7

Evaluative Focal Points

Shelly Kagan

Foundational consequentialists believe that justification in normative ethics is ultimately a matter of appealing to the goodness or badness of the consequences (in some suitably broad sense of 'consequences').[1] Famously, act consequentialists appeal to the consequences directly in evaluating any given act: an act is right just when the consequences of performing the act would be as good as those of any alternative act available to the agent. In contrast, rule consequentialists do not evaluate acts in this manner, that is, directly in terms of the good. Rather, they evaluate the given act in terms of a set of optimal rules, and it is only the rules that are themselves evaluated directly in terms of the good.

Acts and rules are two examples of what I will call *evaluative focal points* (other examples include motives, norms, character traits, decision procedures, and institutions). Act consequentialists and rule consequentialists share the foundational consequentialist thought that justification must be in terms of the good, but they differ in their choice of which evaluative focal point to make primary. The rule consequentialist makes the *rules* the primary evaluative focal point, evaluating the rules directly in terms of the good; she then evaluates the other focal points indirectly, in terms of the rules: thus, for example, acts are not evaluated directly in terms of the good, but only indirectly, via the rules. In contrast, the act consequentialist – as I will be understanding this position – makes the *act* the primary evaluative focal point, evaluating the act directly in terms of the good: other focal points, such as rules, are evaluated only indirectly. (This last claim is potentially misleading, involving, as it does, a bit of nonstandard stipulation; but we will return to it later.)

Since rule consequentialism evaluates the act in terms of the good only indirectly, rule consequentialism is an example of an 'indirect' consequentialist theory. There are other indirect consequentialist theories as well: thus, motive consequentialism takes motives to be the primary evaluative focal point, selecting the optimal motives directly in terms of the good, and then evaluating acts indirectly, in terms of the optimal motives; while decision procedure consequentialism holds that the primary evaluative focal point is the decision procedure, and evaluates acts indirectly, in terms of the optimal decision procedure; and so on. For simplicity, in the bulk of this paper I will be restricting my attention to act and rule consequentialism. But virtually everything I say carries over *mutatis mutandis* to these other consequentialisms as well.

Rule consequentialism has seemed to many to be open to the charge of 'rule worship'.[2] If the best act, as revealed via a direct appeal to the good, differs in some case from that prescribed by the rules, isn't it irrational to continue to insist on compliance with the rules? What is so special about *rules*? If what is of ultimate importance is good consequences, shouldn't we evaluate acts directly in terms of their consequences? (Similarly, motive consequentialism seems guilty of motive worship, norm consequentialism seems guilty of norm worship, and so on.)

Once we start thinking of all consequentialist theories as facing a choice between focal points, however, this charge seems stripped of some of its force. After all, act consequentialism (at least, as I have characterized it) makes its own choice of a primary evaluative focal point – acts – and evaluates the other focal points in terms of the primary one: the best rules, for example, might be those that direct us to perform the best acts. But this means that act consequentialism could with equal justice be viewed as an indirect consequentialist theory as well: the rules are not evaluated directly in terms of the good, but only indirectly, in terms of the best acts. Act consequentialism is admittedly a direct theory with regard to *acts*, but it is an *indirect* theory as far as the other focal points are concerned.

And having said this, we immediately see that rule consequentialism may indeed be an indirect theory as far as *acts* and other focal points are concerned, but it is a *direct* theory when it comes to rules. The language of direct and indirect consequentialism thus seems loaded in favor of act consequentialism: it is only if we are tacitly assuming that it is acts that somehow 'really' deserve to be the primary evaluative focal point that the standard labels will seem appropriate.

To put the same point another way: if rule consequentialism is guilty of rule worship, then act consequentialism is guilty of act worship. ('If the good is ultimately what it is all about, shouldn't rules be evaluated directly in terms of the good, rather than only indirectly, in terms of acts?')

That the two theories are in structurally similar boats can be seen quite easily if we diagram their structures, like this:

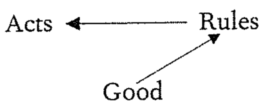

Figure 7.1 Rule consequentialism.

In Figure 7.1 we have a picture of rule consequentialism. I've placed the 'good' at the bottom of the diagram to represent the foundationally consequentialist view that justification in normative ethics is ultimately a matter of appealing to the goodness of the consequences. I've drawn the various evaluative focal points (here, acts and rules) above. The arrow going from the good to rules indicates that in rule consequentialism the rules are evaluated directly in terms of the good; while the arrow going from rules to acts indicates that acts are evaluated in terms of the best *rules*. Thus in rule consequentialism it is only the rules that are evaluated directly in terms of the good, while acts are evaluated only indirectly.

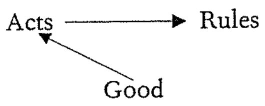

Figure 7.2 Act consequentialism.

In Figure 7.2 we have a picture of act consequentialism. Since it too is a foundationally consequentialist view, the good is again drawn at the bottom of the diagram. But now it is acts (rather than rules) that are evaluated directly in terms of the good, and rules (rather than acts) that are evaluated only indirectly.

Drawn this way, it seems clear that the choice between act and rule consequentialism turns on the question of what reasons there are to select one evaluative focal point over the other to be the primary evaluative focal point. Is there something about either acts or rules that makes them uniquely fit – or unfit – to be the primary evaluative focal point? (Alternatively, we might wonder whether we could somehow avoid elevating *any* focal point to the special status of being primary; this is a possibility I will consider below.)

I want to explore some aspects of this issue in what follows. Although I will be offering a number of arguments, some against rule consequentialism, some against act consequentialism, my primary concern is actually to try to illustrate and partially illuminate this general approach to the topic: I want to advocate thinking of moral theories in terms of the choices between evaluative focal points; and I want to get clearer about the indirect evaluation of secondary focal points in terms of the primary focal point.

Let's begin by focusing on rule consequentialism. The rules are evaluated directly in terms of the good. The optimal rules are those that would produce the most good, the best results overall. But rules – conceived as abstract linguistic items – don't generate results in and of themselves. Rather, they generate results only when they are concretely 'embedded' in some way: they generate results when they are thought about, taught, accepted, disdained, mentioned, mocked, acted upon, flouted, or what have you. So talk of selecting the rules that would have the best results is necessarily shorthand for talk of selecting the rules that would have the best results *if* they were embedded in some specified way. And the rules that are optimal relative to some set of embedding specifications might well not be optimal relative to some other embedding specifications.

Two basic types of specification seem worth thinking about: *ideal* embedding and *realistic* embedding. Ideal embedding is a matter of assuming perfect conformity to the rules. We would then be selecting the rules that would have the best results if they were perfectly conformed to. Here we are assuming that *everyone* conforms perfectly: thus, everyone is *motivated* to conform to the rules, everyone correctly *identifies* the act that is prescribed by the rules, and everyone flawlessly *executes* the identified act. As the name suggests, in assuming full and perfect conformity, ideal embedding makes a highly unrealistic, idealized specification.

Realistic embedding relaxes these assumptions in the direction of more accurately approximating a realistic scenario. It recognizes that for various possible rules, in any given case there may be some number of individuals that would lack the motivation to conform altogether; and some rules might be sufficiently complex or difficult to apply so that cognitive errors might arise in identifying the act that conforms to the rules, or performance errors might arise in executing the identified act. With realistic embedding, then, we do not assume perfect conformity; rather, we assume a more realistic level of conformity, with only imperfect or partial compliance. (Obviously, the rubric of 'realistic embedding' covers a family of distinct embeddings – depending on precisely how far one goes toward making the embedding approximate the current, actual embedding of the rules.[3] But for our purposes there will be no need to have more fine-grained distinctions.)

Among rule consequentialists we find advocates of both types of embeddings. And the choice here clearly matters, since the rules that would have best results if perfectly conformed to will quite likely differ from those that would have best results if we assume a more realistic level of conformity.

Now under rule consequentialism, *acts* are evaluated in terms of the optimal rules. But a further choice faces us when we specify how, exactly, the best or right act is to be identified. One familiar possibility is that the right act is the act that *conforms* to the optimal rules. Another way to view this is to see the right act as the act that would be performed if the optimal rules were ideally embedded, that is, under conditions of perfect conformity.

Given our previous distinction, however, this raises the possibility of a second way of identifying the right act: it is the act that would be performed if the optimal rules were realistically embedded. On this second approach, the right act is not necessarily the act that conforms to the rules, it is rather the act that *actually follows*, given a realistic embedding of the rules.

This second approach is less familiar in the context of rule consequentialism. But there are other moral theories where it is just this approach that seems to be accepted. For example, advocates of motive consequentialism often define the right act as the act that would in fact be performed by someone with the optimal set of motives. (Compare the theory that defines the right act as the act that would in fact be performed by the virtuous individual.) Here, the right act is not defined as the act that 'conforms' to the optimal motives; rather, it is the act that would actually be produced – the causal upshot – if one had the optimal motives.

Or consider one possible version of act consequentialism, which might hold that the best rules are those that would actually produce the best acts. (Under realistic embedding, these need not be the rules that simply direct the agent to perform the best acts.) Here, we are evaluating a secondary focal point (rules) that is causally 'further upstream' than the primary one (acts), and the suggestion is that the best rules are those which would, under realistic conditions, produce the best acts. But if the relation 'the Xs that would realistically produce the best Ys' can be used in evaluation going 'upstream' then it seems that the inverse relation – 'the Ys that would realistically be produced by the best Xs' – could be used when evaluating a secondary focal point that is causally 'downstream' of the primary focal point. It is just this, I have suggested, that is often done by motive consequentialism (and virtue consequentialism), and I see no reason why the same possibility should not be open to the rule consequentialist.

So when it comes to defining the right act in terms of the best rules, there are two possibilities: the right act is the act that *conforms* to the rules, or the right act is the act that would be the *upshot* of the rules (if realistically embedded). (Again, the act that would be the upshot under the assumption of ideal embedding would be the same as the act that conforms to the rules; so hereafter when I talk of the 'upshot' I mean upshot with realistic embedding.)

If there are two basic ways to specify the optimal rules, and two ways to identify the best act in terms of those rules, we have four basic versions of rule consequentialism. The right act is the act that:

1. *conforms* to *realistic* rules; [ideal/realistic]
2. is the *upshot* of *realistic* rules; [realistic/realistic]
3. *conforms* to *ideal* rules; [ideal/ideal]
4. is the *upshot* of *ideal* rules. [realistic/ideal]

What can be said for or against these various versions of rule consequentialism?

Consider, first, the version that defines the right act as the act that conforms to the realistic rules. I think such a theory is implausible. Indeed, I think there is an inherent implausibility in defining the right act in this way. To see this, it helps to notice an easily overlooked fact about how the optimal realistic rules may come to be optimal.

First, a complication. As I have already noted, until they are embedded, rules have – in and of themselves – no results at all, with which to be evaluated from a consequentialist perspective. However, once embedded, a variety of factors can affect the consequentialist 'score' that a rule (or a set of rules) receives. For example, once embedded, rules can have an impact on results that is independent of their impact on *acts*: it might be, say, that merely thinking about a set of rules reassures people, and so contributes to happiness. But for simplicity let us put aside such factors, and focus only on the impact that rules have by virtue of their effect on how we act.

Now when thinking about how it is that the optimal rules come to have the (relatively) high scores that they do, we are apt to focus on only one type of case – cases where as a result of the rules being embedded agents perform acts that conform to the rules, and those acts promote the overall good. But there is a second kind of case that may be relevant as well – cases of nonconformity.

The possibility of nonconformity obviously cannot be dismissed. Given realistic embedding, failures of motivation, identification, or execution can lead to imperfect conformity with any given rule. Given realistic embedding, even the optimal rules may sometimes yield acts of nonconformity. (To be sure, it is *possible* that the best rules would in fact have perfect compliance, but whether this is so is an open empirical question.)

No one is going to deny the possibility of nonconformity. But in thinking about nonconforming acts, we are typically inclined to assume that such acts will *lower* the score of the given rule: failure to conform to the rule will lead to worse results overall, and if the given rules are indeed optimal this is *despite* the 'losses' due to nonconforming acts. In fact, however, not all nonconforming acts need work to the detriment of the rules in this way: the fact that a rule produces a nonconforming act (when realistically embedded) can serve to *raise* the score of the rule. (One way this could happen is this: the rules are 'built' or 'designed' to take into account and match our various shortcomings – motivational, cognitive, and so on – so that we will end up violating them in just the ways that are actually preferable.)

Consider, then, the possibility of the following type of nonconformity case: under realistic embedding the agent would fail to conform to the rule, and this act of nonconformity actually produces *better* results than would be produced by an act conforming to the rule. (We might even throw in – for good measure – the possibility that it is the nonconforming act that strikes us intuitively as the right act to perform in this case.)

Such cases could play a significant role in helping to make it be the case that the optimal rules are indeed the rules with the best results under realistic embedding. In such cases of *desirable nonconformity*, part of the *virtue* of the rule from the perspective of rule consequentialism – part of the reason it gets a high score – will be the very fact that in cases of this sort, promulgation of the rules will produce *nonconformity* to those very rules. In short, one factor that might make a set of rules the optimal rules might be the very fact that in certain cases people will *violate* those rules, where this might be preferable both intuitively and in terms of promoting the good.

It is because of the possibility of such cases that it seems to me implausible to define the right act as the act that conforms to the optimal realistic rules. As just noted, the optimal rules might be optimal in part – indeed, perhaps in considerable part – because in certain cases they will be violated. In such cases, we *selected* the rules because of the fact that they would be violated. The acts of nonconformity are exactly what we are trying to produce – both intuitively and foundationally (that is, in terms of good results). In such cases it strikes me as simply bizarre to call the desired act *wrong*.

That is, it seems bizarre to insist that the right act, the morally preferable act, is the act that *conforms* to the rules – even though the optimal rules are optimal here by virtue of the very fact that they are going to be violated. It is, of course, an empirical question just how often such cases of desirable nonconformity arise for the optimal rules, and just how significant such cases are in determining the overall 'score' of the optimal rules; indeed it is a logical possibility that for the optimal rules such cases never actually arise at all. But to my mind this doesn't dampen the force of the criticism: if we are looking for the rules that would have the best results if realistically embedded, the possibility of such cases of desirable nonconformity is a live empirical one, and it simply seems bizarre to insist that the right act is the act that conforms to the rules – even when the rules may have been selected and designed in part so as to produce nonconforming acts, and it is one of these desirable nonconforming acts that we are evaluating.

In this light, the second version of rule consequentialism may seem more attractive. Here the right act is the act that would be the actual causal upshot of the optimal rules, where these rules are chosen relative to a realistic embedding. In cases of desirable nonconformity, where the agent violates the rules in a way that is both intuitively attractive and promotes the overall good, we will be able to classify the nonconforming act as the *right* act – since it will be the act that will actually be produced given the realistic embedding of the rules. (There is a natural harmony here between the embedding used to select the rules and the embedding

used to evaluate the acts: we assume realistic embedding to evaluate the rules, and then continue with realistic embedding to evaluate the acts. In contrast, the first version of rule consequentialism gets into trouble by demanding perfect conformity to rules that were selected on the basis of their value under conditions of imperfect conformity.)

Unfortunately, this second version of rule consequentialism seems implausible in the face of the possibility of cases of *undesirable nonconformity*. What I have in mind are cases where the rules are violated, but the violation seem unattractive and unfortunate from the foundational consequentialist perspective – since the nonconforming act here fails to promote the good. (Once more, for good measure, we can throw in a reference to our intuitive judgment as well – this time the judgment that the nonconforming act is not the right act.) We would eliminate such acts of nonconformity if only we could. As it happens, we cannot realistically do so; and so such cases may be the inevitable undesirable fallout of our various shortcomings. Here the optimal rules are selected so as to minimize such acts of morally undesirable nonconformity. But to the extent that an ineliminable residue of such acts remains, it seems bizarre to assert that these acts are in fact the *right* acts to perform in the circumstances. Yet this is just what we must say if we define the right act as the act that would in fact be performed – the causal upshot – if the optimal rules were realistically embedded. So I take it that this second version of rule consequentialism should be rejected as well.

In short, if the rules are selected realistically, neither version of rule consequentialism is plausible. Realistic rules cannot provide a plausible standard for right acts.

(I believe that these results can be generalized, and would hold for other indirect consequentialist theories – theories that select some other focal point as primary, and evaluate acts in terms of the primary focal point – given that the primary focal point is selected on a realistic basis: if we define the right act as that which conforms to the optimal focal point, we overlook the possibility of morally desirable nonconformity; if we define the right act as that which would be the actual upshot of the optimal focal point, we overlook the possibility of undesirable nonconformity. Thus, realistically selected focal points cannot provide a plausible standard for right acts.)

What of our third and fourth versions of rule consequentialism, which evaluate acts in terms of *ideal* rules (that is, the rules that would have the best results if ideally conformed to)? The fourth version defines the right act as the act that would be the actual upshot of realistically embedding the ideal rules. The difficulty here is the same as that facing the second version, namely, cases of undesirable nonconformity. If we *realistically* embed the ideal rules, and ask what acts will actually be performed given this embedding, we have to face the possibility that the ideal rules will not themselves be perfectly conformed to, and some of these violations may well be *undesirable* – both intuitively and foundationally. It seems unacceptably bizarre to insist that such acts are right nonetheless. If anything, the objection seems even stronger here: the ideal rules were selected because of the results that they would have under conditions of perfect conformity; it would be bizarre to suggest that a *nonconforming* act that is intuitively unacceptable and that leads to bad results should, for all that, be classified as the right act to perform.

(Generalizing: indirect theories that define the right act as the realistic upshot of the optimal version of the primary evaluative focal point cannot provide a plausible standard for right acts – regardless of whether the primary evaluative focal point is selected on an ideal or a realistic basis.)

This leaves only the third version of rule consequentialism – where the right act is defined as the act that conforms to the ideal rules. Since the fourth version faced the same problem as the second – cases of undesirable nonconformity – we might well wonder whether the third will face the same problem as the first, that is, cases of desirable nonconformity. Do these also plague our final version of rule consequentialism? I believe so, although the problem does not arise in precisely the same way that it did for the first version of rule consequentialism. With the first version, we were dealing with realistic rules; that is, rules selected for the results they would have under realistic embedding: this opened the possibility that the optimal rules might be optimal in part because of situations in which they would be violated. But with our final version of rule consequentialism we are dealing with ideal rules; that is, rules selected for the results they would have under conditions of perfect conformity. Here it cannot be that the optimal rules are optimal in part because they would be violated; since we are selecting the optimal rules under the assumption that they will *not* be violated.

But for all that, I think there remains the possibility of cases where conformity to the optimal ideal rules is morally unattractive (from both a foundational and an intuitive point of view). If even the *ideal* rules can face cases where it is *nonconformity* that is desirable, then defining the right act to be the one that conforms to these rules – even in cases of this kind – seems problematic. Once more, to my mind at least, it seems bizarre to suggest that the right act, the morally preferable act, would be one that conforms to the rules, even though violating the rules would be better both intuitively and in terms of promoting the good.

However, this objection may seem less compelling than the earlier objections. Note that in all of the previous objections the use of realistic embedding – either at the level of selecting the rules, or at the level of evaluating the acts – introduced imperfect compliance, which in turn left cases of undesirable nonconformity and desirable nonconformity as live *empirical* possibilities. Even if the problem cases did not in fact arise for the optimal rules, this was mere 'luck' from the conceptual standpoint. They couldn't be ruled out, yet they led to unacceptably bizarre evaluative claims.

But when it comes to the current objection, it is far from obvious whether the possibility of desirable nonconformity is anything more than a verbally describable cubbyhole. Since our final version of rule consequentialism appeals to ideal conformity to ideal rules, unsettled empirical possibilities don't seem to arise. Given that the optimal ideal rules are those that have the best results when conformed to, can't we rule out a priori the possibility of desirable nonconformity? If so, then the rule consequentialist could admit that such cases – were they genuinely possible – would be problematic; but since they are not genuinely possible, they raise no serious difficulty.

To see whether cases of desirable nonconformity *can* be ruled out, we first need to consider the familiar question of whether rule consequentialism 'collapses' into act consequentialism.

It seems reasonably clear that for versions of rule consequentialism that appeal to realistic rules, the answer will be no. Given realistic embedding, it is not likely that the optimal rule will be a statement of act consequentialism, or some set of rules extensionally equivalent to it.

But what if we are dealing with *ideal* rule consequentialism? Here there is a plausible argument that suggests that the optimal rules *will* be extensionally equivalent to act consequentialism. Roughly, the idea is that if the purportedly optimal rules ever prescribed a nonoptimal act there would be some revised version of the rules that differed only in prescribing the optimal act; since the revised rules would have better results if conformed to, *they* would be the optimal rules (given the assumption of ideal embedding).

It seems to me, however, that this plausible argument is mistaken. Ideal rules are selected for the results they would have given ideal embedding – that is, within a world of ideal and full compliance. In such a full compliance world the directives of the optimal ideal rules will indeed never diverge from act consequentialism.[4] But will the optimal ideal rules contain clauses governing *imperfect* compliance worlds? It is far from obvious that there *will* be such clauses. After all, they would do no work under the assumption of ideal embedding – so why would they be added?

Of course, they will do no harm if added, since under the assumption of perfect compliance, clauses governing imperfect compliance will never be operative. So perhaps such clauses can be added after all. But since clauses governing imperfect compliance worlds will remain inoperative under the full compliance assumption, the perspective of a perfect compliance world offers no grounds for choosing between the alternative possible partial compliance clauses. Perhaps clauses will be added that are stupid and unattractive (whether intuitively, or foundationally). The full compliance world offers no *basis* for selecting plausible or attractive partial compliance rules.[5]

Perhaps this is too quick. Perhaps it will reassure the people in the full compliance world to have rules governing partial compliance situations. Yet why do these people need reassurance? Don't they realize that they are in a perfect compliance situation? (If not, then the rules that would be best for *their* world may need altering in light of that fact; and this moves us to a version of realistic embedding, with all its attending problems.) And at any rate, who says that the partial compliance rules that would be best suited for reassuring people (when inoperative, as under full compliance) would be best (intuitively, or foundationally) for governing *actual* situations of partial compliance? In short, there is no particular reason to think there *will* be rules governing partial compliance; and if there are any, there is no good reason to think that they will be attractive ones. If they are, this is really just a matter of luck.[6] The standpoint of perfect compliance is a poor choice for evaluating rules governing more realistic levels of conformity.

Since *act* consequentialism does entail directives governing partial compliance, directives that are at the very least attractive from the foundational perspective, the argument that rule consequentialism collapses into act consequentialism seems to me unsuccessful. But in the course of diagnosing where the argument fails I think we have established that cases of desirable nonconformity are indeed a live and pressing possibility. Since the ideal rules are those that would be optimal given perfect conformity, there is every reason to worry that the optimal ideal rules would sometimes offer no advice, or disastrous advice, in situations of partial compliance. In the light of this possibility, it seems implausible to *define* the right act as the act that conforms to the optimal ideal rules. (If no clauses explicitly govern partial compliance it will be just luck if applying the full compliance rules doesn't lead to unattractive results; if there are partial compliance rules, it will be just luck if they are attractive ones.)

This is an abstract and general statement of the 'disaster' objection to rule consequentialism: either there are no disaster clauses, or there is no good reason to think that there will be plausible disaster clauses. Of course *realistic* versions of rule consequentialism can easily answer this objection.[7] But – as we have seen – they face their own problems.

I conclude that rule consequentialism is an inherently implausible type of theory – in all four versions. Whether we appeal to ideal or to realistic rules, and whether we define the right act in terms of conformity to the rules or as the causal upshot of the rules, rule consequentialism cannot provide a plausible standard for evaluating right acts.[8]

Does this mean that we should become act consequentialists? No – for act consequentialism (at least, as I have characterized it) is itself an implausible view. Here I can be somewhat more brief. Recall that act consequentialists evaluate acts directly, but rules only indirectly. Yet just as rule consequentialists cannot provide a plausible standard for evaluating the best acts in terms of the best rules, act consequentialists cannot provide a plausible standard for evaluating the best rules in terms of the best acts.

First, as to the evaluation of acts. Considered as abstract entities, act-types (like rules) cannot have results in and of themselves: they need an embedding. But any given act token will obviously come with a 'natural' embedding – the actual one – and so for simplicity I am going to restrict our attention here to versions of act consequentialism that evaluate acts in terms of this actual embedding. So the right act, or the best act, is the one that will in fact lead to the best results. (Other possible embeddings include the 'subjective' embedding – that is, a world in which the agent's beliefs are true – as well as other more exotic or idealized embeddings.)

Next, with regard to the evaluation of *rules*, the two most significant choices are the inverses of the two relations we have already discussed: conformity and upshot. Just as we can look for the act that conforms to a given rule, so we can look for the rule that enjoins or prescribes a given act. And just as we can look for the act that is the actual upshot of a given rule, so we can look for the rule that would actually produce the given act. Thus we have two versions of act consequentialism: the first says that the best rules are those that prescribe the right acts; the second says that the best rules are those that produce the right acts (given realistic embedding). (The first version is equivalent to one that says that the best rules are those that would produce the right acts given *ideal* embedding.)

Let's begin with the first view, according to which the best rules are those that prescribe the right acts. On this view, it should be noted, rules are not to be evaluated with an eye to their success in actually leading us to perform right acts. Indeed, questions about the actual causal upshot of the rules are simply irrelevant. Rather, the best rules are simply those that best enjoin or prescribe the right acts. Is this a plausible basis for evaluating rules?

I think not. To see this, consider the following rule: 'Do the right thing.' In terms of what this rule *says*, it is impeccable. If all agents were to conform to it perfectly, they would perform all the right acts.[9] The rule is completely correct in what it tells us to do, and it gives us correct advice in every conceivable situation. Thus, according to the version of act consequentialism we are currently considering, this rule is perfect; it is the best possible rule.

It is possible, I suppose, that there may be other rules (other than mere notational variants of this first one) that would be ranked just as highly in terms of prescribing all and only the right acts. Luckily, we need not pursue this question, for the point remains that as far as this first version of act consequentialism is concerned, the rule that tells us to do the right thing is *perfect*. It cannot conceivably be improved upon;

it is flawless. It is important to be clear on this point. This first version of act consequentialism holds that rules are to be evaluated solely in terms of the extent to which they enjoin right acts. The best rule, therefore, must be the one that enjoins us to perform all and only right acts. Thus, according to this version of act consequentialism, the simple rule 'Do the right thing' is as good a rule as one could possibly hope for: it is indeed perfect.

But this is an absurd view. Although the advice given by the rule is certainly correct, the rule is virtually useless if it is not supplemented by further rules, rules that help us to identify the right acts. In the eyes of the current version of act consequentialism, however, there is simply no need for such supplementation. Concerns about our *actual* ability to correctly identify right acts are simply beside the point, for in evaluating rules we are, in effect, entitled to assume perfect conformity. That is, in evaluating the rules, we are to assume a perfect ability to identify the acts enjoined by the rules. Thus the *only* relevant question is what rules enjoin the right acts. According to this first version of act consequentialism, then, the rule that tells us to do the right thing is complete in itself, perfect for every situation, every choice, every circumstance.

But this is, as I say, absurd. In evaluating rules we need to at least have open the *possibility* of taking into account our ability to use the rules correctly. No doubt, there may be some purposes for which it is indeed appropriate to assume an ideal ability to identify the particular acts enjoined by a given rule. But for many other purposes, obviously, we will want to make more realistic assumptions about our ability to use the particular act-identifying information provided by the rule. That is, in at least some cases we will want rules that can give us more substantive, concrete, realistically *usable* guidance. But this is a consideration that our first version of act consequentialism is necessarily oblivious to. I conclude, accordingly, that this first version of act consequentialism provides an inadequate basis for evaluating rules.

This suggests, however, that our second version of act consequentialism may be considerably more attractive. Here, we do not simply assume perfect conformity in evaluating the rules. Rather, we assume realistic embedding, and ask what the *actual* upshot of the rules would be. When rules are evaluated in this way, obviously enough, it will hardly be irrelevant to ask to what extent people are actually able to identify the specific acts enjoined by the rules. On this view, in effect, rules are not to be evaluated in terms of what they *tell* us to do, but rather in terms of how *successful* they would be in actually getting us to perform right acts.

According to this second version of act consequentialism, then, rules are evaluated in terms of their ability to actually produce right acts, given realistic embedding. In this way, the second version of act consequentialism escapes the objection that I have just raised against the first. It is important to remind ourselves, however, that this view is still a version of act consequentialism. It is only *acts* that are evaluated directly in terms of the good. Rules are evaluated only indirectly, in terms of their success in producing right acts. According to this second version of act consequentialism, then, the best rule is the one that would actually be *most* successful in producing right acts.

I imagine, however, that anyone who shares the foundational consequentialist thought will think that this concern with *right* acts per se is misplaced. What matters is not whether the right act is performed – what matters is whether the *good* is promoted.

Now this may seem a misplaced criticism. Since (from the perspective of act consequentialism) the right acts *are* the acts that promote the good, the way to promote the good *is* to perform right acts; and so evaluating the rules *in terms of* the right acts can hardly be misguided.

But the distinction is worth maintaining for all that, in light of the fact that under certain circumstances certain *kinds* of acts might be right more often than any other readily identifiable kinds, and yet performing acts of those kinds might do rather little good (as compared to alternatives). In such cases, concern for the *rightness* of the acts per se seems misguided. Accordingly, selecting *rules* on the basis of their success at actually causing us to perform right acts should strike us as misguided as well.

Here is a schematic case to bring out the point. Suppose that I must choose between two acts, A and B, but I cannot tell which of two scenarios I am in. Under the first scenario, if I perform act A, results improve by 1 unit; if I perform act B, there is neither improvement nor loss. Under the second scenario, if I perform A, disaster strikes and results deteriorate by one million units; while if I perform B, there is, again, neither gain nor loss. Finally, suppose that I must play ten times before learning the results of any one round, and I know that in exactly one round (but I can't tell which) I will be in the second scenario (see Table 7.1).[10]

Table 7.1 Schematic case representing acts A and B.

	Act A	Act B
Scenario 1	+1	0
Scenario 2	-1,000,000	0

What rule would be best for dealing with such a case? According to act consequentialism, rules are to be evaluated in terms of the best or right acts. And in particular, according to the version of act consequentialism we are currently examining, the best rule will be the one that is *most* successful in terms of actually causing us to perform right acts.

But what rule will this be? Recall that nine times out of ten I will be in the first scenario, and in this scenario the right act – the act with the best results – is act A. Perhaps, then, the best rule will simply tell me to pick A. Admittedly, when I am in the second scenario the right act is act B. But it is important to keep in mind that this second scenario arises only once every ten rounds. Thus A is the right act to perform *far* more often than B. Apparently, then, from the standpoint of our current version of act consequentialism the best rule will direct me to always choose A. Such a rule would result in my doing the right act nine times out of ten, and so do better in this regard than any other available rule.

This last claim might be disputed. After all, a rule that tells me to always pick A will only result in my doing the right act *nine* times out of ten, rather than all ten times. It will go wrong in the single case where it is B that is actually the right act (rather than A). Wouldn't the best rule be one that directs me to the right act in *all* cases?

But what would such a rule look like? What would it say? We can, of course, imagine rules that tell me to pick A, except when B would have better results, or to pick A, except when I am in the second scenario. But the trouble with such rules, obviously enough, is that I simply cannot tell when I am in the second scenario, so cannot tell when B would have better results. Given the stipulation that I have absolutely no way to tell when I am in the second scenario, rules of this sort will be of no particular use in helping me to perform the right act in all ten cases. Indeed, if a rule like this did ever lead me to pick B rather than A, the most likely result is that I would simply end up performing *fewer* right acts (I am nine times as likely to make a mistake, if I pick B, as I am to get it right).

Provided, then, that we are looking for the rule that would be most successful – given realistic embedding – in leading us to perform right acts, it seems that we must conclude that the *best* rule would simply direct me to always pick act A. This, I take it, is what our current version of act consequentialism must claim.

But this is absurd. It is absurd to suggest that the *best* rule tells me to always pick A, since following such a rule is guaranteed to lead to *disaster* (a net loss of 999,991 units). Admittedly, this rule does direct me to the right act more often than not, but it is obviously a complete failure with regard to the promotion of the overall good. In contrast, a rule that directed me instead to always pick B would be a very poor guide to right acts, getting it wrong nine times out of ten. Yet this would clearly be a preferable rule, since following this rule would lead to dramatically better results.

More abstractly put, the point is this: since we cannot always identify the right act, the rule that does best in this regard may be optimal by virtue of directing us to the right act in the 'unimportant' cases, and so do quite poorly in terms of promoting good results. But it is, accordingly, absurd to suggest that such a rule is indeed the best rule. Yet this is just what the current version of act consequentialism does, in so far as it evaluates the rules in terms of the best acts, rather than directly in terms of the good. In short, this version of act consequentialism does not provide a plausible standard for evaluating rules.

A natural objection to this argument suggests itself. I have assumed that when the act consequentialist asks which rule is best at promoting right acts, all right acts are to be counted *equally*. Thus, the relevant question is simply which rule actually produces the *most* right acts. It is only if this is the appropriate standard for evaluating acts that the act consequentialist must claim – unacceptably – that the rule directing me to always pick A is the best rule.

Perhaps, however, this is not quite the relevant standard. Perhaps we should look for the rule that does best in terms of producing right acts – but only when those acts are *weighted* with an eye to their *significance*. Thus, on the one hand, if a given right act would produce very little good (or avoid very little bad), it should only count slightly in favor of a given rule that it would actually cause us to perform that act. On the other hand, if a rule would cause us to perform a right act that would produce a great deal of good (or avoid a great deal of bad), this should count rather heavily in favor of the rule in question.

If the act consequentialist adopts this revised standard for evaluating rules, then he will no longer claim that the best rule will direct me to always pick A. Admittedly, such a rule will lead me to perform right acts in nine cases out of ten, but these will be – as we have already noted – unimportant cases, and so will count only slightly in favor of the rule; meanwhile, the fact that it will keep me from performing B (even in the case where this is the right thing to do) will count quite heavily against this rule. In contrast, a rule which directs me to always pick B will do much better: although it will cause me to perform *fewer* right acts, when the significance of these acts is factored in this rule will emerge as clearly preferable.

Apparently, then, act consequentialism *can* provide a plausible standard for evaluating rules in terms of producing right acts, provided that we remember to weigh the importance of the right acts with an eye to how much good they produce.

I believe, however, that to do this is tantamount to abandoning act consequentialism. For to embrace this alternative standard is simply to claim that rules should be evaluated *directly* in terms of the goodness of their consequences. And this is to abandon the distinctive claim of the act consequentialist – that it is only acts that are to be evaluated directly in terms of the good, and that rules are to be evaluated only indirectly, in terms of the best acts. Under the proposed revised standard, after all, the fact that a given rule produces *right* acts does no real work whatsoever. The rule is evaluated, rather, simply in terms of the goodness of its results.

This is not to say, of course, that it is implausible to evaluate rules directly in terms of the goodness of their results. It is simply to take seriously the idea that act consequentialists (as I have characterized them) do *not* evaluate rules directly, but only indirectly – only in terms of their connection to right acts. Thus, the revised standard does not represent a genuine version of act consequentialism at all; it is merely masquerading as one.

Apparently, then, the act consequentialist must stick to the original standard, and claim that the best rule is the one that produces the most right acts. This is, of course, an implausible standard for evaluating rules, but it is exactly what emerges from the act consequentialist's insistence that rules are to be evaluated only indirectly – in terms of the best acts. It is for this reason that I conclude that even the second version of act consequentialism does not provide a plausible standard for evaluating rules. And since I have already argued against the first version of act consequentialism, I conclude as well that neither version provides a plausible standard for evaluating rules.

(Generalizing: act consequentialism may provide an inadequate basis for evaluating other focal points as well. An ideal approach runs the danger that the secondary focal point may be unhelpfully specified simply in terms of the very phrase 'right acts', while a realistic approach may end up giving too much weight to right acts that are altogether unimportant.)

Let's review. Rule consequentialism was implausible in its attempts to evaluate acts indirectly – via a distinct focal point – rather than directly in terms of the good. Act consequentialism is itself implausible in its attempts to evaluate *rules* indirectly – via a distinct focal point – rather than directly in terms of the good. The conclusion to draw, I think, is this: if there is a plausible version of consequentialism, it will evaluate *both* focal points – acts *and* rules – directly. Neither focal point will be elevated to the status of primary evaluative focal point; neither focal point will be evaluated only indirectly. A theory that is direct with regard to *all* the focal points might be usefully labelled *everywhere direct* (in contrast to the earlier theories that are direct only with regard to a primary focal point). But for simplicity, let us just call such theories *direct* – reserving the label for this extreme case. My suggestion, then, is that consequentialists should be direct consequentialists.

I can now happily admit that many who have called themselves 'act consequentialists' have actually been direct consequentialists all along, rather than being act consequentialists in my technical sense of that term. Many others, I suspect, have failed to distinguish between direct and act consequentialism, and so may not have had a determinate position in mind at all. (Note, in this regard, that in so far as many who call themselves act consequentialists are simply wed to a standard for right acts – the right act being the act that best promotes the good – this position underdetermines the choice between what I have called act consequentialism and direct consequentialism.) But at any rate, my concern is not to criticize the real or imagined confusions of other self-designated act consequentialists; it is only to make clear that direct consequentialism is indeed a distinct theory from act consequentialism (in *my* sense of the term), and it is the former, not the latter, which consequentialists should embrace.

It is easy enough to illustrate the structure of direct consequentialism (see Figure 7.3):

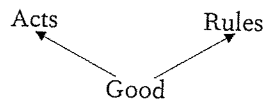

Figure 7.3 Direct consequentialism (with acts and rules as focal points).

And once this is done it is obvious that this theory does indeed differ from act consequentialism (see Figure 7.4).

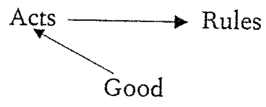

Figure 7.4 Act consequentialism.

Still, it is not at all obvious that direct consequentialism does not fall prey to difficulties of its own, difficulties that will be recognized as problematic even from the perspective of foundational consequentialism. I lack the space to do even a cursory job of addressing potential objections, but let me quickly clear up two points.

First of all, I have restricted our discussion to two focal points: acts and rules. But as I have already suggested, there are many other evaluative focal points that have been endorsed as *primary* focal points – such as motives, norms, institutions, and decision procedures. As I see it, the most plausible version of consequentialism will indeed be direct with regard to *all* of these. In fact, once we free ourselves from the thought that the evaluative focal points must be at least prima facie plausible candidates for the office of *primary* focal point, we realize that absolutely every kind of thing is a potential evaluative focal point (atoms, the weather, sewer systems, suns). So I believe that the most plausible version of consequentialism will be direct with regard to everything. Thus a more accurate illustration of the structure of direct consequentialism would actually look a lot like the sun (see Figure 7.5).

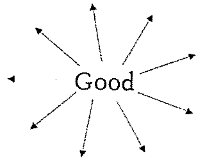

Figure 7.5 Direct consequentialism (with multiple focal points).

But then, second, what are we to say in cases of conflict? If the best rules direct us to different acts from the best motives, which in turn direct us to different acts from the best norms, and so on, what are we to do? Despite the appearance of difficulty, however, this question wears its answer on its own sleeve. If the question is what are we to *do* in the face of such conflicts, then the question is one about *acts*, and we already know the standard by which acts are to be evaluated: directly in terms of the good. Sometimes this means that the right thing to do will be to instill the best motive, and sometimes it will mean that the right thing to do is to promulgate some rule; sometimes we will have to choose between the two, and sometimes we will have to neglect both of them, for there will be something more pressing to attend to. But even if rightfully neglected, the best rule (say) is still the best rule for all that. So long as we remain clear about what *exactly* we are evaluating, the difficulty disappears.

I said at the outset that although I would be offering arguments against both rule consequentialism and act consequentialism, my primary concern was actually to try to illustrate the potential usefulness of thinking about moral theories in terms of their choices with regard to evaluative focal points. In the interests of this more methodological goal, let me make a few more general observations.

As I have already suggested, although only in passing, I believe that many of the arguments I have offered can generalize. Sometimes, perhaps, an issue turns on the specific nature of the focal point in question. But often it does not: I believe, for example, that the arguments I offered against rule consequentialism can be generalized to cover other versions of indirect consequentialism, versions that select some focal point other than rules to be primary.

But sometimes the arguments can be generalized even further. Little essential use was made in my arguments of the fact that we were working with foundational consequentialism. It seems to me that many of the arguments could have been stated in completely abstract terms ('if the given focal point is preferable from the standpoint of the given foundational theory . . .'). It would be helpful to know which of the arguments can be generalized in which ways. At the very least, this would save us the trouble of reinventing the wheel as we move from theory to theory. In short, I believe that evaluative focal points deserve study in their own right.[11]

Here is a quick final nod in that direction. As we have already seen, if we restrict our attention to two focal points, there are three basic types of theories available: two indirect and one direct. Here are the schematic diagrams (see Figure 7.6), abstracting away from the particular choice of foundational theory ('F') as well as the particular identities of the two focal points ('FP_1' and 'FP_2').

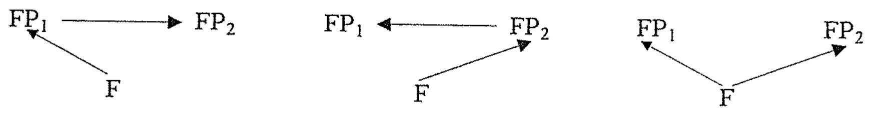

Figure 7.6 Possible theories with two focal points.

Obviously enough, if we introduce a greater number of focal points, the number of possibilities dramatically increases. Here are diagrams of some *sixteen* possible theories that arise once we have a mere *three* focal points (see Figure 7.7, overleaf).

Even here I have not given all the possibilities, since I have made the simplifying assumption that a given focal point is never evaluated on the basis of two other points; and both here and in the two focal point cases discussed above I have assumed that a focal point is never evaluated both directly *and* on the basis of another focal point.

Speaking personally, I have found it helpful to think about moral theories in terms of these structures.[12] I think it would be illuminating to see what kinds of structures are compatible with what types of moral foundations. It would be illuminating to see whether (and, if so, how) different foundational theories make different structures plausible, or whether – as I suspect, but certainly cannot prove – for any plausible foundational theory, the most plausible version of that theory will be a *direct* one.

I leave you, then, with three thoughts. First, if you share the belief in foundational consequentialism, you should be a direct consequentialist. Second, even if you prefer alternative moral foundations, you would still do well to consider seriously the merits of direct versions of those views. And finally, all of us – whether consequentialists or not – could profit from the study of evaluative focal points. This study I heartily commend to you.

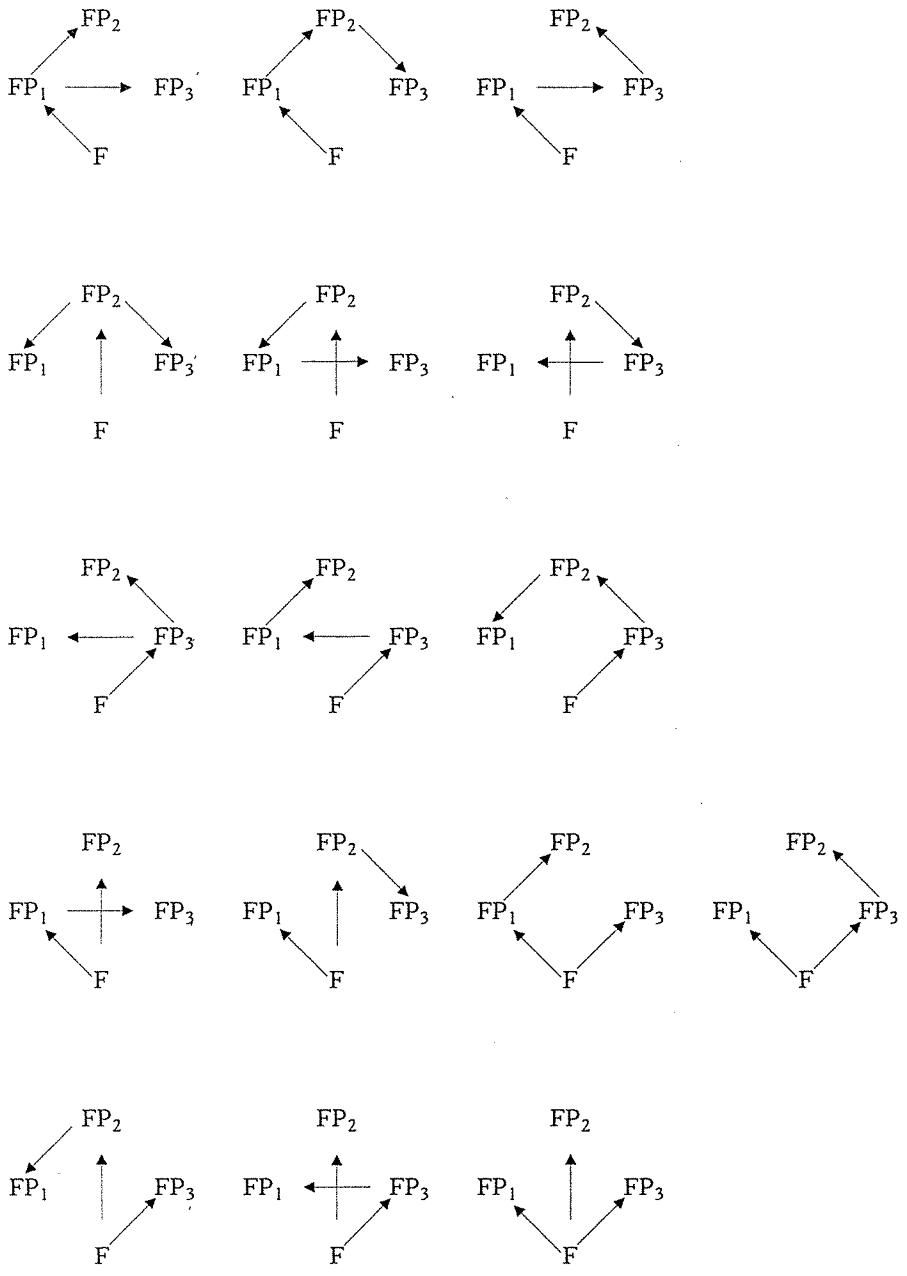

Figure 7.7 Possible theories with three focal points.

NOTES

1. I think we should distinguish between foundational consequentialism and *factoral* consequentialism – the view that consequences are the only factor relevant to determining the moral status of a given act. (I discuss this distinction in 'The Structure of Normative Ethics', *Philosophical Perspectives*, 6 (1992), pp. 223–42, and in *Normative Ethics* (Boulder, CO: Westview, 1998).) Rule consequentialists, for example, are typically only consequentialists at the foundational level, not at the factoral level. Act consequentialists, on the other hand, are typically factoral consequentialists as well (and I will assume so here; but see *Normative Ethics*, pp. 212–23, for discussion of this point).
2. The phrase, I believe, goes back to J. J. C. Smart.
3. At the limit, all rules have the same results under *completely* realistic embedding. Since that is the embedding that they all currently actually have, all rules have the same results – the actual results.
4. For a clear statement of the proof, see Donald Regan, *Utilitarianism and Cooperation* (Oxford: Clarendon Press, 1980). For simplicity, I've neglected the possibility that the rules might diverge from consequentialism by merely permitting rather than requiring the optimal act.
5. For a fuller discussion of this point, see *Normative Ethics*, pp. 228–35.
6. More precisely, the possibilities are these: (1) the ideal rules might be restricted in scope (whether explicitly or implicitly) to situations of ideal conformity, with none of the rules governing partial compliance situations; (2) the various ideal rules may be unrestricted in scope – governing both perfect and partial compliance situations; (3) some rules may govern perfect conformity situations, while other 'supplementary' rules govern partial conformity situations. If (1), rule consequentialism is either incoherent or incomplete. If (2) or (3), there is no reason to think that the rules – tested from the standpoint of perfect compliance – will provide suitable guidance for situations of imperfect compliance.
7. Though they may not be able to answer it adequately. See *Normative Ethics*, pp. 234–5.
8. I have been assuming, for simplicity, that under rule consequentialism the optimal rules are to be 'common' rules – that is, the same for everyone. What if we move to rule consequentialism with individualized rules? The same objections can still be generated – even the last (since ideal individualized rules might fail to provide plausible guidance given the realistic possibility of my past or future failure to conform). Only for a version that requires conformity to optimal ideal rules relativized to the *specific* choice situation can the 'disaster' objection be escaped: but at this point 'rule' consequentialism has indeed *collapsed* into act consequentialism (or, at least, into a theory that is like act consequentialism in evaluating acts directly in terms of the goodness of their consequences).
9. For simplicity, I put aside the possibility that in some cases two different acts might both be permissible, or right, even though we cannot – and need not – do both. In such cases, let us stipulate that 'do the right thing' is to be understood as requiring us to perform only one of the relevant right acts.
10. I owe this example to Joe Pabis.
11. I made this suggestion in 'The Structure of Normative Ethics' as well, using somewhat different examples.
12. For example, a contractarian theory that selects rules directly on the basis of the contract, and then uses these rules to evaluate both acts and institutions, is to be distinguished from a contractarian theory that selects *institutions* directly on the basis of the contract, uses the institutions to generate rules, and then evaluates acts in terms of the rules.