

Teleology, Agent-Relative Value, and ‘Good’*

Mark Schroeder

I. TELEOLOGY AND AGENT-RELATIVE VALUE

A. Introduction

It is now generally understood that constraints play an important role in commonsense moral thinking and generally accepted that they cannot be accommodated by ordinary, traditional consequentialism. Some have seen this as the most conclusive evidence that consequentialism is hopelessly wrong,¹ while others have seen it as the most conclusive evidence that moral common sense is hopelessly paradoxical.² Fortunately, or so it is widely thought, in the last twenty-five years a new research program, that of *Agent-Relative Teleology*, has come to the rescue on all sides. While consequentialism says that every agent ought always to do that action that will bring about the most good, according to Agent-Relative Teleology,

ART: For all agents x , x ought always to do that action that will bring about the most of what is good-relative-to x .³

ART is supposed to allow us to have our cake and eat it as well. It is supposed to both accommodate constraints and retain whatever is attractive about consequentialism—in particular, to avoid the putative

* Special thanks to Doug Portmore, Jeff McMahan, Donald Regan, Tom Hurka, Michael Smith, Paul Pietroski, Brett Sherman, Sarah Stroud, Scott Soames, Jeff King, Jay Wallace, Amy Challen, Sergio Tenenbaum, and Arthur Ripstein, as well as to audiences at the University of Southern California and at the University of Toronto.

1. Bernard Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973).

2. Especially Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1983).

3. I hyphenate “good-relative-to” to make clear that it is a technical term that I use stipulatively to pick out the relation that is appealed to by Agent-Relative Teleology. It is a substantive question that will be of importance for this article whether this relation, the *good-relative-to* relation, has anything to do with ‘good’.

Ethics 117 (January 2007): 265–295

© 2007 by The University of Chicago. All rights reserved. 0014-1704/2007/11702-0018\$10.00

“paradox” of deontology. Agent-Relative Teleologists (henceforth, ART-ists), and nearly every writer who has ever commented on ART in the literature, are agreed both that ART can accommodate constraints and that it can do so in a way that should be appealing for the same reasons that consequentialism is appealing.⁴ Several authors have even held that it is such a compelling moral theory that charity requires the hypothesis that everyone believes it.⁵

But this, I’ll be arguing in this article, is completely wrong. It is true, of course, and not controversial, that ART has the right formal structure to accommodate constraints and special obligations. And since ART sounds a lot like consequentialism, ART-ists and others have taken completely for granted that ART must have the attractive features of consequentialism, even if it may turn out to have other costs. But this, I’ll explain, cannot be taken for granted. A great deal of philosophical spadework will be required for ART-ists to earn the right to suppose that their view can acquire the attractions of consequentialism. They must tell us, in effect, what the *good-relative-to* relation has to do with the word ‘good’.

B. Constraints

Constraints are supposed to pose structural objections to ordinary consequentialism—the view that every agent ought always to do that action that will result in the state of affairs that is best. It is an old objection to consequentialist views that one ought not to murder even in order to prevent two deaths. But this objection turns on taking for granted the consequentialist’s *axiology*, or theory of the good. It turns on assuming that murders are no worse than other deaths. But consequentialists with pluralist axiologies can easily accommodate this case by supposing that murders *are* worse than ordinary deaths—and in particular, that one murder is worse than two deaths. Whether or not this is a plausible view about how bad murders and deaths are, it is a move that shows that consequentialism as such is not directly threatened by such cases.

A better objection to consequentialism is that it may turn out that Franz ought not to murder, even in order to prevent two *murders*. Intuitively, it is at least possible that this might be true. Even if Franz

4. One notable exception is Donald Regan, “Against Agent-Relativity: A Reply to Sen,” *Philosophy & Public Affairs* 12 (1983): 93–112.

5. This is known as *Dreier’s Conjecture*, after James Dreier, “The Structure of Normative Theories,” *Monist* 76 (1993): 22–40. Others accepting this view include Michael Smith, “Neutral and Relative Value after Moore,” *Ethics* 113 (2003): 576–98; Jennie Louise, “Relativity of Value and the Consequentialist Umbrella,” *Philosophical Quarterly* 54 (2004): 518–36; and Campbell Brown, “Consequentialise This” (unpublished manuscript, University of Edinburgh, 2004), but at least one prominent ART-ist demurs (Douglas Portmore, “Consequentializing Moral Theories,” forthcoming in *Pacific Philosophical Quarterly*).

should murder if it would prevent, say, 1 million murders,⁶ it is highly plausible that he ought not to murder in order to prevent two. Or at least, even if this is not true of any actual Franz, it at least seems like a possible scenario. But since the case is constructed with murders on both sides, we can't solve anything by supposing that murders are worse than murders.

Of course, the consequentialist could suppose that Franz's murders really are intrinsically worse than the two murders that he prevents—those of Hans and Jens, say. But this would also be a bad move. For then it would turn out that if Hans could prevent Franz from murdering by committing a murder, then he ought to do so. Indeed, it would turn out that Hans and Jens ought *both* to murder, if that would prevent Franz from murdering. But commonsense intuitions track nothing special about Franz in particular—I pulled him out of a hat. If Franz is under a constraint not to murder even in order to prevent Hans and Jens from murdering, Hans and Jens are likewise under constraints not to murder, even in order to prevent each other and Franz from murdering. Since the situation is symmetric, it won't work to import asymmetries into the consequentialist's theory of value.

The ART-ists' reaction to constraints is to note that what is needed for a consequentialist explanation of Franz's constraint not to murder is the assumption that Franz's murders are worse than Hans's and Jens's put together, that what is needed for a consequentialist explanation of Hans's constraint is to suppose that Hans's murders are worse than Franz's and Jens's put together, and that what is needed for a consequentialist explanation of Jens's constraint is to suppose that Jens's murders are worse than Franz's and Hans's put together. The unfortunate thing is simply that these three assumptions are not jointly consistent. So what ART-ists propose is to grant all three of these assumptions, but to deny that they are inconsistent. They do this by “relativizing” *better than* to agents. So the first assumption is true when understood as relative-to Franz, the second is true when understood as relative-to Hans, and the third is true when understood as relative-to Jens. In this way, ART-ists propose to account for constraints and special obligations, while retaining the advantages of consequentialism.

C. Relativizing

A first observation: note that “relativizing” is a perfectly legitimate formal move. Ordinary consequentialism has the following structure—it invokes an ordering on possible states of affairs, which it uses to induce an ordering on actions available to an agent at a time in the following way: the rank of the action is given by the rank of the resulting state of

6. That is, even if there are no *absolute* side-constraints.

affairs.⁷ This view has the wrong structure to account for constraints, because there is no single ranking on possible states of affairs that will induce the right rankings on actions for different agents. So the ART-istic move is to have a view with a different structure—it invokes one ordering on possible states of affairs *per agent* and ranks the actions available to each agent by the rank of the resulting state of affairs in her ranking on states of affairs. Since this view appeals to different rankings on states of affairs for each agent, there can be no possible problem about a single ranking not being able to do the job. There is no single ranking—only different rankings for each agent.

But importantly, this structural feature of the view doesn't tell us anything about how to interpret these rankings on possible states of affairs. With consequentialism, there is no trouble about how to interpret the ranking on possible states of affairs: it is the *better than* ranking. One state of affairs ranks ahead of another in the ordering on states of affairs appealed to by ordinary consequentialism just in case things would be *better* if the first state of affairs obtained than if the second one did. But the formal move that gives ART the right structure to account for constraints and special obligations does not, by itself, tell us anything at all about how to interpret the orderings on states of affairs that are appealed to by ART. What we know about it, is that it is supposed to be a three-place relation between two possible states of affairs and an agent, such that when we supply an agent to the agent-place, the resulting relation between two states of affairs has the properties of an ordering.⁸ But there is nothing about this formal move that tells us anything more about it.

In principle, there are two possible things that ART-ists can say about this. First, they can say that the interpretation of the *better-than-relative-to* relation is to be given by the theory itself. On this model, the *better-than-relative-to* relation is a theoretical relation, which we know by its fruits. If there were such a relation, and the right things stood in it,

7. More complicated versions of consequentialism may do this indirectly, by using the ordering on possible states of affairs to induce an ordering on "profiles" of states of affairs (cf. Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101 [1991]: 461–82), or on trees of possible outcomes given future choices of the agent (cf. Fred Feldman, "World Utilitarianism," in *Analysis and Metaphysics*, ed. Keith Lehrer [Dordrecht: Reidel, 1975], 255–71), and then using this ordering to induce an ordering on actions available to the agent. These are side issues, for our purposes.

8. I do not mean to be saying that the *better than* relation does not have any additional structure. For example, perhaps it is attributive and requires a *kind* as a further argument (Peter Geach, "Good and Evil," *Analysis* 17 [1956]: 33–42). And plausibly *good*, like *tall* and *fast*, requires yet a further argument over and above this—a comparison class. I take all such further details to be held fixed by the issues under consideration here, and assume that we can safely ignore them.

on this view, then that would give us a highly attractive ART-istic explanation of constraints. Similarly, if there were such things as electrons, and there were pretty much the right ones in order for electromagnetic theory to explain the things that it is supposed to explain, then that would give us an attractive explanation of those phenomena. And that is why we believe in electrons, and what we understand about them—it is not as though we had an independent, pretheoretical grip on or interpretation of what electrons were, and then electromagnetic theory came along and appealed to them in order to explain other things. Everything we know about them is given by the theory, which is attractive on grounds that are independent of our antecedently having at least some grip on what electrons are. Similarly for the *better-than-relative-to* relation, on the view that it is a theoretical relation.

Or, ART-ists could suppose that we *do* have some pretheoretical grip on the *better-than-relative-to* relation—that there is some way in which it fails to be a purely theoretical relation—and that it is this relation that we can already pick out in some other way that they mean their theories to appeal to. In fact, this is what ART-ists universally seem to believe. For example, as Michael Smith says, “If goodness were a relational property of the sort envisaged, then there would be nothing absurd about that. It would simply amount to the familiar view that, as we would put it nowadays, being happy is a relative value, rather than a neutral value.”⁹

Other ART-ists go further and actually attribute views about what is good-relative-to whom to ordinary consequentialists: “A theory is then agent-neutral if and only if it implies that the value of *p* relative to *i* is the same that [*sic*] the value of *p* relative to *j*, for all agents *i* and *j*, for all states of affairs *p*.”¹⁰ “Whereas consequentialism holds that the value of a state of affairs *is something constant for everyone*.”¹¹

And many ART-ists try to emphasize what is plausible about their assumptions about what is good-relative-to whom by trying to draw them out using ordinary language: “Indeed, why should the moral value of the state of affairs as seen from Othello’s position—husband, lover and killer of Desdemona—*have to be* no different from its value as seen from the position of another who is not thus involved?”¹² Explanations in electromagnetic theory require assumptions about where there are electrons and when. But no one supposes that these assumptions are independently plausible, in the way that Sen suggests that the ART-ist’s

9. Smith, “Neutral and Relative Value,” 584.

10. Krister Bykvist, “Utilitarian Deontologies? On Preference Utilitarianism and Agent-Relative Value,” *Theoria* 62 (1996): 124–43, 127.

11. Douglas Portmore, “Combining Teleological Ethics with Evaluator Relativism: A Promising Result,” *Pacific Philosophical Quarterly* 86 (2005): 95–113, 97.

12. Amartya Sen, “Evaluator Relativity and Consequential Evaluation,” *Philosophy & Public Affairs* 12 (1983): 113–32, 118.

assumptions about what is good-relative-to whom are independently plausible.

Some ART-ists even seem to explicitly try to interpret their formal apparatus using ordinary language. The following selection from Campbell Brown is an extreme case (Brown holds that *better than* should be “relativized” to possible worlds and to sets of alternatives, in addition to agents¹³): “A *perspective* is an ordered triple $\langle i, A, w \rangle$, where i is a possible agent, A is a set of alternatives, w is a possible world, and $w \in A$. A *goodness function* is a function F that assigns a goodness ordering to each perspective. We abbreviate $F(I, A, w)$ as $\geq_{i, A, w}$. And we interpret $w \geq_{i, A, w} w'$ as ‘ w is at least as good as w' from the perspective $\langle i, A, w \rangle$ ’.”¹⁴ But if that gloss helped you to understand Brown’s formalism, then your grasp of the English language exceeds mine considerably. My point is: it should be clear from the foregoing quotations that ART-ists are enamored of the idea that we have an antecedent, pretheoretical grasp of the *better-than-relative-to* relation. The important question is why.

D. Antecedent Grasp

The answer is that it is no coincidence that ART-ists typically suppose that it is possible to talk about what is good-relative-to whom in ordinary language. For though the ability of ART to accommodate constraints and special obligations turns only on its formal structure, its attractions, and particularly the idea that it retains whatever is attractive about consequentialism, all turn on the idea that we *do* have some pretheoretical grasp of the *better-than-relative-to* relation. In particular, they turn on the idea that it has something intimate to do with what we are talking about when we talk about what is better than what.

The most general level of this lesson can be drawn out by observing that if the *better-than-relative-to* relation is really just a theoretical relation, on the model of *electron*, then it shouldn’t matter what we call it. The theory should be equally attractive no matter whether we use the words ‘good’ and ‘better’ to express this theoretical posit, or other words, such as ‘orange’ and ‘oranger’. So compare the following two views:

- AR Teleology: For all x , x ought always to do that action that will bring about the most of what is good-relative-to x .
- AR Orangeology: For all x , x ought always to do that action that will bring about the most of what is orange-relative-to x .

13. If you think this is a little bit much, Jennie Louise and Michael Smith advocate also relativizing to times. See Louise, “Relativity of Value”; Michael Smith, “Two Kinds of Consequentialism” (unpublished manuscript, Princeton University, 2006). See also Richard Brook, “Agency and Morality,” *Journal of Philosophy* 88 (1991): 190–212; Frances Kamm, “Harming Some to Save Others,” *Philosophical Studies* 57 (1989): 227–60.

14. Brown, “Consequentialise This,” 26.

You may not understand what it means for something to be orange-relative-to someone. The glib answer to that, on a par with many ART-istic explanations of what it means for something to be good-relative-to someone, is that it is what you get when you relativize *orange* to agents. But the correct answer is that it is just a theoretical relation. If the *good-relative-to* relation is just a theoretical posit, then Agent-Relative Teleology and Agent-Relative Orangeology are really just the same view, simply with different names for the theoretical relation.

But it seems far from obvious that Agent-Relative Orangeology's explanation of constraints and special obligations is deeply more attractive than whatever explanation can be provided by ordinary deontology, that it retains the attractions of consequentialism, whatever those are, and that it avoids the paradox of deontology, much less that it is so deeply compelling that charity requires attributing it to deontologists. If Agent-Relative Teleology and Agent-Relative Orangeology are the same view, then these things should be far from obvious about ART, as well. And if that is so, then using the letters 'g-o-o-d' in the name for this relation amounts purely to persuasive definition. So if anything at all is more attractive about ART than about Agent-Relative Orangeology, it must turn on some pretheoretical grasp that we have of the *good-relative-to* relation, some way or other in which it fails to be purely theoretical.¹⁵

The overarching lesson of Agent-Relative Orangeology is the Preliminary Point of this article. It is just this: that we need some antecedent grasp or other of the *good-relative-to* relation in order for it to turn out that Agent-Relative Teleology has anything over Agent-Relative Orangeology and, hence, in order for it to turn out to be deeply attractive in any way whatsoever, no matter what that way might turn out to be. That is the essence of why ART-ists are so keen on trying to make claims about the *good-relative-to* relation in ordinary, pretheoretical, English. They would like us to think that it is something that we already understood before their theory came along, because only if this is so do we have any grounds to think that their theory is any more attractive than Agent-Relative Orangeology.

But unfortunately, it is far from obvious that this is so. In Section II I'll survey a few of the ways that ART-ists have tried to make claims about the *good-relative-to* relation in pretheoretical English, and argue that they have not succeeded. I'll also show that no one has ever made an uncontroversial distinction between agent-relative and agent-neutral

15. I mean no more by "pretheoretical grasp" than this. It doesn't follow from the argument that we must pretheoretically understand anything about the *good-relative-to* relation; just that it cannot be a purely theoretical relation. It must be one we are able to have intuitions about.

value to which ART-ists could appeal. And I'll argue that this means that there is at least one traditional advantage of consequentialism that ART-ists definitely do not retain.

II. AGENT-RELATIVE VALUE AND 'GOOD'

A. 'Good For'

A first pass would be the simple thought that we can pretheoretically talk about what is good-relative-to whom by talking about what is good *for* whom: "A relativist with respect to intrinsic value holds that a statement ascribing value to some x is incomplete if it fails to indicate *for whom* x is valuable. The logically perspicuous statement ascribing intrinsic value always has the following form: 'X is intrinsically valuable for or to S'."¹⁶ But importantly, though Agent-Relative Teleology is historically modeled on egoism, Agent-Relative Teleology is not simply a version of egoism. According to egoism,

Egoism: For all x , x ought always to do that action whose results will be best for x .

Egoism looks like consequentialism, with claims about what is better than what relativized to agents. Agent-Relative Teleology claims to be just like consequentialism, but with claims about what is better than what relativized to agents. So it is natural to mistakenly hold that Agent-Relative Teleology is just a peculiar version of egoism.¹⁷

But that would be wrong. Egoism appeals to the *good for* relation, with which we are perfectly comfortable in ordinary English (Moore's confusions aside).¹⁸ A tax policy can be good for Dick Cheney's pals without being good or good without being good for Dick Cheney's pals. We know what that means. The *good for* relation, moreover, is not, strictly speaking, *agent*-relative at all but, rather, *subject*-relative, in some extended sense. Rain is good for trees, but trees are not agents. So egoists aren't committed to there being any specifically agent-relative *good* concept at all.

More to the point, if ART-ists did not distinguish the *good-relative-to* relation from the *good for* relation, they would be committed to bizarre claims about what is better for whom. Special obligations, a special case

16. Diane Jeske and Richard Fumerton, "Relatives and Relativism," *Philosophical Studies* 87 (1997): 143–57, 144.

17. Compare, e.g., John Broome, *Weighing Goods* (Oxford: Blackwell, 1991), 6–16; Shelly Kagan, "The Structure of Normative Ethics," *Philosophical Perspectives* 6 (1992): 223–42, 234; Dreier, "The Structure of Normative Theories," 25–29; Smith, "Neutral and Relative Value," 578–85; and Portmore, "Combining," 97.

18. G. E. Moore, *Principia Ethica* (1903), rev. ed., ed. Thomas Baldwin (Cambridge: Cambridge University Press, 1993), secs. 59–60. Moore claimed not to be able to even make sense of something being good for someone without being good.

of constraints, seem to require us to make sacrifices of our own well-being—to do what is worse for ourselves—for the sake both of our loved ones and of those to whom kinship or promises engender other kinds of duties. When special obligations require sacrifice, they ask us to do things that are *not* better for us. So if constraints requiring self-sacrifice are to be accounted for by Agent-Relative Teleology, the *good-relative-to* relation must not be the same as the *good for* relation.¹⁹ Explicit advocates of Agent-Relative Teleology, of course, are fully aware that their theory must appeal to a relation that is distinct from the *good for* relation: “It is important not to conflate what is good *from your perspective* and what is good *for you*.”²⁰ But the point is an important one to keep in mind, as we troll through possible ways of trying to make claims in English about what is good-relative-to what.

B. ‘From the Point of View Of’

We’ve now seen that the *good-relative-to* relation is not expressed by ‘good for’ in ordinary English—it is not the same as the *good for* relation. Some philosophers have tried to elucidate the *good-relative-to* relation in another way that may seem initially more promising. For example, Thomas Hurka tells us that agent-relative goods are things that are good “from the point of view” of some agent or other.²¹ Hurka thinks it is plausible that whether or not Franz’s murders are worse for Franz than Hans’s and Jens’s murders are, they are certainly worse from Franz’s point of

19. Let me be perfectly clear that I am not ruling out the possibility that someone might try to account for constraints and special obligations by appealing to the *good for* relation and claiming that apparent sacrifices required for special obligations are really only illusory, because the well-being of the person to whom you have made a promise really is intrinsically good for you. I’ve only claimed that (1) such assumptions about what is good for whom are highly implausible, if not bizarre, and (2) such a view is not really appealing to a kind of agent-relative value, because ‘good for’ is not relative only to agents—things can be good for trees or even for the ozone layer. So it should be clear that this is not the research program that is actually being advanced, however misleadingly its proponents may be prone to state their views. It is a different view, worth being discussed on a different occasion.

20. Brown, “Consequentialise This,” 21. Also, “there may well be a variety of ways in which we could conceive of the property of being good as a relational property of the required kind. However, my own view is that the best way of doing so is by giving a detailed statement and defense of a particular version of the dispositional theory of value” (Smith, “Neutral and Relative Value,” 591). See also Amartya Sen, “Well-Being, Agency, and Freedom: The Dewey Lectures 1984,” *Journal of Philosophy* 82 (1985): 169–221, 206–8; and Portmore, “Combining,” 97.

21. Hurka refers to the “concept of agent-relative goodness, or *what is good from a person’s point of view* and so gives her (and perhaps only her) reason to pursue it” (Thomas Hurka, “Moore in the Middle,” *Ethics* 113 [2003]: 599–628, 611).

view. Sen and Portmore agree.²² But I don't understand what they are talking about when they say these things.

Now there is more than one way to interpret this talk about "points of view." For example, on one interpretation, this view is to be combined with the view that 'good' has an indexical character and the view that 'from the point of view of' is an operator on characters, so that "X is good from the point of view of A" is true just in case A could speak truly by saying "X is good." I suspect that Sen intends the "point of view" operator to work in something roughly like this way.

This is not a credible view; it has (at least) two problems. First, it is not independent from the view that 'good' has an indexical character, so it can only be as plausible as that view. But I'll explain what is problematic about such views in Section III. And second, even if 'good' does turn out to have an indexical character, if this is genuinely to provide us with a way of talking about the *good-relative-to* relation in ordinary language, then the 'from the point of view of' operator must be one that exists in ordinary language, not just a technical device that Sen invents. But it is not at all plausible that there is such a device in ordinary language. For if there were, then it ought to be able to operate on other sorts of contents. But it does not make sense for me to say, "I published an article in *Philosophy & Public Affairs* in 1982, from Sen's point of view" on the grounds that Sen published an article in *Philosophy & Public Affairs* in 1982. So it seems implausible to suppose that there is an operator in English that does what this interpretation of the "point of view" or "perspective" talk would need it to do.

On another interpretation, the "point of view" talk functions as an operator similar to the tense and modal operators. On this view, points of view must make it in to circumstances of evaluation for propositions, along with worlds and times. Just as "Jim is taller than Andy" can be true at some possible worlds and false at others, and true at some times and false at others, it can be true from some points of view and not from others. But this interpretation does not make very much sense, either. It is one thing to suppose that you understand what it is for "Franz's murders are worse than Hans's" to be true from one point of view but not another. It is another thing entirely to suppose that this is a general feature of propositions, applying to "Jim is taller than Andy"

22. "I would like to explore the possibility that [moral valuations] are coherently interpretable as 'positional' statements, reflecting the view of the state *from the point of view of the evaluator*" (Amartya Sen, "Rights and Agency," *Philosophy & Public Affairs* 11 [1982]: 3–39, 35). And here is Portmore: "S₁ is, objectively speaking, better than S₂ *from the position of an innocent bystander*, and S₂ is, objectively speaking, better than S₁ *from the agent's position*" (Portmore, "Combining," 97).

as well, in the way that times and possible worlds are things with respect to which all propositions need to be evaluated.

So there are a number of ways of trying to interpret what “point of view” talk is supposed to be doing. But there is only one interpretation that seems to stand on its own as helping us to understand what “good-relative-to” talk is all about. And on that interpretation, according to the point of view conception, talk about what is good-relative-to whom should really be understood in terms of talk about what is good that we already understand, together with a special kind of proposition-taking connective, the *point of view* connective, which we can think of as taking propositions as objects in the way that ‘believes that’ and ‘desires that’ do. So on this interpretation it follows that to say that Franz’s murders are worse-relative-to Franz than Hans’s and Jens’s are, is to say that from Franz’s point of view, Franz’s murders are worse than Hans’s and Jens’s are. This way of talking has led some philosophers to wrongly suspect that agent-relative value is a kind of subjective value—value that only exists from points of view and, hence, is only believed to exist.

But this would be a disaster for the Agent-Relative Teleologist to claim. For then it can only explain why Franz is under a constraint not to murder by supposing that Franz’s point of view is wrong about what is better than what. Since the situation is symmetric, it is simply false that Franz’s murders are worse, *simpliciter*, than Hans’s and Jens’s put together. Whether Franz’s point of view is a matter of what he believes, or what he ought to believe, or how he ought to treat things as being—no matter how we interpret the “point of view” operator—the point remains that we get an explanation of constraints and special obligations only if we assume that points of view are systematically and predictably wrong about what is better than what. I conclude that “point of view” talk sheds light on the *good-relative-to* relation only if it undermines the plausibility that constraints and special obligations have anything deep to do with what is good-relative-to whom.

C. An Uncontroversial Distinction?

You are now likely to be wondering about the following objection: mustn’t there be *some* way of talking about what is good-relative-to whom in ordinary language? After all, not only do ART-ists talk about it all of the time, there is an important and uncontroversial distinction in contemporary ethics between agent-relative and agent-neutral value, and as Michael Smith points out, all that the ART-ist needs is to appeal to *that*: “If goodness were a relational property of the sort envisaged, then there would be nothing absurd about that. It would simply amount to the

familiar view that, as we would put it nowadays, being happy is a relative value, rather than a neutral value.”²³

But this, I will now suggest, is simply not so. There is no uncontroversial distinction between agent-relative and agent-neutral values, because no one has ever made such a distinction in a way that was not motivated by trying to give an ART-istic account of cases like constraints. The introduction of agent-relative value presupposes the attractiveness of ART, so it follows that we can’t appeal to our independent grasp of agent-relative value in order to explain why ART is attractive.

How could this be? The answer is very simple. In *The Possibility of Altruism*, Thomas Nagel was the first to begin discussing agent-relative and agent-neutral value, which he there called “subjective” and “objective” value.²⁴ But contrary to popular myth, Nagel never made a distinction between agent-relative and agent-neutral value. What he did was to make a distinction among *reasons*, to make two highly controversial assumptions about reasons that were tantamount to importing teleology, and then to use those assumptions to *posit* that there is such a thing as agent-relative value.

The uncontroversial distinction that Nagel made was between reasons that are reasons for everyone and reasons that are reasons for only some people. He used some unnecessary technical apparatus in order to do so, which led to the talk about “free agent-variables” and so forth. But the idea was that if a reason is a reason for only some people, say, for Tom, the weakest modally sufficient condition for it to be the case that Tom has that reason will have to mention some feature of Tom that distinguishes him from the people for whom that consideration is not a reason. And hence it will have a “free agent-variable.” And if a reason is a reason for anyone, no matter what she is like, then the weakest modally sufficient condition for it to be the case that Tom has that reason will not have to mention Tom—since whatever makes it a reason will make it equally a reason for anyone. And hence it will have no “free agent-variable.”

This uncontroversial distinction has nothing, on the face of it, to do with constraints or special obligations or anything else that might conflict with consequentialism.²⁵ Suppose, for example, that there is a reason for anyone not to murder, no matter what she is like. According to this definition, that would be an agent-neutral reason. But such reasons could explain constraints not to murder. The same reason would

23. Smith, “Neutral and Relative Value,” 584.

24. Thomas Nagel, *The Possibility of Altruism* (Princeton, NJ: Princeton University Press, 1970), 90–96.

25. This is pointed out by David McNaughton and Piers Rawling, “Agent-Relativity and the Doing-Happening Distinction,” *Philosophical Studies* 63 (1991): 167–85.

be a reason for Franz not to murder, a reason for Hans not to murder, and a reason for Jens not to murder—a reason for *anyone* not to murder.

In order to make this uncontroversial distinction track issues that are related to consequentialism, Nagel stipulated that there is no such thing as a reason not to murder. All reasons, Nagel stipulated, are reasons in favor of actions of the form, “promote state of affairs *p*.” So there is no such thing as a reason not to murder. There is only a reason to promote the state of affairs that Franz doesn’t murder, a reason to promote the state of affairs that Hans doesn’t murder, and a reason to promote the state of affairs that Jens doesn’t murder. So by introducing this controversial stipulation, Nagel ensured that his uncontroversial distinction would track the issues related to consequentialism, including constraints and options. Given the stipulation, any reasons that Franz, Hans, and Jens have that would explain their constraints not to murder must not be reasons for everyone—they must be reasons for only them: agent-relative reasons.

This stipulation of Nagel’s is already highly controversial. But then Nagel made a further, highly controversial move. Without ever even trying to distinguish between agent-relative and agent-neutral value, he simply stopped talking about agent-relative and agent-neutral reasons and substituted talk about agent-relative and agent-neutral value. Notice that he *could not have* succeeded at drawing a distinction between agent-relative and agent-neutral values in the same way that he drew his distinction between agent-relative and agent-neutral reasons. For that would involve the claim that agent-relative values are things that are good for some people but not good for everyone and that agent-neutral values are things that are good for everyone. But then it would turn out that agent-relative values were just a special case of the *good for* relation, and we’ve already seen in Section II.A why that is wrong.²⁶

Nagel’s move seemed natural, however, because if agent-neutral reasons are reasons for anyone to promote some state of affairs, *p*, it is natural to think that there must be something good about state of affairs *p* in such cases. So, he concluded, by analogy there must be some sense in which state of affairs *p* is good *relative* to some agent, if there is an agent-relative reason for *that agent* to promote *p*. But this is teleological reasoning. It doesn’t find an uncontroversial distinction between agent-relative and agent-neutral value at all, but just posits agent-relative values to correspond to agent-relative reasons, in the way that it assumes agent-neutral values correspond to agent-neutral reasons.

Nagel didn’t successfully distinguish between agent-relative and agent-neutral value in a non-theory-driven way, and no one has since.

26. Moreover, it is easy to see that things that are good are not necessarily good for everyone. For example, a tax policy might be good but not good for Dick Cheney’s pals.

This means that there is no uncontroversial distinction between agent-relative and agent-neutral value for ART-ists to appeal to, in order to make the case that their theory has the attractions that they claim for it and, in particular, that the assumptions they make about what is good-relative-to whom are independently plausible. And that should undermine our confidence that we *must* have *some* way of talking about the *better-than-relative-to* relation in ordinary language. We need to see what is deeply attractive about ART in the first place, in order to see why it is worth positing the *better-than-relative-to* relation.

D. The Moral

In Section I, I explained why ART can't plausibly be a highly attractive view, unless the *good-relative-to* relation is one of which we have some kind of pretheoretical grasp. And in Section II, I've been starting to assemble a case that it should be far from obvious that we do have any pretheoretical grasp of the *good-relative-to* relation. There don't seem to be ways of picking it out in ordinary English using the word, 'good', as ART-ists so often presuppose, and no one has ever made an uncontroversial and independently motivated distinction between agent-relative and agent-neutral value, which ART-ists could put to work.

Why is this so important? First, because it undermines ART-ists' claims to be espousing a view that is in some way deeply attractive—so deeply attractive, perhaps, that it would be uncharitable to assume that anyone doesn't believe it, according to Dreier's Conjecture. But more; if there is no ordinary-language way of talking about what is good-relative-to whom, it follows almost immediately that at least one of the great attractions of ordinary consequentialism cannot be had by ART.

One of the important attractions of ordinary consequentialism is that consequentialist explanations of what people ought to do appeal to assumptions whose plausibility we can independently evaluate. It is independently plausible, for example, that pleasure is good and that pain is bad. ART can retain this advantage of ordinary consequentialism only if the assumptions that it requires about what is good-relative-to whom are also independently plausible. But these assumptions can be independently plausible, it seems, only if we have some way of saying what they are. So to retain this advantage of consequentialism, ART-ists need to suppose that there is some ordinary language way of saying that something is better-relative-to one person than it is better-relative-to another. And that is what I've just been arguing that it is not at all obvious that ART-ists can do.

But it could be that even if there is no independent evidence for the assumptions that ART needs to make about what is good-relative-to whom, the explanations that ART *could* provide of constraints and special obligations, *if* they were true, are still attractive enough in their own

right to grant ART some deeply compelling quality, just as the attractiveness of the explanations of electromagnetic theory make it worth positing that there are electrons, even though assumptions about which electrons are where are not independently plausible. And it is to that question that we will now turn, in Section III. Can ART capture what is supposed to be deeply compelling about consequentialism?

III. TELEOLOGY AND ‘GOOD’

A. *Avoiding the Paradox of Deontology*

Of course, ART-ists have supposed, ART retains the explanatory advantages of consequentialism. After all, proponents suggest, ART *just is* consequentialism, simply with a new and improved theory of value—agent-relative value. Just as Mill’s utilitarianism improved on Bentham’s by allowing for two kinds of pleasure and Moore’s consequentialism improved on Mill’s by allowing for other basic intrinsic goods, ART-ists claim that their view improves on ordinary consequentialism by simply filling in a more sophisticated axiology.²⁷ On this view, ART *just is* a version of consequentialism.²⁸ So since it is a version of consequentialism, it obviously retains consequentialism’s advantages. Therefore it is supposed to get both the advantages of consequentialism and those of deontology, by accounting for constraints.

This should sound surprising. For constraints were supposed to be a counterexample to consequentialism. They were putatively paradoxical precisely because they appear to be cases in which an agent is required to do what will have a result that is less good than some other available result. And what consequentialism says is that:

Consequentialism: Every agent ought always to do what will lead to the outcome that is best.

Which obviously entails the thesis that Dreier and Portmore have claimed is so attractive:

Compelling Idea: It is always permissible for every agent to do what will lead to the outcome that is best.

But constraints are, on the face of it, counterexamples to the Compelling Idea. That is what has been thought to make them paradoxical. So it would seem to follow that any view allowing for constraints would be inconsistent with the Compelling Idea, and hence inconsistent with consequentialism.

Not so, say Agent-Relative Teleologists. The principal attraction of

27. Portmore, “Combining.”

28. Smith, “Neutral and Relative Value”; Brown, “Consequentialise This.”

their view, Dreier and Portmore claim, is that it allows for constraints while also entailing the Compelling Idea:²⁹

There seems to be something about consequentialism that even its critics find compelling. If not, consequentialism would surely have been dismissed long ago. . . . So what about it is so compelling? Well, it seems to be the very simple and seductive idea that it can never be wrong to produce the best available state of affairs. . . . The thought that it is always permissible to pursue the best available state of affairs is something shared by all teleological theories, both agent-neutral and agent-relative.³⁰

The simple answer we may now give is that every moral view is consequentialist, that we common sense moralists as much as anyone are out to maximize the good. Of course, our understanding of the good may be an agent-centered one, whereas the typical challenger has an agent-neutral understanding, . . . We don't have to be embarrassed by the charge that we are ignoring the good, because the charge is just false.³¹

Similarly, Michael Smith claims that the attraction of ART is that it allows for special obligations (a special case of constraints) while also entailing consequentialism.³² How could this be?

A first pass at the answer to this puzzle is to observe that ART-ists do believe something different:

ART: Every agent x ought always to do what will lead to the outcome that is best-relative-to x .

which entails

Permissible ART: It is always permissible for every agent x to do what will lead to the outcome that is best-relative-to x .

The puzzle about why ART-ists think that they can accept both Consequentialism and the Compelling Idea while also allowing for constraints has something to do with the fact that they do accept these two principles, which bear a vague resemblance to Consequentialism and to the Compelling Idea.

The remainder of the answer is that ART-ists typically claim that some kind of contextualist theory about the semantics of 'good', 'better', and 'best' is correct. Their idea is that given the right semantics for 'best', it will turn out that what consequentialism really says is ART and that what

29. Dreier, "The Structure of Normative Theories"; Portmore, "Combining."

30. Portmore, "Combining," 98–99.

31. Dreier, "The Structure of Normative Theories," 24–25.

32. Smith, "Neutral and Relative Value."

the Compelling Idea really says is Permissible ART. (In the best-case scenario for ART, ART-ists would have provided a semantics for ‘best’ which yielded this result.) On this view, ART-ists get to accept Consequentialism, because they have a creative semantics for the sentence that is used to state Consequentialism, on which it turns out to mean ART. This is the basis for ART-ists’ claim to be defending a version of consequentialism, and the basis for their claim to accommodate the Compelling Idea.

B. *Why This Is Wrong*

There are a variety of problems with this view, however. The biggest is that it is not enough to accept some view, to provide a creative semantics for the sentence expressing it on which you can accept that sentence. To accept some view, you have to accept the sentence expressing that view on the semantics *on which it expresses that view*.³³ Any atheist can accept the sentence ‘God exists’, if given a semantics on which ‘God’ has my wristwatch as its referent, or on which ‘exists’ is synonymous with ‘is preposterous’. This does not make them theists. To be theists, they must accept the sentence ‘God exists’ under the semantics on which it means that God exists.

But there is compelling evidence that any creative semantics for the sentence stating consequentialism on which what it really means is ART would be one on which it means something other than what ordinary consequentialists have always meant by it. One such piece of evidence we have already seen. It is that constraints are widely supposed by competent speakers to be counterexamples to the sentence stating ordinary consequentialism. So if constraints are not counterexamples to ART, then the sentence stating ordinary consequentialism must mean something other than ART.

Another piece of evidence that such a semantics would have to be wrong is that ordinary consequentialists, along with everyone else, have always understood the Compelling Idea in such a way that it validates the following inferences, nominalist qualms aside:

1. It is always permissible for everyone to do what will have the best results.
2. It is always permissible for everyone to do what will have results that rank highest in the *better than* ordering.
3. There is an ordering, the *better than* ordering, such that it is always permissible for everyone to do what will have the results that rank highest in it.

33. See Mark Schroeder, “Realism and Reduction: The Quest for Robustness,” *Philosophers’ Imprint* 5 (2005): 1–18 (www.philosophersimprint.org/005001/) for further discussion.

Sentence 1 states the Compelling Idea. But competent speakers of English have always understood it in such a way as to validate the inferences to 2 and then to 3. But sentence 3 is inconsistent with ART. So that is evidence that ART does not plausibly tell us what ordinary speakers of English have meant by the sentence stating consequentialism all along.

Tom Hurka has suggested, in correspondence, that this argument is question-begging against ART-ists, because it presupposes that the correct semantics for the Compelling Idea is the one which validates the inferences from 1 to 2 and from 2 to 3, but obviously ART-ists believe that this is not so. Understanding his worry is important, in order to understand the structure of what I've just been arguing. I agree with Hurka that it is obvious that ART-ists do not believe this to be the case, but I did not mean to be simply presupposing that they are wrong. What the argument does, is to offer inferences that speakers have always found perfectly natural as evidence about the semantics of the sentence expressing the Compelling Idea. Since speakers of English have always found the inferences from 1 to 2 and from 2 to 3 to be perfectly natural, nominalist qualms aside, I think it follows that when speakers of English find sentence 1 compelling, they are finding something compelling that is inconsistent with ART—not permissible ART, which follows from it. I think this is compelling evidence that ART-ists cannot accommodate the idea that ordinary speakers of English find compelling, when they consider whether it is always permissible to do what will be best.

One final piece of evidence: if ART-ists were right that the sentence stating consequentialism really meant ART, and those ART-ists were right who accept Dreier's Conjecture that everyone really believes ART, then it would follow that everyone accepts consequentialism. To their credit, those ART-ists who both think that ART is a version of consequentialism and accept Dreier's Conjecture endorse this result.³⁴ But it boggles the imagination to suppose that the sentence stating consequentialism could have a semantics that has so eluded the understanding of speakers of English. Competent speakers of English have uniformly, until very recently, supposed that the sentence stating consequentialism expresses a view that is enormously controversial in ethical theory. Any semantics according to which it instead expresses a view that everyone believes must attribute massive error in linguistic competence to speakers of English. So any such semantics is thereby rendered extremely implausible.

So, to recap: in the best-case scenario for ART, ART-ists would have an account of the semantics of 'best' on which the sentence stating consequentialism really means ART and on which the sentence stating the Compelling Idea really means Permissible ART. But for the foregoing

34. Louise, "Relativity of Value"; Brown, "Consequentialise This"; Smith, "Neutral and Relative Value," and "Two Kinds of Consequentialism."

reasons, I think that there are excellent grounds to think that any creative semantics which successfully yielded these results would thereby have changed the subject. It would have shown how to parrot the kinds of things that consequentialists got to say, but not how to say what consequentialists said. ART-ists are still committed to the same putatively paradoxical result as ordinary deontologists—that there are some situations in which agents are required to do what will lead to a(n agent-neutrally) worse result. If they weren't committed to this, then they wouldn't have captured constraints.

C. *We're Not in the Best-Case Scenario*

In any case, we are not even in the best-case scenario for ART. It is not as though ART-ists have actually offered any semantics for the sentence expressing the Compelling Idea on which what it means is Permissible ART, and hence is compatible with their view. In actual practice, what ART-ists have done is simply to assert that they have captured the Compelling Idea, without explaining how. This is a problem, because on the most developed view about the semantics of 'best' that it is possible to glean out of ART-ists' discussions, ART is provably inconsistent with the Compelling Idea. I am going to pause for a few paragraphs, here, to explain why this view cannot possibly work, because I think it yields a general moral about the prospects for ART-ists of being able to successfully get us into the best-case scenario of the last section.

According to Sen and Portmore and Smith, the connection between 'good' and the *good-relative-to* relation is that 'good' expresses the *good-relative-to* relation in a context-dependent way. For example, according to Smith, "If this is right, however, then, as is perhaps already clear, it turns out that 'good' is indeed subscripted in just the way required. For when I judge *p*'s being the case in *C* to be good, I am judging that *p*'s being the case in *C* has a certain relational property. . . . In other words, I am really judging *p*'s being the case in *C* to be good_{me} and you are really judging *p*'s being the case in *C* to be good_{you}. Our judgments are appropriately relational."³⁵ Sen and Portmore also tell us that the truth of evaluations (sentences calling something good) is relative to the evaluator (to the speaker). The concrete view that these remarks most strongly suggest is the view that 'good' is an *indexical*, picking out the relational property, *good-relative-to x*, where *x* is the speaker of the context.³⁶

The indexicalist view, however, is hopelessly wrong. It is wrong on

35. Smith, "Neutral and Relative Value," 591–92.

36. There is also strong further evidence that this is the best interpretation of the view espoused in Smith, "Neutral and Relative Value"; it is the best interpretation, e.g., of why he thinks that his view is subject to a potential problem about disagreement.

independent grounds, of course, because it predicts that if Franz says, “my murders are worse than Hans’s” and Hans says, “my murders are worse than Franz’s,” then we should all (including both Franz and Hans) be able to agree that both speak truly. But even if you can get your head around the idea that there is something appropriate about both saying these things, it hardly seems appropriate for Franz to allow, “Hans speaks truly when he says, ‘Hans’s murders are worse than Franz’s’; nevertheless, Hans’s murders are not worse than Franz’s.” But it is a prediction of the indexicalist account—indeed, of all contextualist accounts—that such assertions will sometimes be appropriate. Compare “I”: there is nothing puzzling about Franz saying, “Hans speaks truly when he says, ‘I am in Boston’; nevertheless, I am not in Boston.” So the indexicalist view makes bad predictions and should be rejected on independent grounds.³⁷

But never mind that. My purpose here is not to explain why contextualist analyses of ‘good’ are bad views, but to explain why they cannot even get ART-ists what they want. The indexicalist view creates even bigger problems for ART, because if it is right, then not only does ART not predict the Compelling Idea, it is outright inconsistent with it. This is easy to see:³⁸ if the indexicalist view provides the right semantics for ‘good’, then the Compelling Idea gets interpreted as follows:

Compelling Indexical: For all x , it is always permissible for x to do what will bring about the most good-relative-to me.

But according to ART:

ART: For all x , x ought always to do that action that will bring about the most good-relative-to x .

Assuming that if you ought to do something, it is not permissible not to do it, it is easy to see that ART and Compelling Indexical are consistent only on the assumption that for all x , what brings about the most good-relative-to x is the same as what brings about the most good-relative-to me. So ART and Compelling Indexical are consistent only if what is good-relative-to each agent is always the same. But the *whole point* of ART was to capture constraints by assuming that what is good-relative-to each agent is *different*. So on the indexicalist view, ART makes the Compelling Idea turn out to be true only if it fails to accommodate constraints.

37. See Mark Schroeder, “Expression for Expressivists,” forthcoming in *Philosophy and Phenomenological Research*, for further discussion.

38. I’ve pointed out this much in Mark Schroeder, “Not So Promising After All: Evaluator Relativism and Common-Sense Morality,” *Pacific Philosophical Quarterly* 87 (2006): 348–56.

It is important not to be misled, as it is hard not to suspect that some ART-ists have been, by the fact that it follows from ART that Compelling Me is indexically valid:

Compelling Me: It is always permissible for me to bring about the most good-relative-to me.

Compelling Me is not strong enough to capture what is Compelling about consequentialism and putatively paradoxical about moral common sense. Someone who accepts Compelling Me but rejects the Compelling Idea is like the nonuniversal egoist. She believes that it is always permissible for herself to bring about the most good but that other people are sometimes required not to. And the Compelling Idea no more follows from the fact that Compelling Me is indexically valid than it follows that everyone is here now, just because ‘I am here now’ is indexically valid. For on the indexicalist view, ‘good’ is *also* an indexical.

D. The General Problem

Of course, the indexicalist view is only one—and perhaps the worst—view on which the word ‘good’ somehow means *good-relative-to*, as it would have to do in order for it to turn out that the sentence expressing the Compelling Idea really means permissible ART. Indexicals are only one model for context-dependent terms in natural languages. Other models include pronouns, graded adjectives, and qualifiers like “local.” But the problem is that none of the other models for context-dependence put ART-ists in their best-case scenario, either.

Each of the other models for context-dependence would predict that there should be some reading of the sentence expressing the Compelling Idea on which it means Permissible ART—so that is good. But each of the other models also yields the prediction that there should be some reading of the sentence expressing the Compelling Idea on which what it means is:

Compelling A: It is always permissible for anyone to do what will bring about the most good-relative-to A.

for some value of “A.” But all such readings are subject to the same problem as the indexical account—they are consistent with ART only on the assumption that the *good-relative-to* relation is invariant across agents, which is precisely what ART-ists need to deny in order to capture constraints.

For example, on the pronominal model, the semantics for the Compelling Idea would look like:

Pronoun Compelling: It is always permissible for anyone to do what will bring about the most good-relative-to her.

This account yields one reading—the anaphoric reading—on which the pronoun is bound by the quantifier, and what it says is Permissible ART. But it also yields a distinct reading on which the pronoun functions deictically and is unbound by the quantifier. To get this reading, try saying it while pointing at the woman on your left when you say “her.” And things go similarly, so far as I can tell, no matter which natural-language model we take for the semantics of ‘good’.³⁹

There are two problems with this. The first is that this means that there are no good models for the semantics of ‘good’ that would actually put ART-ists in their best-case scenario, in which their semantics predicts that the sentence expressing the Compelling Idea really means Permissible ART. So at best, using such models, they can only capture one reading of the Compelling Idea. But the second problem is that this shows that these semantics for ‘good’ are all simply wrong. Ordinary competent speakers of English do not detect an ambiguity in the sentence expressing the Compelling Idea, and so any semantics that predicts that it should be ambiguous is highly suspect, anyway.

So again to recap: the moral of Section III.C was that ART-ists are not in their best-case scenario for validating their claim to have accommodated the Compelling Idea. The moral of Section III.D is that it is unpromising to think that they can get there. And the moral of Section III.B was that even if they did get there, we have more cause to think that they would have changed the subject, than that they would have accommodated the Compelling Idea that was supposed to have been so attractive about consequentialism all along. Moreover, even if the argument has gone astray somewhere, my über-moral remains: the attractions of consequentialism do not come for free to ART; they require some kind of serious work, and that serious work has not been done. The formal move of “relativizing,” or “recognizing the possibility of agent-relative value,” does result in a view with the right structure to accommodate constraints, but by itself it

39. The reason for this is simple. An adequate semantics for ‘best’ must do more than tell us how ‘best’ works in this particular sentence. It must generate a prediction about how ‘best’ works in this particular sentence, on the basis of a general account of how ‘best’ works in all sentences—including ones in which there is no quantifier to bind the agent who something is being said to be best-relative-to. For example, it must deal with the sentence, “It would be best if Franz didn’t murder.” If ‘best’ picks out *best-relative-to*, then there must be some mechanism by which an agent is supplied so that we can interpret this sentence—it could be Franz, it could be the speaker, it could be any contextually salient person—it doesn’t matter. Whatever mechanism allows for this will also potentially be at work in the sentence, “It is always permissible for everyone to do what will have the best results.” So at best, this sentence will receive two readings, only one of which is the one ART-ists want.

does nothing to show why the resulting view is in any way more attractive than Agent-Relative Orangeology.

IV. WHAT ELSE COULD MAKE ART ATTRACTIVE?

A. *Other Attractions*

All of this should make us wonder: if positing agent-relative value is supposed to be merited because it allows for an attractive explanation of constraints, but its explanation of constraints is supposed to be attractive primarily because it accommodates the Compelling Idea, and this turns out to be false, then why should we even believe in agent-relative value in the first place, much less in a moral theory built on it? I say: let's not. I don't think there is anything which deserves to be called "agent-relative value."

Still, though capturing the Compelling Idea—avoiding the paradox of deontology—is widely supposed to be the chief attraction of consequentialism, and thus the chief advantage of ART over ordinary deontology, it is not the only general attraction of consequentialism that has ever been noted. So the natural thing to wonder is whether ART can retain some other generally attractive feature of consequentialism. It is therefore worth taking a look at whether the arguments from Section III will generalize to other proposals about what is attractive about ART.

Such possible attractions can be divided into two categories, for simplicity of discussion. In one category are attractive features of consequentialism whose statement involves words like 'good' or 'value'. The Compelling Idea is just one example of such an attractive feature of consequentialism. Michael Smith tells me that it is independently attractive about consequentialism, and thus about ART, that they explain facts about what people ought to do in terms of facts about *values*. Others, expressing essentially the same thought, say that it is attractive about consequentialism that it explains facts about what people ought to do in terms of facts about what is good:

Explained by Good: For all agents x , the facts about what x ought to do are explained in terms of facts about what is good.

I suspect that the same considerations which show that ART cannot really capture the Compelling Idea will suffice to show that ART cannot capture attractions like these, either. These attractive consequences of consequentialism, after all, are set out using ordinary language terms like 'good', 'better', and 'value'. So it is easy to see why they follow from ordinary consequentialism, which is itself set out in terms of 'good', 'better', and 'value'. But it is hard to see why they follow from ART, which on the face of it has nothing to do with what is good, better, or

of value, but only with what is good-relative-to, better-than-relative-to, or of value-relative-to. So just as with the Compelling Idea, it follows that in order to capture these attractive consequences of ordinary consequentialism, such as Explained by Good, ART-ists owe us some story about the semantics of Explained by Good which validates the prediction that it is both (1) a consequence of ART and (2) plausibly part of what ordinary competent speakers of English have understood to be attractive about consequentialism all along.

But not only have no ART-ists actually done the work to back up their assertions to have captured such attractive features of consequentialism by filling in such a story, it is highly plausible that no such story can be given. In order to capture Explained by Good, ART-ists, it seems, must be able to give it the following reading:

ART Explain: For all agents x , the facts about what x ought to do are explained in terms of facts about what is good-relative-to x .

But these readings, it seems, can't be the ones that consequentialists have found appealing all along. For the same reasons as before, again absent nominalist qualms, consequentialists understand Explained by Good in such a way that it has as a consequence that there is a property, *good*, such that for all x , facts about what x ought to do are explained by facts about what is good. But this does not follow from ART Explain. So though ART-ists may be able to capture ART Explain, that is not enough to capture the same feature that has been thought to be attractive about consequentialism all along.

As before, ART is simply not capable of capturing the features that are thought to be attractive about consequentialism. I see no reason not to think that this same result will hold for any attractive feature of consequentialism that is spelled out using words like 'good', 'better', or 'value'. For any such attraction, ART may, through some creative semantics that ART-ists have yet to supply, be able to mimic the attractive features of consequentialism by allowing for a reading on which the sentence expressing that attractive feature of consequentialism is true. But my suspicion is that the result does generalize: on no such reading will ART be able to capture the right reading of this sentence. If this is right, then it will not turn out to be true that ART succeeded at capturing any such attractive features of consequentialism.

B. Simplicity and Elegance

The remaining attractions of consequentialism, then, will be ones that do not mention anything specifically about 'good', 'better', or 'value'. They will be, it seems, structural attractions of consequentialism—ones

that would be equally attractions of views that were like consequentialism in structure, but did not appeal to the *good*. For example:

Agent-Neutral Orangeology: For all x , x ought always to do that action that will bring about the most orange.

Agent-Neutral Orangeology may not seem to be a highly attractive view to you, in which case you will agree with me that capturing whatever is attractive about it may not be capturing very much. But one of the standard claims about what is so attractive about consequentialism is precisely a feature that it does share with Agent-Neutral Orangeology: its theoretical simplicity and elegance.

One of the main attractions of consequentialism has always been that it is a simple and elegant moral theory that avoids appealing to long and complicated lists of rules, in the way that paradigm historical examples of deontological views did. Agent-Neutral Orangeology is exactly as simple and elegant as consequentialism, so this feature does not decide between them.⁴⁰ I think that it is a dubious honor to share one of the deeply attractive features of Agent-Neutral Orangeology. But even so, it is easy to see that ART does not share this honor, anyway. For ART is no more simple or elegant than the most listlike ordinary deontological views.

In fact, it is far worse. A paradigm deontological view like Clarke's or Ross's will consist, in part, of a list of wrong actions.⁴¹ These are the actions that agents are constrained not to do, even in order to prevent two or more others from doing the same thing. For example, in order to capture the result that every agent is under a constraint not to murder, a deontologist like Ross will have to put murder on the list of wrong actions. Ross's list will have as many items on it as there are actions that everyone is constrained not to do. In order to capture the same result, an ART-ist's list will have to contain this many items per agent. ART-ists will have to assume that Franz's murders are worse-relative-to Franz than Hans's and Jens's put together, but then it will also have to assume—separately, since this is simply a different fact—that Hans's murders are worse-relative-to Hans than Franz's and Jens's put together, and similarly for Jens. So far from being an elegant simple theory like consequen-

40. Although the independent plausibility of requisite assumptions about what is orange in order to get plausible results will certainly be enough to decide between these two views. My point is just that theoretical elegance and simplicity don't decide this.

41. Samuel Clarke, *A Discourse concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*, selections reprinted in D. D. Raphael, ed., *British Moralists 1650–1800* (1969; repr., Indianapolis: Hackett, 1991), 191–225. Also, W. D. Ross, *The Right and the Good*, ed. Philip Stratton-Lake (1931; repr., Oxford: Oxford University Press, 2002).

tialism aspires to be, ART appeals to the same lists as ordinary deontologists do, but it needs to appeal to one such list per agent.

And now we can turn Sen's remarks on their head and wonder, isn't it extraordinary that these lists have so much in common, if *good-relative-to* is just a relation that can either hold between a thing and an agent, or not?⁴² The similarity between the lists for each agent is now something that ART-ists will have to explain. Not that they will not be able to explain it—far from it. But the point is that if what is objectionable about ordinary deontology is that it resorts, at least at first pass, to lists, before it goes on to tell us what the items on the list have in common, Agent-Relative Teleology simply looks worse than ordinary deontology. Its lists are longer, and apparently more than a little bit redundant. For every item on the deontologists' list, the ART-ists need one item per agent. And in any case, there is a long and respectable tradition of deontological stories about what the items on the list have in common. There is no principled reason, short of covertly importing teleological assumptions, why the story that ART-ists can tell about what the items on their list have in common will be any more promising than the story told by the deontologists. So ART cannot even claim to have the advantages of simplicity or elegance over ordinary deontology.

C. *The Main Point*

If ART is in some way deeply attractive but Agent-Relative Orangeology is not, that must be because the *good-relative-to* relation is in some relevant way different from the *orange-relative-to* relation. That was the Preliminary Point from Section I, and it is the lesson to which we have returned. I think that if ART-ists are to successfully make out any deep attractions of ART whatsoever, they are going to have to tell us something about what this relevant way might be. They are going to have to tell us how the *good-relative-to* relation differs from the *orange-relative-to* relation, and why that difference is of the sort to make ART a more attractive ethical theory than Agent-Relative Orangeology.

A first thought might go something like this: okay, so maybe I'm right that 'good' does not express the *good-relative-to* relation in a context-dependent way, and maybe I'm right that that means that they have failed to substantiate the truth of the Compelling Idea. But if so, then the same goes for egoism. If I'm right, then egoists also do not believe that everyone ought always to do the action that will bring about the most good. But, the idea goes, egoism and consequentialism do have something deep in common that is the source of their common appeal, as demonstrated by Sidgwick:⁴³

42. Sen, "Well-Being, Agency, and Freedom," 214.

43. Henry Sidgwick, *The Methods of Ethics*, 7th ed. (Indianapolis: Hackett, 1907).

Consequentialism: For all x , x ought always to do that action that will bring about the most good.

Egoism: For all x , x ought always to do that action that will bring about the most good for x .

And whatever *that* is, the idea goes, it is surely a feature that is shared by Agent-Relative Teleology.⁴⁴

But that is where I get lost. *Good* and *good for*, after all, are concepts that I can understand. I can understand what is appealing about the thought that you shouldn't do something that will be worse for you—how stupid that would be! Irrational, frankly, since you'd be better off if you didn't. And I can understand what is appealing about the thought that you shouldn't do something that will be worse—how stupid that would be! Irrational, frankly, since things would be better if you didn't. I can almost talk myself into Sidgwick's trap, going back and forth between these two ways of thinking.

But since I don't understand what "good-relative-to" talk is all about, I don't understand how it could be appealing to think that you shouldn't do something that will be worse-relative-to you. I don't even understand what that means! Until the ART-ists give me some reason to think that the *good-relative-to* relation is somehow very much like the *good* property and the *good for* relation, I don't see why I should remotely find such an idea deeply compelling. So I think that the fact that egoism and consequentialism have both been sometimes found to be attractive views, and even the fact that Sidgwick found them both attractive, does not establish that ART must also be attractive in whatever way they are. First, I need some evidence that the *good-relative-to* relation is in the relevant way like the *good for* relation and the property of being *good*.

We can take the various putative attractions of consequentialism one by one, as I tried to do in Sections II and III of this article, and try to assess whether ART is ultimately able, after doing some work, to accommodate them. I've been arguing that these attractions don't come for free and that it is not obvious that they come at all. But the Main Point that I am trying to make is this: optimism that ART will be able to accommodate any of the attractions of consequentialism at all has to come from the idea that the *good-relative-to* relation is in some way importantly like *good*. I don't claim to have squashed all optimism that ART-ists will be able to retain some of the attractions of consequentialism, because I don't claim to have shown that the *good-relative-to* relation is not, in fact, importantly like *good*. I just claim to have shown how

44. Doug Portmore and Tom Hurka have both insisted on this in correspondence, and Portmore, "Consequentializing Moral Theories," takes up the idea further.

important it is to the ART-ists' research program, that this turns out to be the case. That is the Main Point.

D. Fitting Attitudes and ART

In response to an early version of the arguments presented in this article, Doug Portmore has constructed an explanation of what it is that *good-relative-to* has in common with *good*.⁴⁵ It is based on an understanding of the *good-relative-to* relation that has been advocated by Tom Hurka, is elaborated in recent work by Michael Smith, and a version of the idea that was spelled out two decades ago by J. L. A. Garcia.⁴⁶ The idea is that agent-relative value and agent-neutral value are both kinds of *value* and that this is confirmed by "fitting attitudes" or "buck-passing" accounts of value. I think that this idea is an important one, and addressing it adequately would require a paper on its own. But in this section I want to briefly explain why the most natural implementation of this idea cannot yield ART-ists what they need. Together, the fitting-attitudes analysis of agent-relative value and the fitting-attitudes analysis of agent-neutral value force the wrong predictions either about what is *good-relative-to* whom, or about what is agent-neutrally good, in order for ART-ists to be able both to accommodate constraints and to accept the assumptions about agent-neutral value that are required in order for constraints to generate the putative paradox of deontology in the first place.

The idea shared by Portmore, Hurka, and Smith is simple. On the fitting attitudes analysis of 'good', things are good when it is fitting to desire them. Fitting for whom? Well, fitting for everyone. So similarly, these theorists say, things are *good-relative-to* Jon when it is fitting *for Jon* to desire them. On this view, agent-relative and agent-neutral value have something in common—a similar structure. Both are analyzed in terms of fitting attitudes, but the *good-relative-to* relation is analyzed in terms of the attitudes that are fitting for some particular agent, while *good* is analyzed in terms of the attitudes that are fitting for everyone. So the view addresses the Main Point, from Section IV.C, because it gives an account of the *good-relative-to* relation which, together with the right account of *good*, explains why they are similar in some important way. And so consequently, it may license optimism that it would turn out that ART would be able to retain one or another of the important attractions of ordinary, agent-neutral, consequentialism (though it would not be enough, by itself, to show that ART did, in fact, retain any of those attractions).

45. Portmore, "Consequentializing Moral Theories."

46. Hurka, "Moore in the Middle"; Smith, "Two Kinds of Consequentialism"; and J. L. A. Garcia, "Agent-Relativity and the Theory of Value," *Mind* 95 (1986): 242–45.

The problem that I have with this idea is simple. It is not that I am opposed to fitting-attitudes analyses of 'good'. It is just that from some very straightforward and compelling assumptions, it follows that the fitting-attitudes analyses of agent-neutral and agent-relative value cannot *both* be true—at least if we are to make the assumptions about what is better than what, and what is better-than-what-relative-to whom, which ART-ists need.

The assumptions that I need are that 'good' is a gradable adjective and that gradable adjectives need to be analyzed in terms of their comparative form. As with 'tall', 'taller', and 'tallest', and 'fat', 'fatter', and 'fattest', it is the *better than* relation that is basic—something is good when it is better than sufficiently many things (perhaps in a contextually relevant comparison class), and it is best when it is better than everything (in some comparison class). So from this assumption, it follows that if a fitting-attitudes analysis of *good* is correct, it must really be a fitting-attitudes analysis of *better than*. And it is easy to construct such an account—A is better than B just in case it is fitting to prefer A to B. Fitting for whom? Well, for everyone. So to have the same structure as this account (which, after all, is the whole point of this idea), the fitting-attitudes analysis of agent-relative value would start with an account of *better-than-relative-to*. It would say that A is better-than B relative-to Jon just in case it is fitting for Jon to prefer A to B.

And now we have enough to generate a problem. For from these two accounts, it follows that A is better than B just in case A is better-than B relative-to everyone. For to be better than B is to be fittingly preferred by each, but being fittingly preferred by anyone is being better-relative-to her. So, in particular, it follows that if A is agent-neutrally better than B, then it is better-than B relative-to, say, Franz. But this conclusion makes it impossible to reconstruct what was supposed to be puzzling about constraints for ordinary consequentialism!

Constraints, recall, were cases like that in which Franz can murder to prevent Hans and Jens from murdering. They were supposed to be paradoxical, because it was supposed to be agreed on all sides that things are (agent-neutrally) worse if Franz does not murder—for then there are two murders, rather than one. Moreover, in order to accommodate constraints, ART-ists must suppose that things are better-relative-to Franz, if he does not murder. So in order to agree that there was a puzzle in the first place, and also have an answer to it, ART-ists need to suppose that constraints are cases in which it is agent-neutrally better for Franz to murder, but agent-relatively better for him not to murder. But from the fitting-attitudes analyses of both agent-relative and agent-neutral value, what we just derived was the thesis that if it is agent-neutrally better for Franz to murder than not murder, then it is better-relative-to Franz for Franz to murder than not murder. Which surely

has to be inconsistent with its being better-relative-to Franz for Franz to not murder than to murder.

The foregoing argument shows that ART-ists cannot have it both ways. They have to give something up. Revising our understanding of comparative adjectives in order to solve the problem sounds to me like a drastic step. But absent that, one of the following has got to go: the fitting-attitudes analysis of agent-relative value, the fitting-attitudes analysis of agent-neutral value, the ART-istic account of constraints, or the judgments about agent-neutral value which create the putative paradox about constraints, in the first place. But it is hard to see how an ART-ist who wants to advocate this response to my Main Point could be happy accepting any of these alternatives.

Again, I don't claim that this point is conclusive. As I said, I think that responses to the Main Point deserve careful consideration, and that there is a great deal more to be said about the one based on fitting-attitudes accounts of value.⁴⁷ But I do think that it shows that it is not at all obvious how ART-ists are to obtain what they need.

V. WHERE THIS LEAVES US

So here is where this leaves us: I haven't been trying to argue in this article that Agent-Relative Teleology is false. And I haven't even been trying to argue that there is no such thing as agent-relative value—the *good-relative-to* relation to which ART appeals (though I don't, in fact, think that there is any such thing deserving the name). I've just been trying to assess the straightforward and common idea that ART should be an attractive view for whatever reasons consequentialism is thought to be attractive. This common and unargued suggestion, I've been arguing, is deeply mistaken. Showing that it is true would require substantial work (as I showed in Sec. I), and this is work that has not been done, despite repeated iterations of this claim by ART-ists. Moreover, I've been assembling a case (in Secs. II and III) that it cannot be done at all, at least with respect to the most obvious attractions of consequentialism, and the ones that have been most vociferously claimed to be ART's main attractions. Though ART does have the right structure

47. For example, one important worry about the fitting-attitudes response on behalf of ART is that fitting attitudes are subject to constraints, just as ordinary actions are. Suppose that A is better than B, but that by preferring B to A, Franz can prevent Hans and Jens from preferring B to A, ensuring that they prefer A to B, instead. Does it follow that it is fitting for Franz to prefer B to A? Obviously not, if the fitting attitudes analysis is supposed to work. But if there is nothing puzzling about these constraints on which attitudes are fitting, such that it needs to be explained by positing some agent-relative kind of value, then why should there have been something puzzling about ordinary constraints on action, such that we needed to posit agent-relative value in order to explain them?

to accommodate constraints and special obligations, it cannot, as ordinary consequentialism can, appeal to the independent plausibility of the assumptions it needs about what is good-relative-to whom (Sec. II), and it cannot avoid the paradox of deontology (Sec. III). Finally, I've given reason to suspect that these arguments will generalize, and drawn out a Main Point which pointed in a direction worth further inquiry (Sec. IV).

Whether ART is attractive at all, and this was the Preliminary Point, must turn on some pretheoretical grasp that we have of the *good-relative-to* relation. If we have no such pretheoretical grasp of the relation, then ART has the same content as Agent-Relative Orangeology, which is not a deeply attractive view at all. As we saw, it does not even have the advantage of Agent-Neutral Orangeology of at least being simple and elegant. Moreover, even if it turned out that we do have some pretheoretical grasp of the *good-relative-to* relation, that would still not be enough to show that ART is any more attractive than Agent-Relative Orangeology, because it wouldn't be enough to show that *good-relative-to* is in any way like *good*. The Main Point was that ART-ists very much need it to turn out that this is so. And finally, even if the *good-relative-to* relation does turn out to be in important respects like *good*, that still doesn't license ART-ists' conclusion that their view retains all of the attractions of consequentialism—it only licenses limited optimism to hope that it will retain some such attractions. Each attraction must be earned in its own right, and some, like accommodating the Compelling Idea, may not be within the reach of ART at all.

Nothing about Agent-Relative Teleology as such requires, of course, that the *good-relative-to* relation to which it appeals be anything of which we have any antecedent grasp, or anything like *good*. It could be merely a theoretical posit, on the model of *electron*. So ART could well be true, for all that I have said. But unlike the case of *electron*, the appealing explanatory benefits that have been claimed for *good-relative-to* all hinge on misleading attempts to formulate the thesis of Agent-Relative Teleology in language that sounds like ordinary language—but is not. The dependence of Agent-Relative Teleologists on connections with ordinary language, therefore, is not mysterious or surprising after all. It is the entire source of interest in their view in the first place.