

Sorensen on Unknowable Obligations

THEODORE SIDER

Utilitas 7 (1995): 273–9

1. Access principles

Vagueness in the phrase “can know” aside, the principle of

Access An act is obligatory only if its agent can know that it is obligatory

is an important principle, worthy of serious scrutiny. Its truth or falsity bears on the question of whether moral rightness, obligatoriness, etc., are a matter of factors “internal” to an agent (such as motives and beliefs), or whether “external” factors (such as consequences) are relevant to determining the moral normative status of acts. Moreover, Access enjoys considerable intuitive support. If I destroy Greensboro in professor Sorensen’s example by pushing the wrong button, I seem to have a good *excuse* to give to anyone who would accuse me of wrongdoing: “I had no way of knowing that this action would be wrong!” But if I have a good excuse, then pushing the button doesn’t seem wrong; and if that action isn’t wrong, then I had no obligation to refrain from pushing that button.

A principle distinct from, but analogous to, Access also deserves our attention:

Access for blameworthiness An act is blameworthy only if its agent could have known that it would be blameworthy

Many philosophers who reject Access proper may be at least initially inclined to accept Access for blameworthiness. Such philosophers might respond to the destroying Greensboro example as follows: the fact that I have an excuse in saying “I couldn’t have known” doesn’t show that my action wasn’t wrong; all it shows is that I don’t merit blame for pushing that button. The availability of such a response is crucial to those who deny Access (otherwise they have no way to explain our intuition that “I couldn’t have known” does in *some* sense excuse); since this response assumes the Access principle for blameworthiness, the Access principle for blameworthiness is an important principle as well.

I think something like the Access principle for blameworthiness is true.¹ I'd like to remain more agnostic about the Access principle itself, because I'm unsure whether wrongness entails blameworthiness. Nevertheless, in these comments I propose to defend Access. Many of Sorensen's arguments against Access seem to work just as well against the Access principle for blameworthiness, which I *do* want to defend; thus, by defending the former principle explicitly, I hope to defend the latter implicitly.

2. The Holism Argument

I'll only have the space to attempt to reply to a few of Sorensen's many ingenious arguments. The first derives from a worry that the Access principle would imply that too many descriptive facts are knowable. Sorensen puts the crucial part of the argument thus:

...if my obligations must be knowable, any necessary background facts must also be knowable.

The general worry is that the Access principle would not be neatly confined to the moral realm, but would have unwanted consequences for what descriptive facts moral agents could know. He illustrates an instance of this worry as follows:

God's existence would be a crucial background fact for anyone trying to ascertain their obligations. So by the access principle, God's existence would be knowable to every human being.

On at least one natural interpretation of the phrase "background fact", the claim that God's existence is a background fact for every obligation is highly dubious. Suppose we take the background facts of a given obligation to be those facts that determine or ground the obligation, according to the true moral theory, whatever that theory is. Facts about an act's consequences are its background facts, if consequentialism is true; the fact that I've made a certain promise is a background fact for various obligations if certain deontological

¹ Perhaps "could have known" should be replaced with "could have had good reason to believe"; perhaps "could have known" should be understood as meaning "could have known at some time" (see the final paragraph); perhaps "could have known" should be replaced by "is not culpably ignorant that"; I leave (some of) these matters undecided.

theories are true. On this construal, the soundness of the argument depends on the truth of some sort of Divine Command Theory (or at least a theory in which the existence of God plays some important role in determining one's obligations).² The defender of Access need not accept such a theory.

But even if we don't accept this part of Sorensen's argument, we're still left with the original worry that the epistemological access guaranteed by the Access principle would extend too far. What if the true moral theory says that the normative status of an action depends on the action's consequences in the distant future, or on the inner mental states of other persons, or on other facts that are sometimes inaccessible to moral agents? In some cases my only way of knowing that I have a certain obligation would seem to be to know the relevant background fact, and so if one of these moral theories is true, Access would rule out the existence of any such obligations.³ But surely a defender of Access has little to fear here. Anyone inclined to accept Access will reject moral theories that make the moral status of an act dependent on such inaccessible facts. It is more likely that the defender of Access will accept some theory according to which the moral status of an action is determined in some way by factors "internal to" the agent (e.g., motives, beliefs, etc.).

Another way to interpret the Holism argument depends on a broader construal of "background fact", on which any fact that must hold in order for a given obligation to exist is a background fact of that obligation. The idea behind the Holism argument would then be that since an obligation presupposes its background facts, I'd need to be able to know those facts in order to be able to know I have the obligation. Consider, for example, the proposition that

(P) the Divine Command Theory is true, and God has commanded that I
not do *A*

where *A* is any given obligation I have. (P) is inconsistent with my being obliged

² Actually, the fact that God exists wouldn't itself be a background fact, if the Divine Command Theory were true. Rather, certain facts that entail that God exists would be background facts—facts such as the fact that God has commanded such and such an action.

³ As Fred Feldman pointed out to me, it would be a mistake to suppose that in *every* case, I could know of the existence of an obligation only by knowing the relevant background facts. I could know inductively that acts of a certain type are obligatory, without knowing what facts about, e.g., utilities, grounded those obligations. Or I could be told by a moral authority I knew to be reliable that a certain act is obligatory. But in *some* cases I'd need to be able to know the background facts. In a case like Sorensen's button pushing example, there is no other source for the knowledge that pushing the correct button is obligatory.

to do *A*, and therefore (P)'s negation is a background fact of *A*. Sorensen might then argue as follows:

- i) I have obligation *A*
- ii) if I have obligation *A* and Access is true, then I can know that I have obligation *A*
- iii) if I can know that I have obligation *A* then I can know that (P) is false (since my having *A* is inconsistent with (P))
- iv) I can't know that (P) is false
- v) Therefore, Access is false

Thus, the Access principle seems to grant too much potential knowledge to anyone with obligations. Surely, there are *some* conceivable circumstances in which I could have an obligation but not be able to know anything about God and the Divine Command Theory.

The problem with this argument is that it bears a striking resemblance to certain arguments for skepticism. Premises iii) and iv) say that since my having obligation *A* is inconsistent with (P), I must know that (P) is false in order to know that I have obligation *A*; since I can't know that (P) is false I can't know that I have obligation *A*. Compare the familiar "My being a brain in a vat in Rochester is inconsistent with my reading my comments in Greensboro, so I can't know I'm in reading my comments in Greensboro unless I can know that I'm not a brain in a vat in Rochester. But I can't know the latter, so I can't know the former." I don't know what is wrong with these skeptical arguments, but I take it that something is wrong with them since I do know that I'm here in Greensboro. Of course, there may be a crucial disanalogy between the skeptical argument and Sorensen's Holism argument. Without a specific diagnosis of what goes wrong in the skeptical argument, it is hard to be sure. So perhaps the best thing to say is that if someone wants to defend the current version of the Holism argument, he or she has the burden of producing a disanalogy between the two arguments (or, alternatively, of defending the skeptical argument).

3. The Indeterminacy argument

It seems clear that there could be cases in which it would be indeterminate whether a given person would be obligated to do a certain thing. For de-

ontologists this might arise from a person making a vague promise; for a consequentialist it could arise from indeterminacy in whether one consequence was better than another. If wrong or blameworthy acts are wrong or blameworthy in virtue of the motives and beliefs of the agent, then indeterminacy could arise both from indeterminacy in just what those motives and beliefs are, and also from indeterminate standards of what counts as an acceptable set of motives and beliefs. So the consequence that it could never be a vague or indeterminate matter whether a given act is obligatory, or blameworthy, would be unwelcome.

Sorensen's argument that Access has just this consequence runs as follows:

All obligations are knowable obligations

No indeterminate obligation is a knowable obligation

Therefore, no indeterminate obligation is an obligation

The first premise appears to be just the Access principle, so if the second premise is necessarily true, then Access seems to imply the conclusion, which appears to be the claim that obligation is never indeterminate. As near as I can tell, the corresponding argument for blameworthiness has equal force.

Trouble begins when we begin to inquire carefully into the nature of the conclusion. It prohibits the existence of an act that is both i) obligatory, and ii) an "indeterminate obligation". Thus, it seems to be the claim that no act that is obligatory can fail to be definitely obligatory. The problem is that this isn't really objectionable, since this is consistent with the existence of an act such that it is indeterminate *whether or not* that act is obligatory—i.e., an act that is neither definitely obligatory nor definitely not obligatory. Thus, contrary to initial appearances, the conclusion doesn't prohibit indeterminacy in whether a given act is obligatory; it merely prohibits indeterminacy in cases where an act *is* obligatory.

Sorensen formulates the argument in a second way:

1. All obligations are knowable. (Assume)
2. It is indeterminate whether I am obliged to vote. (Assume)
3. I cannot know that I am obliged to vote and I cannot know that I am not obliged to vote. (From 2)
4. I cannot know that I am obliged to vote. (3, Simplification)

5. I am not obliged to vote. (1, 4, TF)

If premises like 1 and 2 imply conclusions like 5, Access would indeed rule out indeterminacy of obligation. And the problem with the first argument is absent here, since premise 2 states that it is indeterminate *whether* I'm obliged to vote. But there's a new problem: 3 doesn't follow from 2. The justification for this inference presumably involves the fact that knowledge entails truth: if it is indeterminate whether I'm obliged to vote, then it isn't true that I'm obliged to vote; but I can only know truths; therefore I can't know that I'm obliged to vote. The problem here is that if it is indeterminate whether I'm obliged to vote, it doesn't follow that it is *not* true that I'm obliged to vote, only that it isn't *definitely* true that I'm obliged to vote. If it is indeterminate whether I'm obliged to vote, then it is presumably indeterminate whether it is true that I'm obliged to vote. It may follow from this that *it is not definitely the case that* I can know that I'm obliged to vote, but Sorensen's argument requires the stronger claim that I can't know that I'm obliged to vote.

4. Dumbing down Ethics

Suppose I've inherited money, but I'm worried that my parents' will requires that I donate some fraction of that money to charity. I could burn the will without looking at it, thereby removing all possibility of learning that the will calls for the charitable donation. Let's grant for the sake of argument that the will would be the only source of an obligation to donate the money; thus, it seems that it is within my power to bring it about that I won't be able to know that I'm obligated to donate the money. If Access is true, then it is within my power to bring it about that I'm not obligated to donate the money. Similarly, if Access for blameworthiness is true, then it is within my power to bring it about that failing to donate the money isn't blameworthy. Are these results objectionable?

I think not. Consider carefully what the consequence of Access is:

(c1) *After* burning the will, I have no obligation to donate the money

Think of being in my situation after burning the will. I have no way, then, of knowing whether the will requires me to donate the money. In that situation, *given that the will is gone*, refraining from donating the money doesn't seem

wrong. (Recall the stipulation that there's no source other than the will of an obligation to donate.) What I think *are* objectionable are the following claims:

- (c2) *Before* burning the will, I have no obligation to donate the money
- (c3) I have no obligation to refrain from burning the will

But neither of these is entailed by Access; and the corresponding claims for blameworthiness aren't entailed by Access for blameworthiness. Before I burn the will, it is within my power to know that I have an obligation to donate the will, since it is then within my power not to burn the will. Thus, Access doesn't entail (c2). And since I can know that I'm obligated to refrain from burning the will—at least there's no reason to suppose otherwise—Access doesn't entail (c3).

Sorensen replies to a response something like mine by offering an infinite regress argument. Remember that I'm accepting the consequence of Access that, after the will is gone, I have no obligation to donate the money—my response consists of claiming that intuition is served by the admission that, among other things, I *did* have an obligation not to burn the will. As applied to my version of the defense, the idea behind Sorensen's regress argument is that I could, as he puts it, “keep ignorant” of my obligation not to burn the will, thus alleviating the obligation not to burn the will. Any attempt to respond as before requires a higher order obligation not to keep ignorant in this way; thus, the regress begins.

It is important to be clear that at each stage in the infinite regress, a *new example* appears. In the original example, I do not in fact render it impossible that I know I am obliged not to burn the will; thus I can answer the objection to Access by pointing out that, in that example, (c3) is false—burning the will would be wrong. A new example, one in which I do remove the possibility of knowing that I shouldn't burn the will, can perhaps be constructed. But this new example may then be answered by granting that after I bring about this ignorance, then burning the will wouldn't be wrong, but pointing out that I have an obligation not to bring about this ignorance. If a further case is constructed in which I evade this obligation, this case, just like the earlier cases, may be answered by pointing out that the evasion would be wrong. At best, the regress consists of an infinite sequence of cases, none of which refutes Access. Sorensen responds that violations of esoteric obligations such as would appear at the tenth stage of this regress don't seem serious enough to satisfy

our intuition that such a moral shirker has gone seriously wrong. One problem with this response is that it is far from clear that the regress even gets that far: is there a possible case in which I make it impossible that I know that I ought not to make it impossible to know that I...⟨repeat 10 times⟩...that I know that I ought to read the will? I don't have any idea what such a case would look like. Secondly, if such a case were to be constructed, remember that a defender of Access can deny claims like (c3): *before* the self-induced ignorance, I do have an obligation to refrain from burning the will, and even an obligation to donate the money, because at that time it is in my power to learn about my various obligations.

Let's return to the original case in which all I do to avoid my obligations is to burn the will. I have been defending the consequence of Access that after doing so, failing to donate the money is not wrong; but certain modifications to Access are possible that alleviate the need for such a defense.⁴ For example, the Access principle could be weakened to require of obligations merely that at some time or other they be knowable. Before burning the will, it was open to me to know what the will said, and so this modified principle does not allow me to avoid obligations by burning the will. But I would caution against this way out. Consider an alternate case, in which I could look at the will today, but non-culpably decide to wait until tomorrow. Unfortunately, the will is accidentally destroyed before I look. The Access principle should be formulated to allow my not donating the money in this case, despite the fact that at some time it was possible for me to know that I had an obligation to donate the money. The original Access principle, which has this consequence, therefore seems preferable to the modified principle.⁵

⁴ I thank Walter Sinnott-Armstrong for a helpful discussion of these matters.

⁵ Thanks are due to David Braun, Fred Feldman, Rich Feldman, Walter Sinnott-Armstrong, and Roy Sorensen for their helpful comments.