1. Consequentialism can seem threatened by the fact that choosing the action-option that seems to maximize good consequences all too often does not in fact maximize good consequences. Direct consequentialists do not act as well as some of those who deliberate differently. Those who decide what to do by, say, applying the Golden Rule tend to produce even better consequences than do direct consequentialists. In this respect, direct consequentialism is self-defeating.

In order to cope with this problem, consequentialists typically distinguish between viewing consequentialism as a criterion of right action, and as a decision procedure. If consequentialism is meant to supply us with the criterion of right action, then it simply identifies which characteristics of an action make it right. But accepting the consequentialist account of right-making characteristics, it is stressed, does not commit one to any particular decision procedure. Most importantly, one is not thereby committed to the thought that one should decide what to do by evaluating the consequences of one’s various action-options. One can think that the right action is the one with the best consequences, while being completely agnostic about how one should decide what to do.

Nevertheless, many consequentialists who recognize that direct consequentialism is self-defeating still voice support for some particular way of deciding what to do. They are not agnostic, after all. For example, if deciding what to do in some nonconsequentialist fashion—such as applying the Golden Rule—tends to produce the best consequences, then one should decide what to do by applying the
Golden Rule, rather than by calculating and evaluating consequences. They tend to explicitly favour some form of *indirect* consequentialism (Sidgwick 1981, Parfit 1984, Railton 1984). Even Bales himself, who emphasises as strongly as anyone the distinction between a criterion of right action and a decision procedure, can’t restrain himself from echoing Mill to offer us *some* counsel about how we should decide what to do:

“I confess I don’t know how [the act-utilitarian] is to determine in each and every case which of the acts open to him in fact would maximize utility. I do have a word of advice for him, however: look and see what kind of procedure has tended to work in given kinds of situations. . . .” (Bales, 1971, 263-4)

Procedures that have tended to produce good consequences in the past are likely to produce them in the future, and thus it is wise to decide what to do by adopting them. Such procedures can aid the act-utilitarian agent by “singling out, under immediately helpful descriptions, which of the acts open to him at a given time would [in fact] maximize utility” (264). So by rejecting a consequentialist decision procedure in favour of one more sensitive to the detailed texture of ordinary human life, the indirect consequentialist hopes to take advantage of whatever benefits the employment of a nonconsequentialist decision procedure may provide.

But here I argue that the consequentialist cannot straightforwardly appropriate the decision procedures of those averse to consequentialism in order to avoid self-defeat.
2. Why even think that direct consequentialism is self-defeating in ways that indirect consequentialism isn’t? We can catalogue the reasons direct consequentialism is self-defeating into two camps: those concerning the optimality of the agent’s actions, and those due to a decision procedure’s autonomous effects.

First, the optimality of the agent’s actions. The direct consequentialist advocates deliberating by thinking about which of one’s action-options will produce the most good, and acting accordingly. Call this decision procedure *DPD (Decision Procedure-Direct)*. Yet DPD is only one of a practically uncountable number of possible decision procedures. There are many ways one could decide what to do; for example:

- **DPGR** Treat others as you would have them treat you.
- **DPKant** Act only on those maxims you could will as universal law.
- **DPB** Do what the Bible says
- **DPKoran** Do what the Koran says.
- **DPWWJD** Do what Jesus would do.
- **DPH** Do no harm.
- **DPConv** Do what most people do
- **DPTT** Do what your enemy last did
- **DPF** Do the first thing that comes to mind
- **DPHed** Do what’s most pleasant

...
It would be easy (though not the best use of time) to continue adding to the list. There are very many decision procedures one could employ.

Now the indirect consequentialist notices that, on the whole, one will in fact produce better consequences by employing, not DPD, but some other particular decision procedure. Call this particular indirect decision procedure \( DP^* \). (\( DP^* \) might turn out be DPGR or DPTT or—what’s much more likely—one we haven’t specifically identified.) And let’s call the action that one’s decision procedure recommends \textit{choiceworthy}. If one followed DPD \textit{perfectly}, one’s actions would seem to produce better consequences than would one’s actions if one followed \( DP^* \) perfectly instead. But the follower of DPD does not always perform the choiceworthy action. And so the person who employs \( DP^* \) can produce better consequences than she would were she instead to employ DPD.

Call a decision procedure that one is unlikely to employ successfully \textit{difficult}. Why, then, is DPD difficult? First, it is notoriously hard to identify accurately just which of one’s action-options would in fact maximize good consequences.\(^4\) All too often, one’s guess how to maximize goodness is incorrect. So there are intellectual difficulties in identifying what to do. It’s often possible one can act better by employing some other decision procedure.

Decision procedures like “Maximize good consequences” can be difficult also because one is likely to fail to perform the choiceworthy action, because bias afflicts one’s implementation of the decision procedure. Common forms of bias include 1) favouring yourself over others, 2) favouring friends over strangers, and 3) favouring the
present over the future. And so one may frequently think that some suboptimal option (V-ing) is in fact optimal, because one does better personally if one Vs than one does by acting in whatever way in fact maximizes overall goodness. Bias is less likely to distort one’s employment of other more specific decision procedures, such as “Don’t kill.”

There are also motivational failures that can hinder successful compliance. It is not intellectually difficult to figure out how to implement the decision procedure “Don’t drink in bars after midnight.” Success is easy to verify; it is not tricky determining what a bar is, or when it is after midnight. But by focusing the agent’s attention upon drinking in bars after midnight, the decision procedure may encourage the very thing it forbids. If so, this decision procedure is difficult to follow—the person who would most benefit by complying with it is very likely to fail to comply with it. Much the same might be true of DPD; one may often be unmotivated to do something as abstract as doing what maximizes good consequences. The employment of DPD thus seems likely to lead one to often act akratically.

So, we have seen that there are various reasons why an agent who employs a decision procedure will not always perform the choiceworthy action. Let us call the actual action of an agent who employs a decision procedure — whether an instance of compliance or not — the resultant action. If one were to execute the decision procedure one employs perfectly, the choiceworthy action and the resultant action would be identical. But since one does not always comply with the decision procedure one employs, the resultant action is often different from the choiceworthy action.
Now it seems very likely that employing DPD does not lead one to act so as to produce consequences as good as one would produce by instead employing some other particular decision procedure. DPD is intellectually difficult, is especially vulnerable to bias, and may also be so abstract as to foster akrasia. The resultant actions of employing DPD are likely inferior to the resultant actions of employing some other possible decision procedure. So, DPD is likely to be self-defeating.

3. Using a decision procedure at least partially determines what one actually does. But using a decision procedure also has other effects. Following Gregory Kavka, these effects can be called the decision procedure’s autonomous effects. Consequences over and above the resultant actions (and their effects) of a decision procedure are the autonomous effects of that decision procedure. Even if two different decision procedures were to lead their employers to act identically, their use may have different autonomous effects.

Here are two ways the autonomous effects of a decision procedure affect the desirability of employing it.

First, note that employing a decision procedure takes time, some more and others less. This fact would not be a problem for a theory which prized reflective deliberation above every other endeavour with which it competed for time. After all, someone might insist that reflective deliberation just is the most valuable use of our time. But any such theory is highly implausible: other things being equal, a decision procedure which takes a good bit of time to implement will have worse autonomous
effects than one which can be implemented quickly. And employing DPD obviously takes a great deal of time.\textsuperscript{7}

There is a second sort of autonomous effect of employing a decision procedure. Employing a particular decision procedure influences the menu of action-options one can choose among. Some actions, which are options for those who employ one decision procedure (D2), may not be options for those who employ another (D1).\textsuperscript{8} And so even if the resultant action of some decision procedure is the best of the agent’s options, it still may be true that the agent would act even better (A9) were he to employ some other decision procedure. Different decision procedures structure one’s action-options in different ways, and so some decision procedure may be the best to employ because it opens up action-options not afforded by others. The fact that some decision procedure generates the best of its action-options does not mean that the person who employs it acts better than she would have acted had she employed some other decision procedure.

This bodes trouble for the direct consequentialist, of course, if the best action-option for the person who employs DPD is not as good as the resultant action-option for the person who instead employs some rival decision procedure. The person who deliberates nonconsequentially may have action-options unavailable to the direct consequentialist. Indeed, much ink has been spilt in the service of arguing that the person who deliberates consequentially will be shunned from a wide variety of valuable social relationships.\textsuperscript{9} Because the direct consequentialist pays no attention to the past, her word and her history mean little to others. Most of the joint endeavours that make life bearable are not issuing invitations to direct consequentialists. And so the best that
the direct consequentialist can do may be worse than what the practitioner of ordinary morality can accomplish.

Those are two very different sorts of autonomous effects of the use of a decision procedure, but there are likely many other kinds of autonomous effects of employing a decision procedure that also influence the desirability of its use. The negative autonomous effects of a decision procedure can militate against its use, no matter how effective employing it is in leading us to act well. **Now if DPD in particular has much worse autonomous effects than some other decent available decision procedure, then DPD is probably not the best decision procedure to employ.** The autonomous benefits of using one of the very many other possible decision procedures (e.g. DPGR, DPKant, DPConv, …) may more than outweigh the losses that stem from its suboptimal resultant actions. And so some other decision procedure very well may be the one whose employment produces the best consequences overall.\(^\text{10}\) Thus we see again why it is unlikely that one in fact maximizes good consequences by deliberating about which of one’s action-options maximizes good consequences.

4. The consequentialist convinced of this will be tempted to go indirect, recommending that one deliberate in some particular nonconsequentialist fashion because one thereby *actually* maximizes good consequences.

But in fact this won’t do. The fact that one produces the best consequences by *employing* some particular decision procedure (again, call it DP*) does not entail that one will produce the best consequences by *trying to adopt* that decision procedure. For
the very same factors that make direct consequentialism self-defeating also cast doubt upon the wisdom of indirect consequentialism.

It is crucial to register the fact that the indirect consequentialist’s recommendation to adopt DP* is the upshot of a train of thought. The consequentialist arrives at the conclusion to adopt DP* on the grounds that the employment of DP* produces better consequences than does the employment of DPD. This train of thought takes some time to conduct. And more seriously, there is more to actually adopting a new decision procedure than merely arriving at the conclusion that employing it would be a good idea: it takes incredible energy and moxie to train oneself to deliberate differently. Only in the Bible does Saul become Paul in a day.

In other words, the indirect consequentialist does not instantaneously and effortlessly consider, choose, and implement a decision procedure. Instead he uses a procedure to decide how to decide, a procedure for choosing and implementing a decision procedure—a meta-decision procedure. Just as a decision procedure is a way of deciding how to act, a meta-decision procedure is a way of deciding what decision procedure to employ.

The indirect consequentialist’s meta-decision procedure is: Adopt the decision procedure whose employment in fact maximizes good consequences (MDPD). And because the indirect consequentialist thinks that those who employ DP* in fact maximize good consequences, MDPD recommends that one adopt DP*. But the fact (if it is) that those who maximize good consequences employ DP* does not entail that one maximizes good consequences by employing MDPD.
Much as with DPD, it is necessary to notice that MDPD is only one of very many possible meta-decision procedures one could employ—one might choose a decision procedure on all kinds of different grounds. A brief subset would also include:

- **MDPGR**: Employ the DP that you want others to employ.
- **MDPKant**: Employ the DP that expresses your rational nature.
- **MDPRR**: Employ the DP that revealed religion commands.
- **MPVel**: Employ the DP so that you choose autonomously\(^\text{11}\).
- **MDPG**: Employ the DP God personally instructs.
- **MDPHed**: Employ the DP that feels good.
- **MDPHWJD**: Employ the DP that Jesus would employ.
- **MDPTT**: Employ the DP your enemy employs.
- **MDPMin**: Employ the DP that takes the least time.

…

It would be easy (though not the best use of time) to continue adding to the list. There are a practically uncountable number of meta-decision procedures one could employ. Barring ties, employing only one of them enables one to produce consequences as good as one could produce. It is highly unlikely that MDPD is *that* decision procedure, for two by-now familiar kinds of reasons.

First, when the indirect consequentialist employs MDPD, he may fail to properly identify which decision procedure whose employment in fact maximizes good consequences. Just as the direct consequentialist is likely to fail to calculate which of her action-options really produces the best consequences, so too is the indirect
consequentialist likely to fail to calculate which of his decision procedure-options really produces the best consequences. By-now familiar terminology can describe this feature of meta-decision procedures. The decision procedure that a meta-decision procedure explicitly recommends is the one it deems choiceworthy. But those who employ that meta-decision procedure may in fact wind up employing a decision procedure other than the choiceworthy one. The decision procedure that the employer of the meta-decision procedure in fact ends up using is the resultant decision procedure. If the meta-decision procedure is especially simple, the resultant decision procedure will usually be the same as the choiceworthy decision procedure. But if the meta-decision procedure in question is difficult, then the resultant decision procedure will probably differ from the choiceworthy decision procedure.

Now it is likely that MDPD is a rather difficult one. It instructs its user to identify which decision procedure whose employment maximizes good consequences. The decision procedure whose employment in fact maximizes good consequences is choiceworthy. But the resultant decision procedure of employing MDPD is probably not the same as the choiceworthy decision procedure. And this implies that those employing MDPD will not in fact maximize good consequences.

But even if MDPD’s resultant decision procedure is identical to its choiceworthy decision procedure, there is a second problem vexing MDPD. As with decision procedures, the use of a meta-decision procedure can have autonomous effects, effects other than those associated with the decision procedure it issues in. The unexamined life may have its downside, but we should not overlook the costs of actually figuring out
which decision procedure maximizes good consequences either. Consider only the time it would take to determine just which decision procedure maximizes good consequences. It is no small task to ascertain exactly which decision procedure wins the competition. One wonders whether figuring it out would compensate for the loss of time (and other opportunities) expended conducting the inquiry. So if employing a meta-decision procedure turns out to be a lifelong task, an indirect consequentialist might complete and instantiate the result of his meta-decision procedure only when there is little time left to reap its fruits.

Another worrisome autonomous effect of MDPD is the same as with DPD: just as decision procedures structure the action-options one can choose among, so too do meta-decision procedures structure the decision procedure-options (and thus action-options) one can choose among. If those who employ DPD are excluded from the sort of social relationships that provide one with the best opportunities to produce good consequences, then there is some reason to think that those who employ MDPD will also be so excluded, although perhaps not as severely as straightforward direct consequentialists. Once others discover that a person decides to decide in whatever way happens to produce the best consequences, many of them are likely to mistrust her in the same way they mistrust the direct consequentialist. While a direct consequentialist will keep a promise only if it maximizes good consequence to do so, the employer of MDPD will keep a promise only as long as the decision procedure to keep promises continues to maximize good consequences. And so many people, worrying whether they can really count on that continuing, may prefer to co-operate with those on keep
promises on more stable grounds. So this means that the indirect consequentialist may not even have the best opportunities to act in ways that in fact produce the best consequences.

The indirect consequentialist’s meta-decision procedure, then, may suffer from all the same problems had by the direct consequentialist’s decision procedure. Its employment is unlikely to issue in the best resultant decision procedure, and it is likely to have significant negative autonomous effects. It is very likely that one produces the best consequences when one employs some other possible meta-decision procedure, rendering MDPD self-defeating. So the consequentialist cannot escape all the perils of direct consequentialism by going indirect.\textsuperscript{13}

The consequentialist may respond to these problems by concluding that if employing MDPD does not in fact maximize good consequences, the solution is to go even more indirect: one should adopt whatever meta-decision procedure whose employment in fact maximizes good consequences. There is no reason for the consequentialist to get saddled with an unhappy meta-decision procedure.

But it should be obvious why this move will not work. Figuring about which meta-decision procedure is the best to employ—and then implementing it in your own life—is also a species of practical thought. It is a meta-meta-decision procedure. More specifically, it is the meta-meta-decision procedure to choose the meta-decision procedure that maximizes good consequences (MMDPC). And all of the reasons for thinking that employing consequentialist decision procedures and meta-decision procedures will not in fact maximize good consequences can also be reasons for
thinking that employing meta-meta-consequentialist decision procedures are in the same boat. A nasty regress looms, so the consequentialist cannot safely escape the perils of self-defeat by continuing to go indirect.\textsuperscript{14} So indirect consequentialism evades self-defeat no better than direct consequentialism does.

5. The indirect consequentialist may hope to defend his theory by arguing that his is not the only theory suffering from the problems noted above. If he can show that any plausible rival theory also fails on its own terms—a partners-in-guilt strategy—then he can maintain that my argument above does not score any points against indirect consequentialism in particular.

Thus imagine that the indirect consequentialist launches the \textit{Satan objection}.\textsuperscript{15} Suppose that Satan will wreak havoc upon the world if you continue to employ your favoured decision procedure. Unspeakable horrors will visit every sentient creature if you decide to do things in the usual fashion. In fact, the only way that Satan will stay his hand is if you employ, say, DPD. Continuing to employ your usual decision procedure would have terrible autonomous effects.

What happens next depends upon what kind of meta-decision procedure you employ. Meta-decision procedures can be classified into two categories: first, those that instruct their user to choose a decision procedure at least \textit{in part} because of the good consequences of employing it; and, second, those that instruct their user to choose a decision procedure on nonconsequentialist grounds entirely. If you employ a meta-decision procedure of the second sort, the fact that continuing to employ your favourite
decision procedure has very terrible consequences will not alter how you deliberate. But if you employ a meta-decision procedure of the first sort, then you will surely do your best to decide in whatever way necessary to avert a hellish outcome.

So the fact that Satan will unleash Armageddon if you continue to use your usual decision procedure will sway you only if you employ a meta-decision procedure of the first sort. If you do, then you are concerned about the effects of how you decide to act, even if you are not devoted to the thought that such effects are the only thing that matters. But all the problems that trouble the indirect consequentialist also seem to threaten those who employ meta-decision procedures of the first sort as well, since in other more realistic circumstances, employing a meta-decision procedure that takes consequences into account will, like MDPD, likely yield a resultant decision procedure that differs from the choiceworthy decision procedure.

Thus it’s tempting to embrace a meta-decision procedure of the second sort, one for which the consequences of employing a decision procedure don’t matter at all. Then, the fact that employing such a meta-decision procedure can have terrible consequences produces no paradox. But the indirect consequentialist will remind us about what would happen if Satan were to threaten a person who uses a meta-decision procedure of the second sort: Satan unleashes Armageddon, an outcome most disastrous. The lesson the indirect consequentialist hopes we will take from this possibility is that it is just plain nuts to employ a meta-decision procedure of the second sort. Any acceptable meta-decision procedure will not instruct its user to select a decision procedure-option if doing so is disastrous. Meta-decision procedures of the
second sort may not give rise to any paradoxes, but they have other more serious marks against them. This divide-and-conquer strategy appears to enable the indirect consequentialist to shrug off any worries about the paradoxical situation that his own meta-decision procedure gives rise to.

But this is a false dilemma. One can employ a meta-decision procedure that is neither self-defeating nor shackles one to deciding in a way that has apocalyptically bad autonomous effects. For meta-decision procedures can have successfully implementable disaster clauses. For example, one’s meta-decision procedure could be:

MDPGRA  Employ the DP that you want others to employ, unless doing so has apocalyptically bad autonomous effects, in which case employ the closest DP that avoids such effects.

Employing this meta-decision procedure is not obviously self-defeating, for in all realistic scenarios one can successfully decide what to do in a way that one wants others also to decide, and one can successfully decide differently should one be confronted with a threat wielding demon. Friends of indirect consequentialism have rightly noted that the Satan objection reveals a flaw in employing decision procedures that take no account of their negative autonomous effects. But this does not by itself make it true that MDPD is the best meta-decision procedure, for it’s possible to employ a meta-decision procedure that takes at least some negative autonomous effects into account, but isn’t committed to maximization. And one can build successfully implementable disaster clauses all the way up—into meta-meta decision procedures, and so on—in order to cope with logically possible more sophisticated threat-wielding demons.
By contrast, the indirect consequentialist’s meta-decision procedure really is self-defeating. The source of self-defeat is the meta-decision procedure’s commitment to maximization, the very same element that is the source of self-defeat for the direct consequentialist. Given all the ways one can fail to decide best, it is no surprise that the indirect consequentialist’s meta-decision procedure should actually fall short of its goal. Maximization is a standard that is very difficult to meet, and this goes not only for selecting actions, but also for selecting decision procedures. Thus at least some indirect moral theorists not wedded to maximization should be able to recommend their favoured decision procedures without thereby tripping themselves up. But because the indirect consequentialist will most likely not maximize good consequences, his commitment to maximization means that his favoured meta-decision procedure has all the same problems as the direct consequentialist’s favoured decision procedure. Indirect consequentialism is not really more self-supporting than direct consequentialism after all.

There is also an especially effective ad hominem reply to the indirect consequentialist’s objection available. The consequentialist says that an action is right if and only if it produces better consequences than any alternative action. Now what determines whether some action is a genuine alternative?

Consider the fact that you are hearing this paper now. Does hearing this have better consequences than anything else you could be doing now? Well, it is logically possible for you to discover a cure for cancer, which is obviously better than hearing me. If your action-options are constrained only by what is logically possible, then
hearing me talk (and everything else you do) is not right, because there is indeed something else better you can do. Since this appears to be a reductio, the consequentialist should maintain that one’s options are not constrained by what is only logically possible. The consequentialist should work with a tighter notion of modality.

But if a logically possible alternative is not enough to overthrow the consequentialist’s conviction that some particular action is right, why worry about the logical possibility of a threat-wielding Satan? If the logical possibility is safely ignored in the first case, it is also safely ignored in these other cases as well. It would be a mistake to dismiss mere logical possibilities when evaluating action-options, but rely upon them when evaluating decision procedure-options. Consistency demands the same type of modality in both cases. So threat-wielding demons and the like, if they were to exist, should trouble us greatly. But their non-existence means we can safely and wisely forget about them.

1 Throughout this paper, I restrict my discussion of consequentialists to maximizers.


Indeed, it is usually impossible even to identify just what one’s options are. See Timothy Chappell, “Option Ranges”, *Journal of Applied Philosophy* 18 (2001), 107-118.


For an example, see Parfit’s case of the Transparent Motorist, *Reasons and Persons*, p. 7.


One might be tempted to think that the fact that, barring ties, only one decision procedure maximizes good consequences implies that only one decision procedure avoids self-defeat. But that inference would be mistaken. DPD is self-defeating because it specifically instructs one to choose the act that maximizes good consequences, while employing DPD itself most likely does not maximize good consequences. The fact that employing some other decision procedure also does not maximize good consequences is no knock against it, unless it too instructs one to maximize good consequences. For example, suppose
employing DPGR does not maximize good consequences either. DPGR is not thereby self-defeating, for the goal of DPGR is not to maximize good consequences.


12 Holly Smith has also noted that it is at least as difficult to determine which indirect decision procedures (or higher tier moral codes) maximize good consequences as to determine which actions maximize good consequences. See her “Two Tier Moral Codes”, *Foundations of Moral and Political Philosophy* (1990), eds. Paul, Miller, and Paul (Oxford: Basil Blackwell, 1990), pp. 112-132. Smith concludes that this is a knock against the *usability* of the indirect decision procedure. But this is a problem only for the theorist whose meta-decision procedure heavily prizes what she calls usability. Getting clear on the fact that choosing and implementing a decision procedure is itself a procedure of sorts (viz., a meta-decision procedure) that can have *various* aims is crucial for seeing why indirect consequentialism is no less *self-defeating* than direct consequentialism.

13 One odd result is that it is still possible for a consequentialist to reject DPD on grounds other than that employing DPD does not maximize consequences. In other words, if a consequentialist rejects DPD, not because he employs MDPD, but because he employs some other meta-decision procedure, my argument against indirect consequentialism here does not tell against him. I thank ________ for noting this oddity.

14 Note that the regress problem identified here is not the same supposed problem as the allegation that one always needs to employ a higher-level procedure in order to employ rationally a lower-level procedure. Smith 1991, among others, has successfully debunked this claim.


16 It’s actually stronger than just logically possible. See my (reference deleted for blind review).