

PHIL 260; Spring 2007

The Normativity of Ethics

David O. Brink

Handout #7: Normative Authority and Korsgaardian Rationalism

In The Sources of Normativity (1996) Christine Korsgaard provides a dialectical examination of different conceptions of the sources of normativity or reasons -- conceptions that appeal to voluntarism, realism, and reflective endorsement -- that culminates in her own Kantian or neo-Kantian conception of normativity that is grounded in **autonomy**. Her method is dialectical (Dialectical) inasmuch as her neo-Kantian conception is supposed to reveal the truth or grain of truth in each of the three prior conceptions.

Korsgaard begins Lecture 1 with a statement of the main problems that arise in connection with the normativity of ethics. She factors normative adequacy into answers to two kinds of problems (12-13).

1. **The explanatory problem** is explaining how moral considerations can and do influence us.
2. **The normative or justificatory problem** is to explain why we ought to care about conforming to moral demands or why we are justified in so conforming.

Though the problems are related, they are distinct. They correspond roughly to the issues of moral motivation and the authority of morality, which we have distinguished in the seminar.

Korsgaard also thinks that these two issues can interact. She mentions the possibility of explanatory theses, such as an evolutionary account of moral motivation, that might subvert the normative thesis (14). However, the interaction here is complicated. An evolutionary explanation of moral beliefs might explain why we would hold and be motivated by moral beliefs, even if they were not true or authoritative, namely, because they confer selective advantage. But such etiological claims do not themselves provide reason to think that moral beliefs are not true or lack authority. An **independent** argument for their lack of truth or authority would be required to subvert their justification. All the evolutionary etiology tells you is that you can't argue for their truth or authority just from the fact that you hold moral beliefs.

The primary philosophical problem about normativity concerns the justification or authority of moral demands (20). The rest of Lecture 1 probes voluntarist and realist conceptions of the sources of normativity.

VOLUNTARISM

According to the voluntarist, obligation derives from command or perhaps agreement. You must do X because God commands it or because a political sovereign whom you have agreed to obey commands it. It is also relevant that authorities can and will apply sanctions for noncompliance. On this view, normativity is grounded in authority. Authorities can have reasons for what they command (their commands need not be arbitrary), but it is the fact of their command that creates the obligation.

Pace Korsgaard, if the command can only obligate if its content tracks independent reasons (= is not arbitrary), then the obligation depends on normative reasons that are prior to and independent of the command. In any case, as Korsgaard does argue, the obligation does

depend upon commandment-independent principles about the existence of reasons to obey commands – for instance, a principle recognizing filial duties or the duty to keep one's covenants.

REALISM

The realist says that these normative principles or considerations that underlie the obligations created or imposed by commands are prior to and independent of the commands. They do not derive from command. More generally, the realist posits ultimate or unconditioned normative truths. We can ask of any consideration *X* why we should care about it. Often we can answer by relating consideration *X* to another type of consideration *Y*. But if either *X* or *Y* is to have normative significance, its derived or extrinsic normative significance must ultimately be grounded in something intrinsically normatively significant, whose normative significance does not depend on anything else. If a consideration is of ultimate normative significance, we cannot justify its significance in terms of anything else, for then its significance would not be ultimate. Hence, our normative knowledge must rest on self-evident normative truths.

Korsgaard describes this sort of normative realism as substantive normative realism and contrasts it with procedural normative realism (35-37). She accepts procedural realism but has problems with substantive realism. She raises two problems for substantive normative realism.

1. Are there any ultimate normative truths that are self-evident?
2. Even if there are, why suppose that they support the normative discriminations most of us make?

If an assertion of ultimate normative truths is not to be a mere expression of confidence in our own normative judgments, we must be able to explain the normative authority of factors that we take to be normatively significant. But, claims Korsgaard, the realist cannot do that since that would show that the factors in question did not enjoy ultimate normative significance.

We might wonder whether the realist is barred from a discursive justification of her ultimate normative factors. We should distinguish between metaphysical and epistemic dependence. Derived normative truths are asymmetrically metaphysically dependent on ultimate normative factors. But the epistemic dependence among normative judgments need not be asymmetrical; epistemic dependence can be bi-directional. The epistemology of ultimate normative truths need not be one of self-evidence. A first principle identifies an ultimate normative factor. Consequently, it makes no sense to ask of a first principle that we take to be true

- "In virtue of what further property is *F* an ultimate normative factor?"

For *F* to be an ultimate normative factor is precisely for its normative not to depend upon some further property. But we can still sensibly ask whether *F* is an ultimate normative factor.

- "Is *F* really a first principle?" or "Is *F* really an ultimate factor?"

But we can answer these questions by showing how treating *F* as an ultimate normative factor would sustain and explain our independent beliefs about what we have reason to do.

REFLECTIVE ENDORSEMENT

Korsgaard thinks we can avoid the problems of substantive realism if we ask not whether our actions, projects, and commitments track objective normative facts but rather whether they can be reflectively sustained (Lecture 2). To appeal to reflective endorsement, she thinks, is to appeal to human nature or desire (19). It also appears to be a form of procedural realism inasmuch as it implies that moral considerations have rational authority just in case moral commitments would survive a process of reflective endorsement. Korsgaard's main examples of figures in the tradition of reflective endorsement are Hume, Mill, and Williams. (This reading of Mill largely ignores the perfectionist strands in Mill's ethical and political theory.) This tradition appears to put reflective endorsement within a broadly sentimentalist framework inasmuch as reflective endorsement appears to be a matter of what commitments and ways of living we are prepared, upon clear-eyed and informed reflection, to endorse or live with. Korsgaard's discussion of Hume is instructive (51-66), and our discussion of reflective endorsement will focus on Hume.

HUME ON JUSTICE AND THE SENSIBLE KNAVE

Hume thinks that moral judgments are based on sentiments of approval and disapproval that we feel when we contemplate someone's character from a general or common point of view that aims at public accord rather than private interest (Treatise 581-82; Enquiry sect. ix, pt. 1). While we can explain the basis of moral judgments in this way, Hume seems to think that we can still ask how they obligate, by which he seems to mean what their authority consists in. In the Enquiry he writes

Having explained the moral *approbation* attending merit or virtue, there remains nothing but briefly to consider our interested *obligation* to it, and to inquire whether every man, who has any regard to his own happiness and welfare, will not find his account in the practice of every moral duty [ix 2].

The question seems to be whether we should approve of having a virtuous character and actions that display such a character. In principle we might appeal to any sentiments we have. And Hume does believe that we are sympathetic by nature and that sympathy approves of most of the virtues (Treatise 577-79, 618-19). But here Hume does not rely on sympathy, but instead appeals to self-interest. Our sympathies are limited and weak in comparison with our sense of self-love (487, 586). This appears to lead Hume to assume that our reflective approval of virtuous character and action depends on showing that and how the virtues contribute to one's own happiness. This is easy enough in the case of self-regarding virtues, such as temperance. It's harder in the case of other-regarding virtues. But even here Hume thinks that it is not hard to see how it could be advantageous to be agreeable to others (Enquiry ix 2). The real problem concerns justice. For Hume justice is a virtue defined by rules of forbearance and cooperation the general observance of which is mutually advantageous. Individuals have reason to participate in just systems, because each benefits from the justice of others. But, as the Sensible Knave realizes, just actions are not always beneficial to the agent.

Treating vice with the greatest candour, and making it all possible concessions, we must acknowledge, that there is not, in any instance, the smallest pretext for giving it the

preference above virtue, with a view to self-interest; except, perhaps, in the case of justice, where a man, taking things in a certain light, may often seem to be a loser by his integrity. And though it is allowed, that, without a regard to property, no society could subsist; yet, according to the imperfect way in which human affairs are conducted, a sensible knave, in particular incidents, may think, that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That *honesty is the best policy*, may be a good general rule; but is liable to many exceptions: And he, it may, perhaps, be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions [ix 2].

Just actions are costly. But the existence of a just system does not usually depend upon any one individual acting justly. Even if everyone does better when all are just than when none are, and even if each is usually better off being just than unjust, there are some circumstances in which being just is not necessary to secure the benefits of the system of justice. The Sensible Knave doubts that being just in such circumstances commands reflective endorsement. Hume's answer is puzzling.

I must confess, that, if a man think, that this reasoning much requires an answer, it will be a little difficult to find any, which will to him appear satisfactory and convincing. If his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villany or baseness, he has indeed lost a considerable motive to virtue; and we may expect, that his practice will be answerable to his speculation. But in all ingenuous natures, the antipathy to treachery and roguery is too strong to be counterbalanced by any views of profit or pecuniary advantage. Inward peace of mind, consciousness of integrity, a satisfactory review of our own conduct; these are circumstances very requisite to happiness, and will be cherished and cultivated by every honest man, who feels the importance of them.

Such a one has, besides, the frequent satisfaction of seeing knaves, with all their pretended cunning and abilities, betrayed by their own maxims; and while they purpose to cheat with moderation and secrecy, a tempting incident occurs, nature is frail, and they give into the snare; whence they can never extricate themselves, without a total loss of reputation, and the forfeiture of all future trust and confidence with mankind [ix 2].

After apparently conceding that justice fails to secure the Knave's reflective endorsement, Hume appears to go on to offer various responses to the Knave. Here, Hume appears to assume that the Knave proposes **deceptive injustice**.

Should we make this assumption about the Knave? It's not clear there couldn't be circumstances in which a Knave might be willing to be **openly** unjust, for instance, in openly failing to heed demands for the cooperative production of public goods (because the goods are nonexcludable) or in openly failing to reciprocate in a case of sequential cooperation in which the benefits of noncompliance dwarf the costs of being excluded from future cooperative interaction.

But let's waive this concern and accept Hume's assumption that Knaves propose deceptive injustice. Hume's first response appears to be that honest people will not on reflection approve of duplicity. But that's very close to saying that honesty approves of honesty or that

honesty approves of itself. At points, Korsgaard suggests that this sort of reflexivity test is integral to reflective endorsement. Later, in discussing Hume's standard of taste, she writes

The faculty of taste gives rise to verdicts of beauty. According to this theory a faculty's verdicts are normative if the faculty meets the following test: when the faculty takes itself and its own operations for its object, it gives a positive verdict [62].

Korsgaard goes on to suggest that Hume thinks that the moral sense also passes this reflexivity test (63, citing Treatise Liv.1/180-85). So reflexivity may be a real strand in Hume's defense of the normativity of ethics, in general, and of justice, in particular. Perhaps reflexivity ought to be a necessary condition of normativity. If a given faculty or perspective cannot approve of its own operations, perhaps that is reason to question its authority. But it's hard to treat reflexivity, as Korsgaard does, as a sufficient condition of normative authority. The Knave wonders whether she should care about honesty. It seems like a poor answer to that challenge to vindicate honesty by appeal to a sense of honesty. Not only is that defense of honesty question-begging, it apparently ignores the Knave's appeal to self-love or happiness as the benchmark for approval.

Hume's last response in this passage above does address the Knave's appeal to self-interest. It appears to argue that it is hard to practice injustice on all and only those occasions when it will go undetected. The Knave who practices injustice selectively and engages in deception to exploit others with impunity is likely eventually to miscalculate and get caught out. The penalty for this will be exclusion from the benefits of society (i.e. the forbearance and cooperation of others). But this seems not to be an argument against knavery, but against incompetent knavery. The **Sensible** Knave will disapprove incompetent knavery, but why should she approve justice or honesty per se? Why not approve instead selective injustice and dishonesty?

At one point, Korsgaard suggests that Hume is claiming not that just acts are always more beneficial than unjust acts but rather that it is more beneficial **to be the sort of person** who always practices justice for its own sake than to be the sort of person who practices justice iff it is the most beneficial option available (60). Korsgaard offers fairly little by way of textual support for this interpretive proposal, though one might note that Hume introduces the challenge of the Sensible Knave in the context of a discussion of what sort of "disposition ... he would choose for the foundation of his happiness and enjoyment" (ix 2). The suggestion has actually been explored more fully by David Gauthier (Morals by Agreement and "Three against Justice: the Foole, the Sensible Knave, and the Lydian Shepherd"). This view deserves fuller discussion, but it faces a problem. It may be hard to be the second sort of person -- to be a genuinely Sensible Knave. Perhaps we can't turn attitudes of reciprocation on and off like the tap on a faucet (though that really seems to be an empirical question). If so, our actual choice may be between being a Just Person and an Incompetent Knave. We may well be better-off choosing justice over incompetent knavery. But that provides no reason for thinking that we shouldn't prefer competent knavery to justice. Justice remains, as Glaucon and Adeimantus worry, a second-best option behind undetected injustice. If so, it's hard to see how the Sensible Knave can give her whole-hearted approval of justice or other artificial virtues.

A DILEMMA FOR REFLECTIVE ENDORSEMENT

It was a little hard for me to find a general assessment of reflective endorsement in Korsgaard amidst the details of the Humean, Williamsesque, and Millian exemplars of that

tradition. Here's one reaction that would fit with some of Korsgaard's dialectical picture. The proponent of reflective endorsement imagines an agent reflecting on the authority of an action or disposition to act and asking herself whether she can endorse that disposition or action. Either she appeals to the very same disposition, as in the reflexivity conception of reflective endorsement, or she appeals to a distinct disposition, as the Knave appeals to self-interest in assessing a commitment to justice. If it's the same disposition, it provides an impermissibly circular justification. But if it's a different disposition, we still don't have a justification of the original disposition unless this second disposition is suitably grounded. But now we've just pushed the question of authority or justification back to another level. An appeal to self-love or any other disposition will not ground the authority of other dispositions or commitments unless its authority can be grounded. One might regard substantive realism as an antidote to this threat of regress. But Korsgaard thinks that reflective endorsement can be developed in a Kantian direction.

AUTONOMY, IDENTITY, AND HUMANITY

Korsgaard sees Kant as addressing this kind of worry about reflective endorsement by reminding us that the test of reflective endorsement is a test to see if there is a reason to act (89).

If the problem springs from reflection then the solution must do so as well. If the problem is that our perceptions and desires might not withstand reflective scrutiny, then the solution is that they might. We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do. The normative word 'reason' refers to a kind of reflective success [93].

But this seems merely to repeat the reflective endorsement conception. If reflection is settled by a bare appeal to particular desires or dispositions, there is still reason to doubt the authority of this kind of reflective success. Korsgaard's response is, I think, to ground the normative authority of reflective endorsement in **autonomy** or **agency**. Reflective endorsement should be understood as expressing our capacities as agents to act for reasons and not simply on our kinesthetically strongest desires (94-104). On this view, normative endorsement expresses our capacities as autonomous agents that make us subject to reasons for action in the first place. This is what gives the upshot of reflective endorsement normative authority.

Central to Korsgaard's conception of practical reason is her idea of a **practical identity** (101). This seems to involve the idea that reflective endorsement will both be guided by and yield (perhaps inchoate) conceptions or descriptions of oneself under which one values oneself. Often this will take the form of embracing certain roles -- teacher, father, husband, friend, teammate -- as valuable. Things like rules and commitments can then be assessed for normative significance in terms of their contribution to one's practical identities. There are at least two worries about this neo-Kantian appeal to practical identities as so far described.

One worry is that we seem not to have been given any basis for distinguishing between good and bad practical identities. My identity as philosophy professor, nurturing parent, or decent recreational hockey player may well provide me with reasons to care about and do things. But does the practical identity of the committed (non-self-loathing) lint collector or serial killer also give her reasons for action? Rationalists should want to say that they do not or that, if they do, they are very weak reasons that are easily swamped by her reasons not to so act.

Another worry is that we have many conflicting identities. How do we resolve such conflicts? For instance, how can we show that my practical identity to be a good member of my moral community trumps my practical identity to be a successful scientist when I could advance my career by claiming credit for my lab assistant's discoveries?

Korsgaard is not interested in resolving all practical conflicts (91). And it's unclear, to me at least, whether she is willing to grant any significance whatsoever to nasty maxims or practical identities provided that they survive reflective endorsement. However that issue is resolved, she does seem to want to defend the supremacy of moral demands by connecting them with the practical identity of humanity and arguing that this practical identity takes precedence over or conditions the value of others.

Some commitments or identities are valuable only contingently. They are not intrinsically valuable or valuable independently of being the object of choice. Value is conferred on them by rational endorsement. What is not contingently valuable is one's identity as an agent, since adopting this identity is a condition of undertaking reflective endorsement. Autonomous agency is what confers value on these contingently valuable identities. Like Kant, Korsgaard understands agency as humanity.

This is potentially misleading, since agency or personhood and humanity can easily be understood as different categories. Taken literally, the category of humanity is a biological category. It's not immediately obvious why such a biological category should be normative. By contrast, personhood is, as Locke says, a forensic or normative category. If personhood involves being a responsible agent, capable of acting for the sake of reasons for action, then it seems clearer why personhood carries normative authority. For these reasons, it's better to eschew focus on humanity and keep the focus on agency or moral personality.

However, to value humanity or agency as such is to care about rational agents as ends in themselves. This reasoning appears to suggest that Kant's Humanity Formula of the Categorical Imperative has necessary normative authority and conditions the authority of one's other practical identities. This conclusion would apparently imply not just that moral demands have supreme authority but that contra-moral demands and identities have no authority whatsoever (113-22, 129-30).

One question about Korsgaard's favored conception of normativity, which grounds other-regarding duties in considerations of agency or autonomy, is how it differs from substantive realism. Why isn't her view representable as a form of realism in which autonomy figures as the ultimate value? Perhaps she will claim that autonomy is a procedural, rather than substantive, value. But in this context I don't know what that distinction would amount to.

FROM HUMANITY TO MORALITY

As Korsgaard herself recognizes, there is a concern about this defense of the authority of other-regarding morality (121, 130, 132-45). It is perhaps clear how autonomy or agency is a value that is presupposed by rational endorsement and so is rationally compulsory. We might even concede that autonomy, so conceived, conditions the value of other projects and identities. What is unclear is why it is agency or autonomy as such, as opposed to an agent's own agency or autonomy, which has this transcendental value. Perhaps every agent must be concerned with his own rational agency. Why should or must he be concerned with other rational agents as such?

Korsgaard wants to argue that one cannot value one's own humanity (agency) without valuing the humanity (agency) of others (132-45). She treats the challenge as one which grants the existence of private reasons but denies that you can get from private reasons to public reasons

(134). Her response is to deny that there are private reasons and to insist that all reasons are essentially public (134-36). She defends the essentially public character of reasons by analogy with Wittgenstein's private language argument (136-38). As I understand Wittgenstein, or at least Korsgaard's Wittgenstein, the central idea is something like this.

1. Meaning is a normative notion.
2. Hence, linguistic meaning presupposes correctness conditions.
3. The correctness conditions must be independent of a particular speaker's utterances.
4. Hence, correctness conditions must be established by the usage conventions of a community of speakers.
5. Hence, a private language is not possible.

Even if one accepted something like this argument for the public nature of semantic normativity, one might well wonder whether that was an apt model or analogy for practical reason. For one might think that interpersonal communication was essential to language in a way that a shared point of view is not essential for practical reason. But even if we accept the analogy, we might resist the conclusion by denying this argument against the possibility of a private language. The move from (3) to (4) seems problematic. We can satisfy (3) without commitment to (4). If there were Platonistic truth conditions, then we could have correctness conditions not established by shared conventions about usage. Perhaps Platonism is implausible about meaning, precisely because interpersonal communication is essential to language, and transcendent conceptions of meaning threaten communication. But it's less obvious that Platonism is an inappropriate model for the normativity of practical reason. In any case, we can deny the inference from (3) to (4) without embracing Platonism. (3) can be satisfied without resort to (4) if correctness conditions are fixed by diachronic intrapersonal agreement or convergence in usage. Such private languages may be of very limited value in navigating a social world, but they don't seem impossible provided there are stable principles for intrapersonal usage. But that would of course defeat the private language argument.

Korsgaard elaborates on the lessons of the private language argument for practical reason by returning to voluntarist themes. Practical laws requires two parties a legislator and a legislatee.

Since it is a relation, and indeed a relation in which one gives a law to another, it takes two to make a reason [138].

But surely there's been no argument that legislator and legislatee need be different people. Self-legislation is not only possible, but is supposed to be the basis of Kantian autonomy. But then it's hard to see how private reasons could be illusory or posterior to public reasons.

Korsgaard's failure here parallels Kant's own. Rational agency is a potentially ambiguous commitment. Kant explains well why each must regard his rational self as superior to his empirical self but he provides no good explanation of why one rational self should be intrinsically and unconditionally concerned with other rational selves. If we are to address this gap in the authority of other-regarding concern, we must either find a more cogent defense of impartial practical reason or look to traditions outside Kantian ethics, such as the Greek eudaimonist tradition or the ethics of self-realization found later British idealism, that attempt to

show why someone concerned with his own rational agency should be non-instrumentally interested in the rational agency of others.