

Why Must We Treat Humanity With Respect? Evaluating the Regress Argument¹

Act so that you treat humanity, whether in your own person or in that of another always as an end and never as a means only.

-- Immanuel Kant (Kant 1990, p. 46/429)

The idea that our most basic duty is to treat each other with respect is one of the Enlightenment's greatest legacies and Kant is often thought to be one of its most powerful defenders. If Kant's project were successful then the lofty notion that humanity is always worthy of respect would be vindicated by pure practical reason. Further, this way of defending the ideal is supposed to reflect our autonomy, insofar as it is always one's *own* reason that demands that one treat humanity with respect. In this article, I consider what I take to be one of the most important and compelling attempts to defend the Kantian project. I draw the disappointing conclusion that this attempt does not succeed. The reasons this attempt fails shed some light on the difficulties facing any attempt to defend the Kantian project.

I. The Regress Argument.

Perhaps the most influential argument for the Enlightenment ideal is what is usually referred to as the "regress argument." The regress argument begins from the premise that rational agents can stand back from their impulses and ask whether they are worth acting upon. As rational agents we often take this reflective attitude, make choices on the basis of our reflection, and take those choices to be rational. Insofar as we take our choices to be rational, it seems that we also must take the objects of those choices to be good in the sense of being choiceworthy. Given that we take the objects of our rational

¹ Many thanks to Karen Jones, Rae Langton, Geoff Sayre-McCord and Michael Smith for helpful comments and encouragement.

choices are good, though, what should we suppose makes them good? The proponent of the regress argument suggests that the best explanation of the goodness of an object of an agent's rational choice is just that it *is* the object of rational choice. For apparently the only alternative explanations involve an untenable form of moral realism, according to which the objects of rational choice must be valuable apart from being valued. The proponent of the regress argument argues that this would be epistemologically mysterious² and metaphysically extravagant.³

One way of framing this stage of the argument is as posing a question very much like the question Socrates asked Euthyphro.⁴ Socrates' question was, "Are pious things pious because the gods love them, or do the gods love them because they are pious?" The proponent of the regress argument asks, "Are good things good because we rationally choose them, or do we rationally choose them because they are good?" The regress argument favors the first answer to this question on the grounds that the alternative answer commits one to a picture of moral properties that is metaphysically dubious and leaves us with no plausible epistemological story to tell about how we could come to know anything about such properties. Another Kantian motivation for this answer to the Euthyphro-like dilemma is the idea that the opposite answer would commit one to a form of heteronomy, insofar as the rational authority of morality would be imposed on the agent by moral facts "out there in the world." So long as we assume that rational motivation must in some sense be self-legislated and not externally imposed, this

²See Harman 1977, chapter one, for an argument in this vein.

³See Mackie 1977 for a classic presentation of this worry. Korsgaard seems to follow both Mackie and Harman in these respects, though she does not think Harman and Mackie's worries undermine what she calls "procedural realism," a view I briefly discuss in the text below.

⁴See Plato 1961b.

would give us reason to prefer the Kantian account of the sources of normativity. David

Cummiskey makes this point in pressing the regress argument:

In choosing to pursue an end, I am affirming that it is pursuit-worthy, that it is either instrumentally valuable or valuable as an end. But as we have already seen, the action cannot be practically justified by an external law, because then it would not provide a reason for me to do the action. I thus cannot become obliged simply by recognizing a natural or non-natural property. The determining ground must be an internal source. (Cummiskey 1996, p. 72)

Once we allow that we must take this horn of our Euthyphro-like dilemma, the Kantian argues that we can then infer that rational nature is an end in itself. Christine Korsgaard puts the point as follows:

Kant saw that we take things to be important because they are important to us - and he concluded that we must therefore take ourselves to be important. (Korsgaard 1996b, p. 122)

Elsewhere, Korsgaard puts the point slightly differently:

Thus, regressing upon the conditions, we find that the unconditioned condition of the goodness of anything is rational nature, or the power of rational choice. To play this role, however, rational nature must itself be something of unconditional value - an end in itself. (Korsgaard 1996d, p. 123)

The basic idea is that for anything to be the unconditioned condition of the value of everything else, it must be unconditionally good. If rational nature is this unconditioned condition, it would follow that it must be unconditionally good.

It will help to have a more formal reconstruction of the regress argument around which my discussion is structured:

- (1) In making a rational choice, a rational agent rightly judges the object of her choice to be good. [an analysis of what it is to make a rational choice]
- (2) The best explanation of the object of a rational agent's rational choice's being good is simply that it *is* an object of a rational choice. [for the alternative explanation involves an epistemologically and metaphysically problematic form of moral realism]
- ∴ (3) The object of a rational agent's rational choice is good simply because it *is* the objects of a rational choice. [from 1 & 2, by way of an inference to the best explanation]
- (4) Anything that is good is either the object of a rational choice or is rational agency itself. [follows from the rejection of realism that underwrites step (2) of the argument]

.∴ (5) Rational agency is the unconditioned condition of the goodness of anything else.
[from 3 & 4 via conceptual analysis of what it is for something to be the unconditioned
condition of the goodness of anything else]
(6) For anything to be the unconditioned condition of the goodness of anything else, that
thing itself must be unconditionally good. [(?)]
.∴ (7) Rational agency is unconditionally good. [from 5 & 6]

The argument is intuitively powerful. It is plausible to suppose that if things are valuable
then they must be valuable because they are appropriately valued. It would, however,
seem odd to suppose that the appropriate valuations of rational agents underwrite the
value of everything which is valuable, without rational agency itself being valuable. If
rational agency is the condition of the value of everything else, then it seems plausible to
infer that rational agency is itself good, and unconditionally so.

II. The Euthyphro Problem Revisited.

In the first stage of the regress argument (steps (1) through (2)), premise (2) is clearly the
key premise. For once 'good' in (1) is read as simply as having roughly the same
meaning as 'choiceworthy' that premise should be fairly uncontroversial. Whereas (2)
seems to be a rather controversial philosophical thesis about the metaphysics of
choiceworthiness. Recall that (2) asserts that the object of a rational choice is good
because it is rationally chosen rather than the other way around. The standard defense of
(2) assumes that anyone who takes the other horn of this Euthyphro-like dilemma must
embrace *some* form of problematic moral realism. I argue that this is not the case; we can
embrace the other horn of the Euthyphro-like dilemma without embracing moral realism.

First, we must get a somewhat more clear idea of what 'realism' and its rivals
claim. 'Moral realism' and its cognates are terms of art that are used very differently by
different philosophers so we should see just what the proponent of the regress argument
has in mind. Korsgaard is admirably clear on this point:

The realist's response is to dig in his heels. The notion of normativity or authority
is an irreducible one. It is a mistake to try to explain it. Obligation is simply
there, part of the nature of things. (Korsgaard 1996b, p. 30)

Moral realism on this account seems to be the thesis that moral properties (like the property of being obligated to perform a given action) are irreducible ones. So, in particular, moral properties are not reducible to any natural properties. Moral realism in this sense is therefore a species of G.E. Moore's famous doctrine of non-naturalism (though Moore focused especially on the notion of being good as opposed to the notion of being obligated). Indeed, Korsgaard cites Moore (as well as Samuel Clarke) as a paradigm instance of a realist view (see Korsgaard 1996b, p. 33)

The defender of the regress argument claims that if one rejects premise (2) of that argument then one is committed to moral realism in this sense. To assess this claim, we must first see what the alternatives are to moral realism. There is at least one other option worth considering here: what Simon Blackburn has termed "quasi-realism." The quasi-realist holds that when we are doing serious metaphysics we should not suppose that there really are moral properties out there in the world. The quasi-realist shares the Kantian's worries about non-naturalism (moral realism, in Korsgaard's terms) and argues that naturalism fares no better for a variety of reasons. Following more traditional non-cognitivists like Hare and Stevenson, the quasi-realist holds that the primary function of moral utterances is to express our pro- and con-attitudes rather than to describe the way the world is. The point is often put in terms of the "direction of fit" of evaluative judgments. While there are many different ways of cashing out the "direction of fit" metaphor, the basic idea is that beliefs are made "to fit the world," whereas pro-attitudes are made "to make the world fit them."⁵ If I believe that p but p is not the case, then it is my belief which needs to be changed. Whereas if I desire that p but p is not the case, it is the world that needs to be changed. I shall use 'pro-attitude' to refer to any state with the direction of fit of desire.⁶ With this explanation of our moral practice in hand, the quasi-

⁵ For a much more detailed investigation of the notion of a direction of fit, see Humberstone 1992 and Smith 1987. The classic treatment can be found in Anscombe 1957.

⁶ The choice is somewhat stipulative in part because one could just as easily have used the term 'con-attitude' here. For there may be no important difference between having a pro-attitude toward p and having

realist goes on to show how we nonetheless can “earn the right” to the realist-sounding discourse of ordinary language on a metaphysically slender basis.⁷

We are now in a position to see where the case for premise (2) of the regress argument goes wrong. For advocates of quasi-realism can consistently say things like, “The objects of our rational choice are good independently of our choice, and we rationally choose them because of their goodness.” For on the quasi-realist account, such claims are themselves first-order claims and are to be understood as the expression of certain attitudes. They might, for example, gloss such a claim roughly as the expression of a suitable pro-attitude toward norms holding that certain options are to be chosen independently of whether anyone actually has chosen them.⁸ Or perhaps they would gloss such a claim as the expression of a pro-attitude toward the promotion of certain ends regardless of whether anyone has a preference for the promotion of those ends. These may not be the only ways in which a quasi-realist might gloss such claims, but they are representative. After all, a chief virtue of quasi-realism is supposed to be its ability to mimic realist talk without taking on any of the difficult metaphysical and epistemological burdens that a full-blooded realist must embrace. Whereas the defense of the regress argument's premise (2) depends on the *only* alternative explanation of thing's value being a problematic form of realism. So if the defender of the regress argument wants to maintain this then she must explain why the quasi-realist is mistaken in thinking that they can say all these things.⁹ As it stands, this is an argumentative burden that defenders of the regress argument have not explicitly tried to discharge.

a con-attitude toward not-p. Since ‘pro-attitude’ is the more usual umbrella term for states with this direction of fit, I simply follow this orthodoxy.

⁷Simon Blackburn is the primary defender of quasi-realism. See Blackburn 1984 and Blackburn 1993.

⁸Here, for example, is Allan Gibbard's discussion of an anorexic who prefers her starvation to a figure plump enough to sustain life: “We do not think what she is doing is irrational simply in virtue of our own tastes or preferences in the matter. Nor do we think it is irrational in virtue of the norms we accept. The norms we accept prohibit starving for a trim figure, regardless of what one prefers or what norms ones accept.” (Gibbard 1990, p. 166) See also Blackburn 1993.

⁹It is worth noting a contrast with subjectivist forms of cognitivism like the views defended by John McDowell and David Wiggins here. On their views, values are mind-dependent in an important sense. This is one of a number of important respects in which quasi-realist expressivism differs from subjectivist

It might be objected that if quasi-realism were true then moral rules would be heteronomous in a way that Kantians suppose is problematic. For it might seem that on the quasi-realist view our moral judgments are externally imposed on us. After all, the quasi-realist apparently supposes that our inclinations are our moral judgments, and Kant famously thought that our subjective inclinations are not part of our “true self” unless we rationally endorse them, thereby incorporating them into our will. This objection relies on the potentially problematic assumption that there is a deep distinction to be drawn between our inclinations and our will. Indeed, most quasi-realists are sympathetic to views in action theory that would naturally lead them to question this distinction. However, even if we grant the Kantian that such a distinction can and should be drawn the objection does not seem to work. We might also grant the seemingly implausible Kantian assumption that our inclinations are external to us in some problematic way unless we endorse them as worthy of satisfaction and thereby incorporate them into a maxim and grant the legitimacy of concerns about heteronomy (though it seems unlikely that anyone who denies those theses is thereby irrational). The crucial point is that the quasi-realist position, properly understood, is consistent with all of this. For at least one possible version of quasi-realism would be the view that moral judgments are commitments of the will. On this view, when I say “X is wrong,” I express a commitment of my will to avoid actions like X, rather than a mere sentiment, inclination, or desire not to X. So long as the will has the same “direction of fit” as a desire, rather than the direction of fit of a belief, this will not be inconsistent with quasi-realism, broadly construed. For quasi-realism's primary claim is that moral utterances serve to express pro- and con-attitudes rather than beliefs, where the defining mark of such attitudes is their direction of fit. So long as the Kantian will has the right direction of fit, it will qualify in this very broad sense as a pro- or con-attitude. Moreover, it is hard to

theories, although these theories are often run together. Thanks to an anonymous referee for drawing me out on this point.

see how anything recognizably described as “the will” could fail to have this direction of fit. On any plausible account, willing is fundamentally more like desiring than it is like believing even if it is logically distinct from both.

Alternatively, it might be objected that the quasi-realist's reading of premise (2) of the regress argument gives it a first-order reading whereas it should be given a metaphysical or meta-ethical reading. Indeed, insofar as rationalist arguments in general go from meta-ethical premises to first-order conclusions, in evaluating such arguments one must be vigilant for equivocations between meta-ethical and first-order readings. If (2) is glossed as making the metaphysical claim that value judgments are in some metaphysically weighty sense “made true by” rational-choice facts, then the quasi-realist might simply deny that (2) has such a reading in ordinary language and argue that the attempt to give philosophical sense to it in a way that will help the rationalist is unlikely to succeed. However, the crucial point is what the quasi-realist should say about (2) *if* it is granted that it has an interesting meta-ethical reading. For clearly the quasi-realist must deny (2) thus understood. They must deny (2) on this reading because their quasi-realism commits them to holding that there are no “truth-makers” for value judgments at all – our moral judgments do not represent a way the world is or might be, given their direction of fit, and hence are not the sorts of judgments that have truth-makers. Similarly, if (2) makes the meta-ethical claim that 'good' means 'rationally chosen', then a quasi-realist need not accept (2), for quasi-realism entails no such analysis. So *regardless* of whether we read (2) as (i) a first-order claim about the conditions under which things are good, or (ii) a metaphysical claim about truth-makers of value judgments, or (iii) a meta-ethical claim about the meaning of 'good', quasi-realists can happily reject (2), for all that has been said so far.

Further, it is unclear how much room someone like Korsgaard, the most outspoken advocate of the regress argument, has for criticism of quasi-realism. For her own view seems to *be* a form of quasi-realism, though she would no doubt resist this

characterization. Korsgaard endorses the stock criticisms of realism but also clearly does not want to embrace naturalism (she seems to think that something like Moore's Open Question Argument against naturalism is sound), apparently leaving her with quasi-realism by default. One might look to Korsgaard's distinction between procedural and substantive realism for some guidance as to whether she is a quasi-realist, as she accepts procedural realism but rejects substantive realism. Here is her account of that distinction:

The procedural moral realist thinks that there are answers to moral questions *because* there are correct procedures for arriving at them. But the substantive moral realist thinks that there are correct procedures for answering moral questions *because* there are moral truths or facts which exist independently of those procedures, and which those procedures track. (Korsgaard 1996b, p. 36-37)

The problem with this account is that it simply pushes the question back a stage. For now the obvious question for the procedural realist is what makes it true that one procedure is correct. Indeed, one might ask what it *means* to say that one procedure for answering moral questions is "correct." If the proceduralist strategy is redeployed to answer these questions, then an infinite regress threatens. If it is not redeployed then the only other alternative to quasi-realism seems to be substantive realism at one remove.

So the first problem with premise (2) of the argument is that it seems simply to overlook the quasi-realist option - an odd oversight since in many respects quasi-realism is a view that Kantians should find attractive. The Kantian defense of (2) faces another problem, though. In rejecting the realist account of value in (2), the Kantian is committed to supposing that nothing could be valuable apart from its being rationally valued. For according to the argument to suppose otherwise is to land oneself in the epistemologically and metaphysically murky waters of moral realism. However, at the end of the day, it might seem that the Kantian does want to say *just that*, for the Kantian seems to suppose that humanity would be valuable even if nobody actually valued it. For humanity's value is supposed to be unconditional, where something's value being unconditional consists in its being of value in all possible circumstances. This, however,

seems to entail that humanity is valuable quite apart from any agent's rationally making it their end, insofar as rationally being made an end is presumably part of the circumstances in which humanity is sometimes contingently found. This seems to commit the Kantian to supposing that something can be valuable apart from its *actually* being rationally valued, though, *contra* the Kantian defense of (2). So the Kantian's own position may face the same objections as those pressed against the moral realist. To be consistent, the Kantian must say that rational agency is valuable *only* if it is the object of rational choice, but this abandons the idea that it is good *unconditionally*. The worry is structurally similar to a worry that faces some popular versions of the cosmological argument for the existence of God. Just as some versions of the cosmological argument rely on the premise that everything must have a cause to the conclusion that there must exist a first cause that is itself uncaused, the regress argument seems to go from the premise that everything that is valuable must be valuable only because some condition is met (being the object of rational choice) to the conclusion that something must be valuable unconditionally. In both cases, the argument's conclusion is inconsistent with its main premise. Something has gone badly wrong.

Korsgaard has addressed this point in a symposium on her work in *Ethics*:

...in some of the essays in *Creating the Kingdom of Ends*, I talk about the value of humanity as if it were a form of intrinsic value, distinguishing the value of humanity in this way from the value of our particular ends...More recently, especially in *The Sources of Normativity*, I have come to think that the value we place even on ourselves is conferred...there is a continuity between the value of humanity and the value of other things: they are all the result of our own acts of conferring value. (Korsgaard 1998, p. 63)

Perhaps this is compatible with the idea that humanity's value is unconditional after all. For the idea may be that though the value of humanity is conferred it is not conditional because it is *impossible* for a rational agent to fail to value humanity. Actually, this still seems to make humanity's value conditional but such that the condition on its value necessarily is satisfied whenever it exists. Let us put this worry to one side, though. For

this move would at least allow Korsgaard to claim that humanity is valuable in all possible circumstances, which is impressive and important enough. For on this account, wherever one finds humanity, one would find humanity having rationally been made its own end. However, the claim that it is literally impossible to fail to value humanity, would require a further argument. The regress argument at most establishes that it is *irrational* not to value one's own humanity, not that it is *literally impossible*. For not valuing one's own humanity intuitively does seem to be possible, and at least some Kantians would not want to deny this.¹⁰ Indeed, Kantians have done an especially good job of emphasizing the importance of self-respect. Since failures of self-respect are most plausibly understood as failures to value one's own humanity, the Kantian should be the last person to hold that it is literally impossible not to value humanity. In any event (so far as I know) the advocates of the regress argument have provided no further argument for this impossibility claim. Indeed, if such an argument could be given then the regress argument itself would be otiose. For in that case we could just appeal to that separate argument to show that insofar as we are rationally agents we literally cannot help but value humanity.

III. Moorean Organic Unities.

Whatever one thinks of the first stage of the argument, the second stage is more clearly problematic. Let us for the sake of argument grant (4). We may even grant that (5) follows from (3) and (4), so long as all we mean by that is that the thing, X, is itself a condition of the goodness of everything else and that there is no *other* condition that X must meet to serve as a condition for the goodness of everything else. The key remaining step of the argument is premise (6), which claims that if something is the unconditioned condition of the goodness of everything else, then it must itself be unconditionally good. In reconstructing the regress argument, I have flagged (6)'s justification with a question-

¹⁰For example, Alan Donagan, at some points does seem to think a Kantian can allow for the possibility (though not the rationality) of complete indifference toward humanity. See Donagan 1977.

mark. For though (6) is clearly the main premise of the second stage of the argument, it is obscure what its justification is. Though it sounds plausible, it is not as if it is self-evident. One can perfectly well understand what (6) says and still be unsure of its truth. Why should one suppose that simply because X is the condition for the goodness of everything else that X itself must be good *at all*, much less unconditionally good? G. E. Moore's famous discussion of "organic unities" makes it clear that one need not draw such an inference (see Moore 1903). As his discussion illustrates, one can consistently hold that something is part of a larger organic unity that is good, and be essential to that unity's being good, without itself being good *at all*. This is a limiting case of the more general point that the goodness of an organic unity is not reducible to the sum of the goodness of its parts.¹¹ (6) implicitly rejects this possibility, but it is unclear why we should reject it.

Consider, by way of analogy, the view that the unconditioned condition of the goodness of everything is being the object of an inclination. On this view, for anything other than an inclination that is good, it is good only insofar as it is the object of an inclination. If we were to accept (6), we would have to infer from this that inclinations themselves must be unconditionally good, but this consequence just does not follow. It is perfectly consistent for someone to hold the view that so long as someone actually does have an inclination for something, that thing is ipso facto good, but also to hold that inclinations themselves are not only not good, but positively bad. Indeed, some of the Stoics seemed to have held a view very much like this. Ironically, Kant himself seems to have held a view close to this, remarking that, "The inclinations themselves as the source of needs, however, are so lacking in absolute worth that the universal wish of every rational being must be indeed to free himself completely from them" (Kant 1990, p. 45/428). It does not seem that my supposition that something is good only on the

¹¹ Even critics of Moore's more general thesis grant the possibility of such limiting cases as an uninteresting (even "trivial"). See, for example, Zimmerman 1999.

condition that somebody has an inclination for it commits me to supposing that inclinations themselves are good at all, much less unconditionally good. For all that has been said so far, I might hold either that the inclinations themselves are indifferent, or even bad, in themselves. Similarly, it is coherent to suppose that rational agency might be the unconditioned condition of the value of everything else, but itself not have unconditional value. So being the unconditioned condition of the goodness of anything else does not require unconditional value; the alternative is intelligible and even plausible in some cases.

It is perhaps no surprise that Korsgaard criticizes Moore's theory of organic unities (though she never considers the preceding objection itself). In particular, she argues that, "Moore's view, and the intuitionistic method of isolation, veil or obscure the internal relation within the organic unity in virtue of which the organic unity has its value." (Korsgaard 1996g, p. 271) This may be a serious problem for Moore, given his particular characterization of intrinsic value as non-relational in its character and his famous attachment to an "isolation test" for such value.¹² On its face, though, this is no reason to deny the possibility of organic unities, but is rather a complaint that Moore's view obscures the internal relations of those unities. Let us grant that Moore himself could not explain the internal relations in virtue of which an organic unity (or its components) has its (or their) value. It would in no way follow that such an account cannot be given if we reject Moore's account of intrinsic value and the associated "isolation test." So long as we understand the notion of an organic unity as the idea of a whole whose value is not reducible to the sum of the value of its parts, we can hold onto that notion without Moore's other views. Only if we unnecessarily "load up" the concept of an organic unity with lots of Moorean assumptions that are actually inessential to it

¹² Very roughly, according to Moore's isolation test, one determines whether something has intrinsic value by imagining a universe containing nothing but it with all its intrinsic properties intact, and then seeing whether it seems like there is anything of value in that universe - whether it would be rational to prefer it to a universe without the thing in question. See Moore 1903.

would worry about those assumptions translate into worries about the very idea of an organic unity. For all that has been said so far, one could reject the Moorean baggage Korsgaard and others have attacked (i.e., his theory of intrinsic value and associated "isolation test," as well as his non-naturalist realism and intuitionist epistemology) and still suppose that one thing (A) might be valuable *in virtue of* its relation to something else (B), even where this something else (A) is not itself unconditionally valuable. Different philosophers may explain the relevant "in virtue of" relation in different ways, but there is no apparent reason to suppose that any of the plausible specifications of the "in virtue of" notion would preclude the sorts of value structures Korsgaard rules out. Indeed, the *ex ante* denial of the possibility of such value structure may itself, "veil or obscure the internal relations within the organic unity in virtue of which the organic unity has its value." (Korsgaard 1996g, p. 271) Whereas Moore may not make room for claims about the importance of the internal relations of an organic unity, the Kantian view seems to suppose without argument that such internal relations cannot take certain forms.

IV. **Agent-Relativity, Agent-Neutrality, and Practical Identity.**

Recall that the regress argument is meant to show that an agent must treat humanity, *whether in his own person or that of another*, as an end in itself. However, the argument as I have reconstructed it seems only to support the conclusion that humanity has value of some sort, but leaves it open whether that value is agent-relative or agent-neutral. By speaking of a value as agent-relative I mean to indicate the principle underwriting that value makes ineliminable, non-trivial, pronominal back-reference to the agent for whom it is valuable, whereas a value is agent-neutral if that principle involves no such back-reference. If, for example, the principle underwriting the value of pleasure is of the form, "An agent's pleasure is always valuable *for that agent*," then the value of pleasure is agent-relative. If, by contrast, the principle in question were simply, "An agent's pleasure is always valuable," then the value of pleasure is agent-neutral. The former principle gives an agent reason to promote pleasure only if the pleasure is her own but the

latter principle provides an agent with reason to promote pleasure no matter who whose it is. For all that has been said so far, the regress argument might be taken to show that each agent must value his or her own humanity, but not show that one must also value the humanity of *others*. For one might take the lesson of the argument to this point to be that insofar as I view something as good-to-me that I must do so on the basis of my rationally making it my end, and that I must therefore see my own rationality as good-to-me too. One might allow that this argument goes through for every possible rational agent, and thereby respect a kind of universalizability. Just as I must see my humanity as good-to-me, you must see your humanity as good-to-you, etc. Still, the sceptic might insist, none of this delivers the conclusion that each of must see humanity, *whether in one's own person or that of another*, as always good as an end, simpliciter. The value of humanity might be purely agent-relative.

Korsgaard is well aware of this concern. She argues (in Korsgaard 1996f) from a doctrine that she refers to as "intersubjectivism" to the conclusion that reasons for acting are not agent-relative. As I understand her reply, it comes in two stages. First, she appeals to Wittgensteinian considerations to argue for the view she calls "intersubjectivism." (Korsgaard 1996f, p. 278)¹³ Second, she argues that intersubjectivism entails that all reasons must be reasons we can share, and so cannot be agent-relative. One might naturally infer from this that Korsgaard holds the view that reasons are agent-neutral, given that the distinction between agent-relative and agent-neutral reasons is exhaustive. On the other hand, it might also seem a little perverse to interpret Korsgaard as arguing that all reasons are agent-neutral given that the subtitle of her paper is, "an attack on the distinction between agent-relative and agent-neutral values." Actually, there seems to be a contradiction at the heart of Korsgaard's paper.

¹³Here I am piecing together Korsgaard's arguments in *The Sources of Normativity* (Korsgaard 1996b) and the argument of "The Reasons We Can Share: An attack on the agent-relative/agent-neutral value distinction." (Korsgaard 1996f).

On the one hand, she characterizes the conclusion of her argument as the thesis that, "values are neither subjective nor objective, but rather are *intersubjective*." (Korsgaard 1996f, p. 276) What, though, is meant by 'subjective' and 'objective'?

In contemporary jargon, the question is whether reasons and values should be understood to be agent-relative or agent-neutral, or whether reasons and values of both kinds exist. In slightly older terms, the question is whether reasons and values are subjective, existing only in relation to individuals, or objective, there for everyone. (Korsgaard 1996f, pp. 275-276)

This passage, combined with her characterization of her thesis as the doctrine that "values are neither subjective nor objective, but rather are *intersubjective*," commits Korsgaard to arguing that the agent-relative/agent-neutral distinction is not exhaustive. On the other hand, Korsgaard also explicitly allows that the distinction *is* exhaustive:

James Dreier has pointed out to me that in styling my project [sic.] an attack on the distinction between relative/subjective and neutral/objective, I might give the impression that I think this logical distinction is not logically exhaustive, which it obviously is. (Korsgaard 1996f, p. 303, n. 6)

In fact, Korsgaard not only "gives the impression" that she thinks the distinction is not exhaustive, she commits herself to arguing that it is not, as her subtitle would suggest when she goes on to hold that 'values are neither subjective nor objective'. A more charitable interpretation might insist that Korsgaard is using 'subjective' and 'objective' in a different sense than 'agent-relative' and 'agent-neutral' when she claims that values are neither subjective nor objective, but she never actually explicitly glosses 'subjective' and 'objective' in any way other than as being 'older terms' for the distinction marked in contemporary jargon by 'agent-relative' and 'agent-neutral'. So the discussion of Dreier's point seems to vitiate the intended conclusion of the argument on its most obvious reading. Moreover, Dreier is right that the relative/neutral distinction is an exhaustive one; for any reason, either its articulation involves the relevant sort of

pronominal back-reference or it does not. It seems that Korsgaard really should be arguing that the reasons to respect humanity are agent-neutral rather than arguing that the agent-relative/agent-neutral distinction is not exhaustive. Indeed, at one point she comes very close to endorsing this reading herself, commenting that,

It might be thought that I am defending this position: that persons have agent-neutral value, while all other values are agent-relative. And then I add that you express your sense of the neutral values of others by sharing in their agent-relative ends. This *is* close to the Kantian position, but it is a misleading way to put it. It makes the value of persons a metaphysical reality, perhaps in need of a metaphysical defense... (Korsgaard 1996f, pp. 300-301)

The reason Korsgaard finds this gloss on her view misleading is that the agent-neutrality of the value of persons ‘makes the value of persons a metaphysical reality’ but this does not actually follow from Nagel’s official gloss of what it is for a reason to be agent-neutral, which Korsgaard endorses. Even a quasi-realist like Blackburn can allow that reasons are agent-neutral in Nagel’s sense, as that will just be a first-order remark about whether reasons for one agent entail reasons for others. So nothing metaphysical *follows* from the agent-neutrality of the value of persons. Korsgaard’s discussion seems to equivocate between Nagel’s official gloss of agent-neutrality and a more objectionably metaphysical conception which she finds implicit in Nagel’s substantive discussion of agent-neutrality. That these two notions of agent-neutrality are both in play in her discussion is also suggested by her reply to Dreier, in which she goes on to remark that, “My quarrel...is really with Nagel’s account of the source of these reasons,, which suggests that values and reasons originate either from personal, idiosyncratic desires or from metaphysical realities of some kind.” (Korsgaard 1996f, p. 303, n. 6) If this is right then the issue becomes why Korsgaard’s critique of these accounts of the sources of values and reasons would, if successful, entail that humanity is valuable has agent-neutral value

in a sense which is not based on any problematic metaphysics, which seems to be her own view. Carefully distinguishing these different notions of agent-neutrality and agent-relativity and explaining how they are related to one another would have perhaps brought out some of the suppressed premises in Korsgaard's argument. In any event, Korsgaard is committed to arguing that the reasons to respect humanity are *not* agent-relative in any of the senses she canvasses, and my objection to her larger argument is that she does not successfully defend this negative conclusion. We can therefore avoid the question of just what positive characterization of such reasons Korsgaard should be defending.

Let us grant that Wittgensteinian considerations establish intersubjectivism in Korsgaard's sense. On its face, this still leaves open the question of agent-relativity. For intersubjectivism is an interpretation of what non-agent-relative reasons would be like if there are any. Intersubjectivism is defined via a contrast with a view Korsgaard calls "*Objective Realism*," according to which "values would exist in a world devoid of creatures who see and respond to reasons." (Korsgaard 1996f, p. 278) The intersubjectivist account, by contrast, interprets those values as neutral with respect to the identity of individual agents, but not as existing independently of all agents as such.

Korsgaard's strategy seems to be to survey all the plausible sources of agent-relativity and argue that, in light of intersubjectivism, each fails to support agent-relativity. Intersubjectivism apparently is the doctrine that each of us can always, in principle, make our values intelligible to others. Suppose, for example, that I have rationally made it my end to climb a mountain. Nagel takes this as a paradigmatic kind of agent-relative value, insofar as I have reason to promote my climbing the mountain, but nobody else has any reason to promote my doing so *simply in virtue of the fact that I*

have made it my end. On Korsgaard's account, it seems that if I am rational, I must take it that mountain-climbing has value which anyone could appreciate, and simply wish to be the one who instantiates it. To think otherwise would run afoul of intersubjectivism, for it would seem to entail that my values could not be made intelligible to others. Here is Korsgaard:

There are reasons for caring about these things, reasons which are communicable and therefore at least potentially sharable. Ask a mountain climber why she climbs and she need not be mute: she may tell you things about the enlarged vistas, the struggle with the elements, the challenge about overcoming fears or surpassing physical limitations. She takes her desire to climb mountains to be a motivated desire, motivated by recognizably good features of the experience of climbing. She does not take the value of the climb to be conferred on it simply by her desire to do it. Someone who says, 'I *just* want to' is not offering you a reason; he is setting up a bulwark against incomprehension. (Korsgaard 1996f, p. 289)

One problem I shall here put to one side is that this passage seems to imply the opposite answer to the Euthyphro question to the one given in the first stage of the regress argument. For the natural reading of this passage would be that the mountain climber rationally chooses her project because of it is independently good (in virtue of the scenic vistas, the struggle with the elements, etc.) rather than its being good because the climber rationally chooses it. Let us put this *prima facie* tension to one side, though. Korsgaard in the following way: (a) intersubjectivism entails that all reasons must be communicable, (b) a reason is communicable only if it is sharable, and (c) an agent-relative reason is not sharable. Therefore, she concludes, reasons for acting are not agent-relative.

The first move is plausible enough, but the latter stages are problematic. For mere communicability does not entail sharability in any sense in which sharability would entail non-agent-relativity. Granted, if you have reason to promote your climbing a mountain then you should be able to tell me what that reason is and even tell me why it counts as a reason for you to do it. All of this seems consistent with my still supposing that I

nonetheless have *no* reason to promote your climbing the mountain. For your articulation of your reason and why you have it might, for all that has been said so far, be purely agent-relative. You might even explicitly invoke some agent-relative theory (egoism, say) in explaining why you have the reason you do, in which case I need not see your reasons as generating reasons for me, as I might be forced to do if the reasons were not agent-relative. I might even fully appreciate your reason, and would aim to act on such a reason if I ever found myself in circumstances just like your own. We can accept intersubjectivism without abandoning agent-relativity.

Following Korsgaard, we might at this stage try to supplement the regress argument by appealing to the idea of a "practical identity." The main idea here is that insofar as we are rational each of us must have a practical identity, where a practical identity is a description under which one values oneself. However, Korsgaard argues, if we are to have any practical identity at all, at least part of that identity must be as human beings, and from this it is supposed to follow that we must value humanity, at least in our own person. Korsgaard supports this latter claim by way of emphasizing that our ability to reflect on our ends and ask whether they really provide us with reasons is essential to our identity as rational agents. If we are wondering about what to do, we must initially reject nihilism. As Korsgaard puts it, if we are not to be complete normative sceptics, we must think it is possible for some of our actions to be worth doing. It is somewhat tendentious to put this point in terms of each of us necessarily having a practical identity, where a practical identity is a conception of oneself under which one values oneself. For the latter characterization suggests that one values *oneself* as an end, but that does not follow from any platitudes about what follows from the rejection of nihilism. I might, e.g., reject nihilism and be a hedonistic utilitarian, and value only pleasure rather than people. Still, in an attenuated and stipulative sense we might grant that even the hedonistic utilitarian has a conception of himself under which he values himself, though strictly instrumentally – namely, as a potential promoter of happiness.

The next step in the argument is to establish that we must take our humanity as part of our practical identity in this sense. Here Korsgaard's idea seems to be that our identity as rational agents is not contingent. There is an important truth here. It is true that to have a practical identity, in the relevant sense, is to value oneself under some description, and to do this one must be capable of valuing. Insofar as it is plausible to suppose that only rational agents are capable of evaluative judgments, it is plausible that only rational agents can have practical identities. However, it would be a mistake to infer directly from this that one must value one's rational agency. Here it is important to note that talk of one's 'practical identity' can be ambiguous. On our canonical interpretation, a practical identity is a conception of an agent under which the agent values herself. Another related but distinct thing we might mean by talk of a 'practical identity' would be a conception of oneself *that makes it possible for one to value anything at all*. The preceding line of thought reveals the plausibility of supposing humanity is part of one's practical identity in the second sense, but it does not follow from this that one's humanity is part of one's practical identity in the first sense. There seems to be a fallacy of equivocation between two senses of 'practical identity'.

Or perhaps there is no fallacy, and Korsgaard is relying on the suppressed premise that anything that is part of one's practical identity in the second sense must also be part of one's practical identity in the first sense. In this case, the suppressed premise stands in need of defense. *Prima facie*, it seems perfectly coherent to regard a feature as essential to one's being able to value things without concluding that the feature in question is itself valuable. Perhaps Korsgaard would argue that we must accept this relationship because we must come down a certain way on the Euthyphro problem. Since we must see things as valuable because we value them, rather than the other way around, we must "see ourselves as important," and so must value our rational agency. It should be clear from the preceding sections why this sort of move is problematic. The appeal to practical identity faces precisely the same kinds of objections facing the regress argument. Nor

should this come as a surprise, given that the appeal to practical identity is “just a fancy new model of an argument that first appeared in a much simpler form, Kant’s argument for his Formula of Humanity.” (Korsgaard 1996b, p. 122)

Conclusion.

Sadly, the regress argument cannot vindicate the Kantian project. At each stage the argument faces serious objections, and what seem to be the Kantian’s best replies to these objections are not compelling. Nor does supplementing the regress argument with an appeal to the idea of a "practical identity" seem promising. If Enlightenment morality is true, its truth is not demonstrated by the regress argument.

Bibliography

- Anscombe, E. (1957), *Intention*, Oxford: Basil Blackwell.
- Blackburn, S. (1984), *Spreading the Word*, Oxford: Oxford University Press.
_____. (1993), *Essays in Quasi-Realism*, Oxford: Oxford University Press.
- Brink, D. (1989), *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
- Cummiskey, D. (1996), *Kantian Consequentialism*, New York: Oxford University Press.
- Donagan, A. (1977), *The Theory of Morality*, Chicago: University of Chicago Press.
- Gibbard, A. (1990), *Wise Choices, Apt Feelings*, Oxford: Clarendon Press.
_____. (1999), "Morality as Consistency in Living." *Ethics*.
- Humberstone, L. (1992), "Direction of Fit", *Mind*, pp. 59-83.
- Kant, I. (1990), *The Groundwork of the Metaphysics of Morals*, Translated into English by Lewis White Beck. New York: Macmillan Publishing Company.
- Korsgaard, C. (1996a), *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.
_____. (1996b), *The Sources of Normativity*. Cambridge: Cambridge University Press.
_____. (1996c), "Morality as Freedom", in Korsgaard 1996a, pp. 159-187.
_____. (1996d), "Kant's Formula of Humanity", in Korsgaard 1996a, pp.106-132.
_____. (1996e), "Kant's Formula of Universal Law", in Korsgaard 1996a, pp. 77-105.

- _____. (1996f), "The Reasons We Can Share: An Attack on the Distinction Between Agent-Relative and Agent-Neutral Values", in Korsgaard 1996a, pp. 275-310.
- _____. (1996g), "Skepticism About Practical Reason." In Korsgaard 1996a, pp. 311-334.
- _____. (1998), "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind", *Ethics*, 109, pp. 49-66.
- Moore, G.E. (1903), *Principia Ethica*, Cambridge: Cambridge University Press.
- Nagel, T. (1970), *The Possibility of Altruism*, Princeton: Princeton University Press.
- O'Neill, O. (1989), *Constructions of Reason*. Cambridge: Cambridge University Press.
- Plato. (1961a), *Plato: The Collected Dialogues*. eds. Hamilton, E. and Cairns, H. Princeton: Princeton University Press.
- _____. (1961b), "Euthyphro." Translated by Lane Cooper. In Plato 1961a., pp. 169-185.
- Smith, M. (1987), "The Humean Theory of Motivation", *Mind*, pp. 36-61.
- _____. (1994), *The Moral Problem*, Oxford: Oxford University Press.
- _____. (1999), "Search For the Source", *Philosophical Quarterly*, 49, pp. 384-394.
- Zimmerman, M. (1999), "Virtual Intrinsic Value and the Principle of Organic Unities", *Philosophy and Phenomenological Research*, pp. 653-666.