10-1-2009

# Another Step in Divine Command Dialectics

Alexander R. Pruss

# ANOTHER STEP IN DIVINE COMMAND DIALECTICS

Alexander R. Pruss

Consider the following three-step dialectics. (1) Even if God (consistently) commanded torture of the innocent, it would still be wrong. Therefore Divine Command Metaethics (DCM) is false. (2) No: for it is impossible for God to command torture of the innocent. (3) Even if it is impossible, there is a non-trivially true *per impossibile* counterfactual that even if God (consistently) commanded torture of the innocent, it would still be wrong, and this counterfactual is incompatible with DCM. I shall argue that the last step of this dialectics is flawed because it would rule out every substantive metaethical theory.

## Introduction

I shall take Divine Command Metaethics (DCM) to be the view that what it is for an action to be obligatory or wrong is for it to be, respectively, commanded or forbidden by God. Nothing I shall say would, however, be affected if one replaced "commanded or forbidden" by, say, "willed or nilled".

I will begin by summarizing three steps in the dialectics of the criticism of DCM that can be put in the form of the question: "What if God commanded something horrible?" I will then add a fourth dialectic step where I will argue that this whole line of criticism of DCM is equally a criticism of just about every substantive meta-ethical theory.

While I am defending DCM against a particular objection in this paper, this does not imply that I accept DCM—in fact, I think DCM is false, but for other reasons.

## Step I: Against DCM

Throughout, I will use "torture of the innocent" as my example of a horrible action, one that is plainly and uncontroversially wrong no matter what. If one does not take this to be uncontroversial, one can use rape, genocide, etc. instead. The first step of the dialectics is now as follows. Consider the conditional:

(1) Even if God consistently commanded it, torture of the innocent would still be wrong.

To "consistently" command something is to command it without forbidding it. The "consistently" qualifier is needed to rule out uninteresting cases where the consequent of (1) holds because God both commands and forbids torture of the innocent. There is something deeply plausible about

(1)—torture of the innocent is necessarily *categorically* wrong, and in particular wrong no matter who commands it.

We now fill out the argument based on (1). There are some possible worlds where God does consistently command torture of the innocent. Now, on any plausible logic of subjunctives, if we have a true subjunctive conditional "were *p* to hold, *q* would hold," and *p* holds in at least one world, then there is at least one world where both *p* and *q* hold.[1] Therefore, if (1) holds, then there will be a possible world where God consistently commands torture of the innocent but torture of the innocent is still wrong. But this is incompatible with the DCM claim that the obligatory and wrong are defined as what God commands and forbids, respectively. Hence, DCM is false.

### Step II: A Defense of DCM

A standard response is that God's nature—perhaps in respect of omnibenevolence—is such that it is metaphysically impossible for God to command torture of the innocent. Thus, the argument in Step I fails, because there is no possible world where the antecedent of (1) holds.

### Step III: Against DCM, Again

In separate recent works, Wes Morriston and Erik Wielenberg[2] have defended a very natural suggestion, namely that the opponent of DCM can defend Step I by interpreting (1) as a *per impossibile* counterfactual, thus not begging the question against the defender of DCM who thinks it is metaphysically impossible for God to command horrible things. It will not matter much for my criticism of Step III how exactly this line of argument is developed, but for the sake of definitiveness, I will develop it as follows[3]:

(2)   Claim (1) is non-trivially true as a *per impossibile* counterfactual.

(3)   If (1) is non-trivially true, then DCM is false.

---

[1] If we use → for subjunctive conditionals, and ⇒ for entailment, the following pair of axioms are very plausible: (i) if $p \Rightarrow q$ and $p$ is possible, then $p \rightarrow q$; and (ii) if $p \rightarrow q$ and $p \rightarrow \sim q$, then $p$ is impossible. Given these two axioms, it follows that if $p \rightarrow q$ holds and $p$ is possible, then we must have $\sim(p \rightarrow \sim q)$, since if we had $p \Rightarrow \sim q$, we would also have $p \rightarrow \sim q$ by (i), thereby contradicting (ii). But if we have $\sim(p \Rightarrow \sim q)$, then there must be a possible world where both $p$ and $q$ hold.

[2] Wes Morrison, "What if God Commanded Something Terrible?" (paper presented at the University of Texas at San Antonio Philosophy Symposium, San Antonio, Texas, 2008). http://colfa.utsa.edu/ecpc/symposium_papers/what-if-God.pdf. Erik Wielenberg, *Virtue and Value in a Godless Universe* (Cambridge: Cambridge University Press, 2005), pp. 41–43, 48–49.

[3] My development differs slightly from Morriston's (ibid.) in that Morriston's *per impossibile* counterfactual does not have a consistency requirement in the antecedent in the way (1) does. I take (1) to be a friendly amendment of what Morriston considers, but in any case my defense of DCT would work just as well if I used Morriston's version.

Of course, it follows from (2) and (3) that DCM is false. (Note that the conditional in (3) is an ordinary material conditional.)

The reason for my putting "non-trivially true" instead of just "true" is that on typical possible worlds semantics for counterfactuals, if the antecedent of a counterfactual is necessarily false, as Step II claims in the case of (1), then the counterfactual is automatically true. Thus, (2) requires either a denial of possible worlds semantics for counterfactuals, or else one needs a distinction between a *per impossibile* counterfactual that is merely *trivially* true (i.e., true simply because the antecedent is necessarily false), such as

(4)  If some object were blue all over and green all over, then a circle would be square

and one that is non-trivially true (i.e., true not simply because the antecedent is necessarily false), such as

(5)  If some object were blue all over and green all over, then some object would be blue all over.

The idea behind (2) is that torture of the innocent (or whatever evil you want to put in its place) is so horrible that if, *per impossibile*, God were to consistently command it, the right thing to do would be to disobey.

On the other hand, the idea behind (3) is something like this. According to the DCM, what defines wrongness is divine prohibition and what defines permissibility is lack of divine prohibition. But then if something is consistently commanded by God, by definition it is permissible given DCM. Hence, the argument goes, if DCM is true, we should either say that (1) is a false *per impossibile* counterfactual, or that if it is true, it is true for no further reason beyond the fact that all *per impossible* counterfactuals are at least trivially true. Therefore, if (1) is true and not merely trivially true, DCM is false.

*Step IV: The DCM defended, Again*

I shall now argue that there is something wrong about the argument in Step III, by noting that the argument proves too much—it equally applies against every substantive metaethical theory. I shall say a metaethical theory is substantive provided it gives us a non-circular definition of terms like "obligatory" and "wrong", and does not do so by giving an infinite list of all the obligatory and wrong actions. There are thus two kinds of non-substantive theory: one is a refusal to define terms like "obligatory" and "wrong", and the other defines "obligatory" and "wrong" by infinite disjunctions. I will begin by considering non-consequentialist theories.

For concreteness, take metaethical Kantianism as our example of a substantive non-consequentialist metaethical theory. According to metaethical Kantianism, something is obligatory provided that reason requires it and forbidden provided that it is contrary to reason. Say that reason *consistently* requires $A$ provided reason requires $A$ and $A$ is not contrary to reason (if one thinks, as Kant does, that what reason requires cannot be contrary to reason, then reason requires $A$ if and only if reason consistently requires $A$). Now, exactly parallel to (1) we have the following conditional:

(1K) Even if it were categorically required by reason, torture of the in-
nocent would still be wrong.

This seems plausible for exactly the same reasons for which (1) is plausi-
ble—namely, the horribleness of torturing the innocent. We can now re-
prise Steps I–III. First, one might take (1K) to refute Kantianism directly.
Then one might respond that the antecedent of (1K) is impossible—torture
of the innocent cannot be required by reason, since it violates the Categori-
cal Imperatives. Finally, one can run Step III, and affirm:

(2K)  Claim (1K) is non-trivially true as a *per impossibile* counterfactual.

(3K)  If Kantianism is true, then (1K) is not non-trivially true.

Again, these claims seem to have exactly the same plausibility as (2) and
(3). Torture of the innocent is so horribly wrong that even if, *per impossibile*,
it were consistently required by reason, it would still be wrong—the right
thing to do in such a case would be to go against reason.

Exactly the same thing can be said about other substantive meta-ethical
theories. Exactly parallel to (1), we have claims like:

(1K2) Even if it were consistently required by both categorical impera-
tives, torture of the innocent would still be wrong.

(1R)  Even if free negotiation under a veil of ignorance were to settle on
a system where the innocent are tortured, torture of the innocent
would still be wrong.

(1L)  Even if it were a loving thing to do, torture of the innocent would
still be wrong.

And, again, we can come up with parallels to (2) and (3).

One might be a bit worried about (1L), because the conceptual connec-
tion between something between torture and an action being unloving
might be felt to be too close, and so one might deny (1L). But I think one
can still come up with parallels to (1L) if one replaces "torture of the in-
nocent" by something else that is clearly always wrong, like "torture of
the innocent in order to save many lives" or "rape".

There is thus surely something wrong with Step III, since it would re-
fute every non-consequentialist substantive meta-ethical theory. Note also
that Step I *by itself* refutes typical consequentialist meta-ethical theories,
since the following is true:

(1C)  Even if it were to have the best consequences, torture of the inno-
cent would still be wrong.

The typical consequentialist is, after all, unable to avail herself of the Step
II response that the antecedent of (1C) is impossible. There may be an
atypical consequentialist who can say that the antecedent of (1C) is impos-
sible (e.g., one who thinks that, necessarily, due to the goodness of God,
situations where horrible actions produce the best consequences are im-
possible), but then the consequentialist is subject to the Step III criticism.
Hence, collectively, Steps I and III apply against every substantive meta-
ethical theory, consequentialist or not. That is, surely, proving too much.

At the same time, as Morriston has himself noted, there is a way in which consequentialists are somewhat better off here than divine command metaethicists. For the consequentialist can say that even if it were to have the best consequences, torture of the innocent would still be the doing of something that is *prima facie* wrong. But the defender of DCM is apt to have a harder time saying that if God consistently commanded torture, the torture would still be *prima facie* wrong.[4]

One might also object that not all of (1K), (1K2), (1R) and (1L) are on par with (1). Each of the meta-ethical views implies a conditional of the form:

(6)   If $F(A)$, then $A$ is obligatory.

Now, we might distinguish two kinds of meta-ethical views. On some meta-ethical views, (6) is claimed to be conceptually necessary.[5] For instance, the connection between being categorically required by reason and being obligatory will be conceptually necessary if Kant is right. And it may be that other meta-ethical views merely claim that (6) is metaphysically necessary. I am myself sceptical of the distinction between conceptual and metaphysical necessity, but let me play along. The idea, then, is that the general argument form being examined works better against one class of meta-ethical views and more poorly against the other class. Perhaps it works more poorly against the ones where the criterion $F(A)$ is connected with the obligatoriness of $A$ of conceptual necessity, because in the case of such a view, we may be no more troubled by the claim that if, *per impossibile*, we were to have $F$(torture of the innocent), then this torture would be obligatory than we are troubled by the tautologous claim that if, *per impossibile*, torture of the innocent were obligatory, then this torture would be obligatory.

However, I actually think each of (1K), (1K2), (1R) and (1L) is still troubling even if each is taken to be conceptually necessary. But in any case, the defender of DCM can similarly claim that there is a conceptually necessary connection between divine command and moral obligation.

At the same time, if the meta-ethical views where the connection in (6) is merely metaphysically necessary are seen as less troubling vis-à-vis this argument, perhaps because in those cases one might actually with some plausibility either deny the conditional in (1K), (1K2), (1R) or (1L), or claim that it is merely trivially true, then the defender of DCM can likewise adopt the metaphysical necessity version of DCM.

Another potential difference between the views, one that we have already alluded to in the case of the love view in (1L), is that on some of the views, the antecedent of the *per impossibile* conditional (1K), (1K2), (1R) or (1L) might be taken to be *conceptually* impossible. And this could be taken as a point in favor of these theories. For we cannot do *more* justice to the strength of our intuitions that torture of the innocent is wrong than by saying that it is a conceptual truth that torture of the innocent is wrong. For this to count as an objection to DCM, it would have to be the case that it is not conceptually impossible for God to command torture of the innocent,

---

but it is, say, conceptually impossible for torture of the innocent to be a loving thing.

This in turn forces a distinction between different DCM theories. According to DCM, to be obligatory is to be commanded by God. Now, we can understand "God" as a name devoid of descriptive content or as either a definite description or a name that has descriptive content ("United States of America" could be such a name—it might be a conceptual truth that the United States of America contains at least one state). If we understand it as a proper name, then it seems to be conceptually possible for God to command torture of the innocent. Hence, we have reason to prefer those meta-ethical theories on which it is a conceptual truth that torture of the innocent is wrong to a DCM on which "God" is taken to be a proper name.

But a DCM where "God" is taken to be a proper name seems implausible. It would force us to accept as non-trivially true other strange *per impossibile* counterfactuals, such as:

(7)  Were God a finite being who did not create us, and who had limited knowledge and not much goodness, what he commanded would still be obligatory.

It would also be strange if one could define obligation in terms of something—e.g., the fact that a being $x$ commanded the action—which does not carry anything of great moral significance on its sleeve.

A more plausible DCM will take "God" to be a definite description or a name with descriptive content. But for the DCM to be plausible, this description or descriptive content may well include the fact that the being is all-knowing, all-powerful, perfectly reasonable, perfectly non-malevolent, and perfectly loving. And if the description is such, then it will be at least as plausible that it is conceptually necessary that this being does not command torture of the innocent as it is that it is conceptually necessary that torture of the innocent is unloving, unreasonable, contrary to the categorical imperatives, or rejected in negotiations in the original positions. And if we think it is not self-contradictory that such a being could command torture of the innocent (e.g., because we can imagine some radical circumstances), then for the same reason (e.g., in the same circumstances) we could probably imagine that torture of the innocent might be loving, reasonable, in accord with the categorical imperatives, and agreed-on in the original position.

## Conclusions and Speculations

The line of argument started by considering horrible actions which would still be wrong even if God consistently commanded them, then, seems to apply against every substantive meta-ethical theory. This by itself shows that there is something fishy with this line of argument, but it is unsatisfactory since it does not show *what* is fishy about it. And, of course, there is always the possibility that someone will embrace this line of reasoning as an argument against the very possibility of a substantive meta-ethics, along the lines of Moore's open question argument.

I want to end with several speculations on what might have gone wrong in Step III. The first is that we reason poorly about outlandish

counterfactuals like (1) and its analogues, and *very* poorly about *per impossibile* counterfactuals.

The relevant problem seems to me to be that when we are very strongly sure that something is true and its truth is very important to us, we have a tendency to carry it over into counterfactual situations, even when doing so is inappropriate, as it may be in (1), (1K), (1K2), etc. Consider the joke[6]: "I am glad I don't like broccoli, because if I did, I would eat it, and that would be terrible." One can almost imagine this being seriously said, the speaker's hatred of broccoli being so vivid that it is carried into the counterfactual situation even when it is expressly denied there. Indeed, phobias provide real-life examples of this fallacious reasoning. I am afraid of dogs, and I don't want to lose my fear of dogs, because were I to lose my fear of dogs, I would no longer avoid them.

In fact, the defender of DCM can give an explanation of why we have a tendency to inappropriately carry over the prohibition on, say, torturing the innocent into the consequent of counterfactuals like (1), (1K), (1K2), etc.: God has implanted a belief in this prohibition in us so deeply that we find it very difficult to think ourselves into considering the impossible worlds where it doesn't hold. Alternately, perhaps, God has implanted in us a tendency to carry the prohibition into the consequent of ordinary counterfactuals in general, because we should see the torturing of the innocent as wrong *no matter what* (it's wrong even if it saves hundreds of lives, etc.), and then we wrongly continue to carry the prohibition into the consequent of extraordinary counterfactuals like (1), (1K) or (1K2).

Second, one might simply question the idea of non-trivial truth of *per impossibile* counterfactuals altogether. They are, after all, hard to make sense of, though valiant attempts have been made.[7]

Third, following a suggestion by C. Stephen Evans (in correspondence), one might note that what allows (1) to have plausibility as a non-trivial *per impossibile* counterfactual is our ignorance of the nature of God. If we really knew God's perfectly good nature, we would not take (1) seriously. In fact, we can compare (1) to counterfactuals like this:

(8)  Were a diamond to have the molecular structure of $H_2O$, it still wouldn't be water.

Surely (8) is no challenge to the theory that to be water is to be $H_2O$. Why not? Well, the defender of the theory that water is $H_2O$ has several options. She can just say that (8) is merely trivially true—just as the defender of DCM, the Kantian or the contractarian can respectively say that (1), (1K) or (1R) is only trivially true. Or she can come up with a theory of counterfactuals on which (8) is non-trivially true, because we get to assert *per impossibile* counterfactuals where the consequent is a necessary truth that is central to our conceptual scheme (such as that torturing the innocent is wrong or that diamond is not water). But on such a theory of counterfactuals the non-trivial truth of (8) is no challenge to the identity of water

---

[6]I got it from Richard Gale who heard it said by a comedian whose name I don't remember.

[7]E.g., Nicholas Rescher and Robert Brandom, *The Logic of Inconsistency* (Oxford: Blackwell, 1980).

and $H_2O$, just as the non-trivial truth of (1), (1K) or (1R) is no challenge to DCM, Kantianism and contractarianism.

Let me end on a personal note. I actually think DCM is false, and through what seems to be a kind of intellectual *akrasia*, I still find myself *feeling* that (1) provides evidence for the claim that DCM is false, in a way in which I do not feel that, say, (1K) or (1K2) provides evidence for the claim that Kantianism is false (though to be fair to myself, I think I have other reasons to oppose DCM). Perhaps the right diagnosis of this akratic feeling is that it is easier to imagine the antecedent of (1) being true than it is to imagine the categorical imperatives consistently requiring torture of the innocent. But if that is what is going on here, then the feeling may be due to the fact that the necessity of God's benevolence, and hence the necessity of the antecedent of (1) being false, is simply hard for us earthly beings to see, since we do not have a grasp of the divine essence, and so I may mistakenly feel that God's commanding something horrible is not *utterly* impossible.[8]

*Baylor University*

---