure, and so on: I ought to maximize *my* pleasure, you ought to maximize *your* pleasure, others ought to maximize *their* pleasure, and so on. We cannot properly characterize the value-making features in question if we do not mention agents or evaluators. *Hedonistic utilitarianism* differs from hedonistic ethical egoism precisely in that it holds that the intrinsic value-making feature associated with welfare is neutral: the intrinsic value-making feature associated with everyone's obligations is pleasure, without regard to whose pleasure it is. I ought to maximize pleasure without regard to whose, you ought to maximize pleasure without regard to whose, others ought to maximize pleasure without regard to whose, and so on.

Some philosophers think that hedonistic ethical egoism, so understood, is an incoherent position (Moore 1903). They point out that, on the assumption that hedonistic ethical egoism really is supposed to be a version of consequentialism, the effect of my actions on my pleasure can play a role in fixing what I ought to do only if my pleasure has value, and that the effect of your actions on your pleasure can play a role in fixing what you ought to do only if your pleasure has value. But they then go on to argue that, if just this much is agreed, then it is obscure why the effect of my actions on your pleasure doesn't likewise play a role in fixing what I ought to do as well as playing a role in fixing what you ought to do. The effect of my actions on your pleasure is, after all, an effect that my actions has on something that is admitted to be of value. To suppose that what I ought to do is a function of the effect of my actions on my pleasure alone is thus to suppose that what I ought to do is a function of the effect of my actions on only *some* of what's of value, not *all* of what's of value. And the same goes for the idea that what you ought to do is a function of the effect of your actions on your pleasure alone.

According to this line of argument, consequentialism is therefore committed to the view that if pleasure has value at all, then it has neutral value, not relative value: consequentialists who were inclined to accept hedonistic ethical egoism at first blush should, on reflection, reject that view in favor of hedonistic utilitarianism. And what goes for pleasure, conceived of as a relative value, goes for all of the other values that we might have thought

were relative values too, for a similar line of argument will show that it is incoherent to combine consequentialism with the view that there are any relative values at all. The upshot, according to this line of argument, is that consequentialists must suppose that all values are neutral. If this is right then that has a very important impact on our main question. For if consequentialists must suppose that all values are neutral then they will have to say that the value-making features associated with promissory obligations are neutral too.

III. CONSEQUENTIALISM, NEUTRAL VALUES, AND PROMISSORY OBLIGATION

What would a global consequentialist account of promissory obligation grounded in neutral value look like? We get one possible answer to this question if we focus on the account of promissory obligation proffered by a welfarist theory like global hedonistic utilitarianism.

Note, to begin, that if global hedonistic utilitarianism is correct, then it is a contingent empirical matter whether an agent's keeping his promises is obligatory at all. This is because it will be a contingent empirical matter whether an agent's keeping any particular promise he makes will maximize pleasure (here the focus is on hedonistic *act-utilitarianism*). Having said that, however, we must immediately add that, even if it isn't obligatory for an agent to keep all of the promises that he makes, it may still be the case that he ought to accept the rule that each person should keep all of the promises that he makes (here the focus is on hedonistic *rule-utilitarianism*), and that it may also be the case that he ought to be motivated to keep all of his own promises (here the focus is on hedonistic *motive utilitarianism*). This is because the amount of pleasure that is caused by an agent's acceptance of the rule that he keeps his own promises, and that caused by his being motivated to keep all of his promises, may be quite different from the amount of pleasure that would be caused by any of the particular acts that accord with those rules or that he would be thus motivated to perform. To repeat, however, all of this depends on empirical contingencies.

Do matters look very different if we abandon welfarism and adopt instead some nonwelfarist account of neutral values? Suppose, for example, we adopt the most favorable nonwelfarist account of neutral value from the point of view of accounting for promissory obligation: suppose we hold that a kept promise is itself an intrinsic neutral value-making feature. Would matters look very different? The answer is that this would make some difference, but not all that much. For it would still be a contingent empirical matter whether an agent's keeping any particular promise he makes maximizes the

number of kept promises. Perhaps there would be more kept promises if an agent were to ignore the promises that he makes and saw to it that others kept the promises that they make instead. If so, then he ought to ignore his own promises and see to it that others keep theirs. This too will depend on empirical contingencies. Moving to a nonwelfarist account of neutral values, even one that is most favorable to promissory obligation, thus leaves us with a consequentialist account of promissory obligation that is, in crucial respects, a lot like the account we get from the version of consequentialism that is committed to a welfarist account of neutral values.

Of course, this doesn't show that there is anything wrong with these accounts of promissory obligation. For all that we have said, one or another of these accounts of promissory obligation might be the correct one to adopt. Such will certainly be the case if the neutral values in terms of which these accounts of promissory obligation are given are the only values that there are. For in that case these neutrally characterizable outcomes are the only outcomes that reason is on the side of our desiring or approving of: we quite literally only have reason to care about our keeping our own promises to the extent that our keeping our own promises brings about such neutrally characterizable outcomes. Having said that, however, it must be admitted that neither of the accounts of promissory obligation described are accounts of the kind that we might have expected from a consequentialist *prior* to having any idea about what values there are. Moreover there is, I think, an important reason why we haven't been given an account of the kind that we might have expected. For what's remarkable about both of these accounts of promissory obligation, grounded as they in neutral values, is that an agent's obligations turn out not to be especially targeted on the promises that *he* makes. This is remarkable because promissory obligation, at least as ordinarily understood, *is* so targeted. As ordinarily understood, promissory obligation is such that I ought to keep *my* promises, you ought to keep *your* promises, and others ought to keep *their* promises; to suppose otherwise is to suppose that there is nothing distinctive about promissory obligation at all.

Ex ante, then, we might have expected a consequentialist account of promissory obligation to look for something distinctive about promissory obligation, and the obvious way for it to do this would be to model promissory obligation on the sorts of obligations postulated by hedonistic ethical egoism, rather than on those postulated by hedonistic utilitarianism, or any of its nonwelfarist counterparts. For the obligations postulated by hedonistic ethical egoism are structurally exactly like promissory obligations, as ordinarily understood. They too are targeted on features of the agents in question: I ought to maximize *my* pleasure, you ought to maximize *your* pleasure, and others ought to maximize *their* pleasure. What we might have expected from the consequentialist, then, prior to having any idea what

values there are, is an account of what's distinctive about promissory obligation in terms of the particular *relative* value that is promoted by an agent's keeping the promises that he makes, just as hedonistic ethical egoism attempts to account for what's distinctive about our obligations vis-à-vis our own pleasure in terms of the distinctive relative value of our own pleasure. But, of course, this is exactly what the earlier arguments for the incoherence of relative values tell us cannot be done. There are no distinctive relative values promoted by an agent's keeping the promises he makes because there are no relative values full stop.

To sum up, if the argument for the incoherence of combining consequentialism with the view that some values are relative given earlier is correct, consequentialism looks like it will be unable to give an account of promissory obligation as being in any way distinctive. Though it may seem to be a distinctive feature of an agent's promissory obligations that they are targeted on the promises that he himself makes, this turns out to be a mistake, as according to accounts of promissory obligation grounded in neutral values, such obligations are never so targeted. To repeat, this isn't an objection to such accounts of promissory obligation. For if there are only neutral values then reason is only ever on the side of our desiring or approving of outcomes that are neutrally characterizable. Reason isn't on the side of our desiring or approving of anything that is targeted on ourselves. But it does suggest that a rather large burden is being carried by the earlier argument for the incoherence of combining consequentialism with the view that some values are relative. The question we must ask is therefore whether the argument is really convincing. If it is not really convincing—if there is some latent confusion or equivocation—then the way is clear to look again at whether there is some distinctive relative value promoted by an agent's keeping the promises that he makes.

IV. RELATIVE VALUES RECONSIDERED

Here is the crucial part of the argument given earlier:

Some philosophers... point out that, on the assumption that hedonistic ethical egoism really is supposed to be a version of consequentialism, the effect of my actions on my pleasure can play a role in fixing what I ought to do only if my pleasure has value, and that the effect of your actions on your pleasure can play a role in fixing what you ought to do only if your pleasure has value. But they then go on to argue that, if just this much is agreed, then it is obscure why the effect of my actions on your pleasure doesn't likewise play a role in fixing what I ought to do as well as playing a role in fixing what you ought to do. The effect of my actions on your pleasure is, after all, an effect that my actions have on something that is admitted to be of value.

Despite its apparent plausibility, it seems to me that this step in the argument is crucially flawed. It is flawed because it proceeds without making explicit what it is for something to be of value. Yet when we do make that explicit, it becomes clear that the argument turns on a crucial equivocation (see also Smith 2003).

Let's once again follow Sidgwick's lead and say that something is of value if and only if and because there is a requirement of reason that we desire it, or approve of it, or aim at it. We might initially try formulating this claim as follows:

$\forall x$ (p is valuable iff and because x is required by reason to desire that p, or approve of p, or aim at p).

But a moment's reflection makes it clear this is at best a very sloppy formulation. What we have is here is a purported reduction of facts about value to facts about the desires, attitudes of approval, and aims that are required by reason. Desirers are therefore explicitly mentioned on the right-hand side, but where exactly are they mentioned on the left-hand side? The answer is that they must be mentioned at least implicitly in our talk of something's being valuable. To properly formulate the claim we should therefore make the mention of desirers on the left hand side explicit as follows:

$\forall x$ (p is valuable_x iff and because x is required by reason to desire that p, or approve of p, or aim at p).

This reformulation is required because it makes it explicit that the concept of something's being valuable is itself a relational concept. The subscript explicitly signals the relational element.

Armed with this account of what it is for something to be valuable, let's now reconsider the argument for the incoherence of relative value. According to hedonistic ethical egoism, in order to properly characterize the value-making feature that explains why I ought to maximize *my* pleasure, you ought to maximize *your* pleasure, and others ought to maximize *their* pleasure, we must mention an agent or evaluator: *my* pleasure in the case of my obligations, *your* pleasure in the case of yours, the *others* pleasure in the case of the others. They therefore conclude that the values in question are relative, not neutral, a conclusion we can formulate as follows:

$\forall x$ (x's pleasure is valuable_x)

or, equivalently given our paraphrase of Sidgwick,

$\forall x$ (x is required by reason to desire that x has pleasure, or to approve of x's having pleasure, or to aim at x's having pleasure)

But once we formulate their idea in this way, it becomes clear that the earlier argument for the incoherence of hedonistic ethical egoism turns on a crucial

equivocation. For what follows from the fact that I ought to maximize my pleasure is not that my pleasure is valuable, but that my pleasure is valuable_{me}, and what follows from the fact that you ought to maximize your pleasure is not that your pleasure is valuable, but that your pleasure is valuable_{you}. And since it does not follow from the fact that my pleasure is valuable_{me} and that yours is valuable_{you} that your pleasure is valuable_{me} or that my pleasure is valuable_{you}, so it follows that there is no way to make the equivocation innocent. Your pleasure simply isn't relevant to my obligations, nor is mine relevant to yours, because though reason is on the side of my desiring or aiming at my pleasure, it isn't on the side of my desiring or aiming at yours, and vice versa.

Is it surprising that facts about my obligations are fixed by what's valuable_{me} and that facts about your obligations are fixed by what's valuable_{you}? The answer is that it isn't surprising at all, given that we are each obliged to do is, fixed by what we have normative reason to do, and what we each have normative reason to do is fixed by what we are each required by reason to desire, or approve, or aim at (or, to put the point in Scanlon's terms, what we would each desire or approve or aim at if we were maximally sensitive to all of the reasons that there are for desiring or approving or aiming at things). It simply reflects the fact that I may be required by reason to desire a different state of affairs to that which you are required by reason to desire: I may be required to desire that I have pleasure, whereas you are required to desire that you have pleasure. Absent some additional argument or consideration, it is thus simply false that either of us has any normative reason at all, hence any obligation, to bring about each other's pleasure. The quite general argument given earlier for the incoherence of relative value thus collapses.

V. CONSEQUENTIALISM, RELATIVE VALUES, AND PROMISSORY OBLIGATION

The way is therefore clear for a consequentialist to construct an account of promissory obligation grounded in distinctive relative values. But how might he go about doing this? How, in other words, might he try to identify the distinctive relative values that are realized by the fulfillment of our promissory obligations? The consequentialist's best strategy, I think, is to take a leaf out of the book of those who claim to construct nonconsequentialist theories of promissory obligation. Let me briefly illustrate the strategy I have in mind by looking at Scanlon's account of promissory obligation.

According to Scanlon, our obligation to keep the promises we make is explained by (very roughly) the impermissibility of our failing to meet the reasonable expectations that we knowingly create when we make promises.

Though these expectations are ones that we might create without explicitly making promises, promising is one conventional way such expectations can be created. He formulates this idea in the following principle, a principle he dubs "Principle F":

If (1) X voluntarily and intentionally leads Y to expect that X will ϕ (unless Y consents to X's not doing so); (2) X knows that Y wants to be assured of this; (3) X acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) Y knows that X has the beliefs and intentions just described; (5) X intends for Y to know this, and knows that Y does know it; and (6) Y knows that X has this knowledge and intent; then, in the absence of special justification, X must ϕ unless Y consents to ϕ 's not being done. (1998, p. 304)

Principle F is a nonconsequentialist principle, according to Scanlon, because we cannot explain why it is true simply by appealing to the value of the outcome of X's ϕ -ing. But is this true?

There is, I think, a real problem in supposing that it is true. After all, Scanlon presumably has to think that there is a reason for each person to want that (very roughly) when he knowingly creates certain reasonable expectations, he meets the reasonable expectations he thus creates. Each person has to think that there is such a reason because, absent such a reason, people would have no normative reason to do what Principle F tells them they must do. Scanlon's own Principle F would in that case break the connection between what agents are obliged to do and what they have normative reason to do. The trouble is, however, that if there is such a reason then it follows immediately that there is value_{each} in each's meeting of the reasonable expectations each knowingly creates. This follows immediately because, to put the point in Scanlon's terms, all that p's being valuable_{each} amounts to is each's having a reason to desire that p (or, in terms of our paraphrase of Sidgwick, all that p's being valuable amounts to is each's being required by reason to desire that p).

The upshot is that Principle F itself seems to presuppose that there is a nonwelfarist *relative* value in an agent's meeting of the reasonable expectations that he knowingly creates. The value in question is nonwelfarist because there is no reason to suppose that meeting such expectations contributes, as such, to the agent's well-being. And the nonwelfarist value in question is relative because the desires that there is reason for me to have concern the reasonable expectations that *I* knowingly create; the desires that there is reason for you to have concern the reasonable expectations that *you* knowingly create; and the desires that there is reason for others to have concern the reasonable expectations that *they* knowingly create. Moreover, Principle F's truth depends on the existence of such relative values because, as we just saw, the existence of such relative values is

required for agents to have a normative reason to do what Principle F says that they must do. Far from Principle F's being a nonconsequentialist principle, then, it turns out that Principle F is itself an implicitly consequentialist principle. It is an implicitly consequentialist principle because the facts about obligation that it describes are all ultimately grounded in, and thus reduce to, facts about the relative value of agents meeting the reasonable expectations that they knowingly create.

Scanlon might reply that this argument is predicated on a false assumption about Principle F. He might suggest that the argument assumes, falsely, that Principle F is an ultimate moral principle. But, he might say, Principle F is not an ultimate moral principle, but is rather a principle that is derived from the ultimate *contractualist* moral principle. As he puts it elsewhere:

in order to decide whether it would be wrong to do X in circumstances C, we should consider possible principles governing how one may act in such situations, and ask whether any principle that permitted one to do X in those circumstances could, for that reason, reasonably be rejected. (1998, p. 195)

Principle F, he might say, is a derived moral principle because it is one of the principles governing how one may act that no one could reasonably reject. What makes it the case that violations of Principle F are wrong is thus, ultimately, the fact that no one could reasonably reject acting in accordance with it.

The trouble with this reply, however, is that it has no affect on the cogency of the argument just given. For all it shows is that, though we do indeed have a reason to desire that we meet the reasonable expectations that we knowingly create, this desire is itself derived from something much more general that we have reason to desire, namely, that we each act only on those principles that no one could reasonably reject. But the fact that we each have a reason to desire this itself shows that the relative value that grounds the truth of Principle F is itself derived from a much more fundamental relative value, namely, the nonwelfarist relative value of acting in accordance with principles that no one could reasonably reject. This is another nonwelfarist value, because there is once again no reason to suppose that meeting such expectations contributes, as such, to an agent's well-being. And the nonwelfarist value in question is once again relative because what I have a reason to desire, at the most fundamental level, is that *I* act in accordance with principles that no one could reasonably reject; what you have a reason to desire, at the most fundamental level, is that *you* act in accordance with principles that no one could reasonably reject; and what others have a reason to desire, at the most fundamental level, is that *they* act in accordance with principles that no one could reasonably reject. What we all have a reason to desire, at the most fundamental level, is thus that we can justify *our own* conduct to

others. Once again, it seems that Scanlon has no choice but to admit that there are such reasons, and hence to admit that there are such relative values, because, absent such reasons, and hence absent such relative values, we would have no normative reason to do what, according to the contractualist principle, we have an obligation to do: namely, to act in accordance with principles that no one could reasonably reject.

Note that this discussion of Scanlon's Principle F and of his more general contractualist principle are meant to be merely illustrative. All I have tried to indicate is how a consequentialist who has an open mind about the possibility of there being relative values that explain the distinctive nature of promissory obligation might go about trying to figure out what those relative values are. Nothing I have said shows that there is any relative value in agents' meeting the reasonable expectations that they knowingly create, and nothing I have said shows that there is any relative value in agents' acting in ways that they can justify to others either. What the discussion does suggest, however, is that, to the extent that we find Scanlon's Principle F and his deeper contractualist principle independently plausible, our finding them so amounts to our finding it independently plausible that there are indeed such relative values. To put the point somewhat contentiously and ambitiously, what the argument suggests is that the very best arguments for nonconsequentialism will all turn out, on closer inspection, to be arguments for consequentialism, albeit versions of consequentialism that are grounded in relative values that had hitherto escaped our attention.

ACKNOWLEDGMENT

An earlier version of this essay was presented at Promises and Agreements, a workshop held at Rice University, October 2008. I would like to thank all of workshop participants for their helpful comments during the workshop and Hanoch Sheinman for his very helpful written commentary.

REFERENCES

- Griffin, James. 1998: *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Oxford University Press).
- Hooker, Brad. 2000: *Ideal Code, Real World* (Oxford: Clarendon Press).
- Jackson, Frank. 1991: "Decision Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101: 461-482.
- Mackie, J. L. 1977: *Ethics: Inventing Right and Wrong* (Harmondsworth, England: Penguin).
- Mill, John Stuart. 1861: *Utilitarianism* (Toronto: University of Toronto Press).
- Moore, G. E. 1903: *Principia Ethica* (Cambridge: Cambridge University Press).

- Pettit, Philip, and Michael Smith. 2000: "Global Consequentialism." In Brad Hooker, Elinor Mason, and Dale E. Miller, eds., *Morality, Rules, and Consequences: A Critical Reader* (Edinburgh: Edinburgh University Press), pp. 121-133.
- Railton, Peter. 1984: "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13: 134-171.
- Raz, Joseph. 1986: *The Morality of Freedom* (New York: Oxford University Press).
- . 1999: "Agency, Reason, and the Good." In Raz, *Engaging Reason* (New York: Oxford University Press), pp. 22-45.
- Scanlon, Thomas M. 1998: *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).
- Slote, Michael, and Philip Pettit. 1984: "Satisficing Consequentialism." In *Proceedings of the Aristotelian Society*, supp., 58: 139-176.
- Smart, J. J. C., and Bernard Williams. 1973: *Utilitarianism: For and Against* (Cambridge: Cambridge University Press).
- Smith, Michael. 1994: *The Moral Problem* (Oxford: Wiley-Blackwell).
- . 2003: "Neutral and Relative Value after Moore." In *Ethics*, Centenary Symposium on G. E. Moore's *Principia Ethica* 113 (), pp. 576-598.
- . 2005: "Metaethics." In *Oxford Handbook of Contemporary Philosophy*, edited by Frank Jackson and Michael Smith (Oxford: Oxford University Press), pp. 3-30.
- Sumner, Wayne. 1996: *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press).

Chapter 9

Act-utilitarianism and Promissory Obligation

Alastair Norcross

Abstract

Act-utilitarianism is often criticized for failing to account for the strength of promissory obligation. Worse still, it is not even clear that the institution of promising could exist in an act-utilitarian society. Rule utilitarianism is often claimed to be in a better position than act utilitarianism with respect to providing an account of the moral status of promising (or rather keeping one's promises). In fact, the move from act utilitarianism to rule utilitarianism is often motivated by the desire to justify the obligation to keep promises. However, the act utilitarian has multiple resources available to explain the moral status of promissory behavior. I argue that act utilitarianism provides a fully satisfactory account of the moral considerations relevant to promising.

The standard maximizing act-utilitarian account of rightness is attacked both for being too demanding and for being too permissive. The theory is said to be too demanding, because it counts the interests of all equally. Thus, if the \$10 I propose to spend on a movie can produce more good for others than the good I will get from seeing the movie—which it undoubtedly can—I should forgo the movie and spend the \$10 on improving the lot of others. The theory is said to be too permissive, because it permits agents to break intuitively appealing moral rules whenever more good (on balance) will come from breaking them than from keeping them. In fact, it doesn't merely permit agents to break rules in such circumstances, but it demands that they do so.

One of the most common examples of this latter criticism concerns the rule that one should keep one's promises. Few philosophers, with the exception of some (hopefully) imaginary Kantians, claim that this rule has