

PHIL 260; Spring 2007

The Normativity of Ethics

David O. Brink

Handout #4: Normative Authority and Error Theories, Relativism, Minimalism, Response-Dependence, and Strategic Egoism

Another way to think of the normativity of ethics is, not in terms of whether morality does motivate, but in terms of whether it **should**. One way to understand this aspect of normativity is in terms of the **authority** of morality. But the authority of morality raises its own puzzles. It is common to think of morality as both objective and impartial, in particular, as containing various other-regarding duties of cooperation, forbearance, and aid that apply to agents independently of their own aims and interests. Most of us also regard moral obligations as authoritative practical considerations. But heeding these obligations appears sometimes to constrain the agent's pursuit of his own interest or aims. If we accept instrumental or prudential conceptions of rationality, we may wonder whether moral conduct is always rationally justifiable. We can capture this tension in terms of a puzzle about the authority of morality.

1. Moral requirements include impartial other-regarding obligations that do not apply to agents in virtue of their aims or interests.
2. Moral requirements necessarily provide agents with reasons for action.
3. Rational action is action that advances the agent's aims or interests.
4. Fulfilling other-regarding obligations need not advance the agent's aims or interests.

(1) expresses a commitment to morality's **objectivity** and **impartiality**; (2) is a **rationalist** thesis about the authority of morality; (3) reflects common assumptions that **practical reason is instrumental or prudential**; and (4) reflects a common view about the **independence** of different people's aims and interests.

As with the puzzle about moral motivation, we can think of various familiar positions about morality and practical reason as responses, express or tacit, to the tension underlying this puzzle about normative authority. A motley crew rejects the assumptions about objectivity and impartiality in (1), including **error theorists, relativists, minimalists and response-dependent views**. They all believe that moral requirements must be relativized to and further the agent's aims and interest in some way. **Anti-rationalists** reject (2), claiming that immoral conduct need not be irrational. Those who believe that **practical reason can be impartial** reject (3) and claim that an agent has non-derivative reason to benefit others. **Metaphysical egoists** reject (4) and claim that, properly understood, people's interests are interdependent such that acting on other-regarding moral requirements is a counterfactually reliable way of promoting the agent's own good.

We'll begin with the first family of solutions to the puzzle about normative authority – those that deny objectivity and/or impartiality and relativize moral requirements to the agent's aims or interests in some way.

ERROR THEORIES, RELATIVISM, AND MINIMALISM

We might appeal to (2)-(4) to reject (1). If our rejection focuses on the objectivity in (1), it might make sense to view the resulting position as a kind of **metaethical** view. This may be one way of understanding some forms of error theories and some forms of relativism.

A. Mackie. For instance, we might make sense of J.L. Mackie's error theory in Ethics: Inventing Right and Wrong (ch. 1) within this framework. Mackie thinks that it's part of our ordinary concept of morality that there are other-regarding moral requirements that apply to agents and that give them reason – indeed, overriding reason – for action independently of their contingent inclinations and desires. He thus accepts both (1) and (2). However, he thinks that these ordinary assumptions are mistaken because there are no such objective values. It's not that the belief in objective values is incoherent; it is just false. His skepticism about objective values is like the skepticism that most of us have about witches and that atheists have about God. We have reason to deny that there are objective values, because they would be both metaphysically and epistemologically queer. Instead, the only kind of reasons that Mackie recognizes are hypothetical imperatives. Specifically, the only kind of reason he seems to recognize are instrumental reasons, reasons grounded in one's desires. Prudential reasons, which are desire-independent, seem just as mysterious to him as objective moral values. So he defends a special case of (3), and he seems to take the independence assumption in (4) for granted.

Mackie's skepticism about ordinary morality, as he sees it, together with his acceptance of instrumentalism about practical reason, make it natural to interpret him as denying (1). I'm less sure that his constructive claims about morality support this interpretation. For it may be possible to interpret his constructive Humean claims about the point and content of morality in part II of his book as suggesting a positive (revisionary ?) conception of morality as consisting in behavioral norms the general acceptance of which is mutually advantageous. As we will see later, such a conception of morality will recognize the existence of moral norms against unprovoked aggression, theft, unfairness, and infidelity. While compliance with such norms will often be beneficial and we can explain why agents will often care about compliance with such norms, there is no guarantee that compliance will satisfy an agent's desires. If not, then Mackie's constructive account of morality might be best interpreted as a form of anti-rationalism, similar, at least in spirit, to Foot's anti-rationalism.

I don't take this possible anti-rationalist interpretation to show any incoherence, as such, in Mackie. It seems possible for him to combine skepticism/nihilism about ordinary moral commitments with an anti-rationalism about his own favored view of morality. Incoherence would result only if ordinary morality can or should be understood in terms of mutual advantage. And, of course, either of Mackie's claims (negative or positive) is only as plausible as his arguments against objective values and categorical imperatives and his arguments in favor of instrumentalism about practical reason. These assumptions certainly deserve scrutiny.

B. Harman. A related metaethical view that fits naturally within this framework is Gilbert Harman's defense of moral relativism in "Moral Relativism Defended". He combines rationalism about moral requirements (expressed by "inner" judgments) with an instrumental conception of practical reason to yield a "soberly logical" relativist conclusion that moral duties must be relativized to the desires of the agent.

1. Moral requirements entail reasons for action.
2. Reasons for action are desire-dependent.
3. Hence, moral requirements are desire-dependent.

Harman is admirably clear about his commitment to rationalism and to instrumentalism about practical reason. Unfortunately, he offers no real defense of either commitment.

Harman's form of relativism appears to be a morally revisionary view insofar as it tailors an agent's duties to her desires. In this way, it's as much a **minimalist** moral view as it is a metaethical view. Here the boundaries between metaethics and moral theory are not sharp.

Another view within this framework that has both metaethical and moral dimensions is Callicles' position in Plato's Gorgias. Callicles resolves the apparent conflict between the demands of virtue and the agent's own interest by arguing that "real" or natural justice does not require the agent to help others or forbear from harming them, as conventional morality supposes (482de, 483ab, 488b-490a). The naturally just person satisfies her own unrestrained desires (488b).

1. Justice is a virtue.
2. Virtues must benefit their possessor.
3. Conventional justice benefits others, not the agent.
4. Hence, conventional justice is not real justice.
5. Hence, real justice consists in benefiting oneself, not others.

This is a kind of error theory insofar as it says that conventional justice is not a genuine virtue, but it is a substantive and revisionary moral theory insofar as it claims that genuine justice does not recognize obligations to benefit others (that cannot be justified instrumentally).

D. The General Worry. Any ethical theory that holds an agent's obligations hostage to the promotion of her immediate or unreflective interests or desires seems committed to a revisionary moral view; it seems unable to explain why those who do not have other-regarding attitudes should cultivate them and why those who happen to have them should maintain them. These are worries about the **scope** and **stability** of such conceptions of morality.

RESPONSE-DEPENDENCE

We might try to avoid worries about the revisionary implications of relativism and minimalism by grounding moral requirements in **idealized desire**. On such views (e.g. Firth), we give a **response-dependent** account of moral requirements as the actions that a suitably idealized observer would endorse or desire. If this form of response-dependence is to be true to its instrumentalist assumptions about practical reason, then the process of idealization must itself be purely formal and morally neutral; it must be limited to full non-normative information, vivid representation of this information, good reasoning, etc. For instance, we might model this sort of response-dependence on the sort of idealized desire account that some have offered of well-being or the personal good.

Such desiderative conceptions of well-being have a distinguished pedigree. In Utilitarianism Mill at least suggests an idealized desire conception of happiness when he explains the intrinsic, and not just instrumental, superiority of higher pleasures by appeal to the preferences of a competent judge.

If I am asked what I mean by difference of quality in pleasures, or what makes one pleasure more valuable than another, merely as a pleasure, except its being greater in amount, there is but one possible answer. If one of the two is, by those who are compe-

tently acquainted with both, placed so far above the other that they prefer it, even though knowing it to be attended with a greater amount of discontent, and would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superiority in quality so far outweighing quantity as to render it, in comparison, of small account [Utilitarianism ii 5].

At one point in The Methods of Ethics Sidgwick proposes that we understand a person's overall good to consist in

What he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realized in imagination at the present point in time [Methods 111-12].

In A Theory of Justice John Rawls adapts Sidgwick's proposal and identifies a person's good with a rational plan of life.

It is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in light of all the relevant facts, what it would be like to carry out all of these plans and thereby ascertained the course of action that would best realize his more fundamental desires [TJ 417].

In A Theory of the Good and the Right Richard Brandt identifies a person's well-being with what it would be rational for her to desire, and he understands rational desire as desire that would survive a process of cognitive psychotherapy that requires full and vivid exposure to logic and the relevant facts.

However, recent work within this tradition suggests some modifications in the classical informed desire theory of well-being. Recognizing that even in a more idealized state we might have desires that we do not endorse or identify with, David Lewis proposes that something is good just in case one would, under conditions of full imaginative acquaintance with the alternatives, desire to desire it. Peter Railton notices that an ideal appraiser is likely to be very different from the actual self that it idealizes and that, consequently, what my idealized self may want for himself may not be appropriate for me. For instance, education appears to be a good for my actual self, but because my idealized self is already fully informed, he may not desire (or desire to desire) to get an education. To remedy this source of potential alienation, Railton proposes that we appeal to what the ideal appraiser would want his actual self to want.

[A]n individual's good consists in what he would want himself to want, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality ["Facts and Values," 16].

On this view, A's good consists in what A+ would want A to want.

These desiderative forms of response-dependence are conceptions of the personal good. But notice that they place no restrictions on the content of the agent's desires and, in particular, make no requirement that the object of desire be related to the appraiser herself. This has led some people to claim that such views confuse **what interests an appraiser** and **what is in her**

interest or what she values and what is valuable for her. That's a problem in a theory of well-being, but not necessarily in a moral conception. So we might entertain such desiderative views as conceptions of morality.

But there are various worries about laundering actual preferences in this way. Some worries concern commitments introduced by idealization, specifically appeal to an appraiser who is fully informed about all of his opportunities and vividly represents their various features, so that he is omniscient with respect to all the experiential and non-experiential aspects of the options available to him. An ideal appraiser must evaluate different possible lives. But one question is whether it is possible to combine wildly disparate lives and perspectives into one overall evaluative perspective (Sobel, Rosati). The conditions that make a vivid appreciation of one perspective accessible may make a vivid appreciation of a very different perspective inaccessible. For example, the conditions that make a naïve or insular perspective accessible, such that one can appreciate its attractions, may make a cosmopolitan perspective inaccessible, and vice versa.

Furthermore, even where diverse possibilities are jointly accessible from a common perspective that does each phenomenological justice, we may wonder whether the effect of vivid representation is normatively significant. One can't rule out the possibility that full confrontation with the facts wouldn't extinguish desire or shape it in ways that one would pre-theoretically identify as pathological (Gibbard). Perhaps the weakness of altruistic impulses is typically due to an inadequate appreciation of the suffering of others. But vivid exposure to the enormity of suffering involved in world hunger may overwhelm or de-sensitize appraisers so as to suppress, rather than elicit, sympathetic response. Here, vivid representation produces what are intuitively exactly the wrong normative results.

Moreover, content-neutral idealization seems unable to remove worries about the revisionary implications of response-dependent views. For as long as the idealization is formal and content-neutral, as instrumentalism about practical reason requires, then it must be a contingent fact what ideal desirers will care about. Even if most people are such that their idealized selves would want their non-idealized selves to care about other people as morality seems to require, there's nothing to guarantee this. Those whose ideal selves would fail to have such desires can apparently be under no other-regarding moral obligations. And those who happen to have these desires can always ask why they should persist in having them, especially if more self-confined desires are easier to satisfy. These are limitations in the **scope** and **stability** of moral demands that appear quite revisionary.

STRATEGIC EGOISM

One attempt to meet this challenge is embodied in that strand of social contract theory -- including Epicurus, Hobbes, and Gauthier -- that understands the scope, content, and authority of morality in terms of rational agreement. In its classical form, this view appeals to a prudential or egoist conception of rationality and argues that it is in the long-term interest of agents to develop, maintain, and act on other-regarding norms and attitudes. The general story is familiar enough and can be illustrated with an example from Hume.

Your corn is ripe today; mine will be so tomorrow. 'Tis profitable for us both, that I shou'd labour with you today, and that you shou'd aide me tomorrow. I have no kindness for you, and know you have as little for me. I will not, therefore, take any pains upon your account; and shou'd I labor upon my own account, in expectation of a return, I know I shou'd be

disappointed, and that I shou'd in vain depend upon your gratitude. Here then I leave you to labour alone; You treat me in the same manner. The seasons change; and both of us lose our harvests for want of mutual confidence and security [Treatise III.ii.5/520-21].

Hume's example involves a one-shot case of sequential cooperation. Where the cooperation would be simultaneous, the situation can be represented as a **prisoner's dilemma** (PD).¹ If the benefit to each of the other's contribution is 10 and the cost to each of his own contribution is 5, then the pay-offs are as follows.

		B	
		Contribute	Don't Contribute
A	Contribute	A: 5 B: 5	A: -5 B: 10
	Don't Contribute	A: 10 B: -5	A: 0 B: 0

For every rational individual, noncooperation strictly **dominates** cooperation; whatever the other does, each is better-off not cooperating. Every rational individual would reason in this way, with the result that rational individuals cannot, in principle, secure mutual benefits.

But typically A and B have opportunities for ongoing mutually advantageous interaction with each other and others. Here we have a so-called **iterated** PD situation, which changes the pay-off structure of the original game (so that we no longer have a genuine PD). If A contributes, it is better for B to contribute, because the benefits of contribution and the costs of noncontribution are higher than in the single-shot PD; if B fails to contribute, A will cease to deal with him and/or retaliate and B will suffer bad reputational effects with others.

Insofar as moral interaction can be modeled as an iterated PD situation, the rational egoist can perhaps justify other-regarding moral behavior. Much of impartial other-regarding morality involves norms of cooperation (e.g. fidelity and fair play), forbearance, and aid. Each individual has an interest in the fruits of interaction conducted according to these norms. Though it might be desirable to reap the benefits of others' compliance with norms of forbearance and cooperation without incurring the burdens of one's own, the opportunities to do this are infrequent. Noncompliance is generally detectable, and others won't be forbearing and cooperative toward those who are known to be noncompliant. For this reason, compliance is typically necessary to enjoy the benefits of others' continued compliance. Moreover, because each has an interest in others' cooperation and

¹For the classical PD-situation, substitute "stay silent" for "contribute" and "rat" for "don't contribute" and put the following values in boxes : NW box = A and B each serves 2 years in prison; SW box = A goes free and B serves 12 years; NE box = B goes free and A serves 12 years; and SE box = A and B each serves 10 years. It's essential, of course, that prisoners consider the value only of jail terms -- for instance, they are not to consider the negative reputational effects of being a "stool pigeon" -- in other words, the pay-offs should really be in terms of utilities.

restraint, communities will tend to reinforce compliant behavior and discourage noncompliant behavior. If so, compliance is often necessary to avoid such social sanctions. Whereas noncompliance secures short-term benefits that compliance does not, compliance typically secures greater long-term benefits than noncompliance. The structure of this justification is something like this.

1. It is in an agent's interest to receive the benefits of norms of cooperation and restraint.
2. Others won't be cooperative and restrained toward those who are known to be noncompliant with these norms; the benefits of systems of cooperation and restraint are available only to those who appear to be cooperative and restrained.
3. Given the costs of detected noncompliance and the psychic costs (e.g. anxiety) that noncompliance will be detected, the least costly way to maintain the appearance of compliance is to be compliant.
4. Hence, it is in an agent's interest to be cooperative and restrained.
5. Hence, there are other-regarding duties that have prudential foundations.

In this way, a strategic conception of morality can try to justify other-regarding morality and avoid unacceptably minimalist conclusions while denying the claim that these moral requirements apply to agents independently of their interests and desires. But there are several worries about this account of the authority of morality.

A. (1) is not always true. The price of detected noncompliance is exclusion from the benefits of the system of cooperation and restraint. But the penalty may not always outweigh the rewards of noncompliance.

B. (2) is not always true. In the case of public goods that are nonexcludable, the appearance of compliance is not necessary to receive the benefits of compliance by others, though fairness typically seems to require compliance.

C. (3) is sometimes false. Selective noncompliance may go undetected where it is difficult to monitor compliance and detect deception. Where detection is unlikely, it would be unreasonable to be terribly anxious about detection. In such cases, compliance would not be the least costly means of maintaining the appearance of compliance.

D. (4) and (5) may be too weak. For even if the defense of other-regarding moral behavior was extensionally adequate, it would still be counterfactually fragile. On this justification of compliance with other-regarding norms, compliance involves costs, as well as benefits; it must remain a second-best option, behind undetected noncompliance, in which one enjoys the benefits of others' compliance without the costs of one's own. But then if one had some way of ensuring that one's own noncompliance would go undetected -- for instance, one had sole access to the ring of Gyges -- one could enjoy the benefits of others' compliance without the burdens of one's own, and one would have no reason to be compliant. Whether this is a limitation in morality itself or in the rational authority of morality depends on how we understand the proposed link between morality and self-interest.

If we understand morality in act-egoist terms, then morality extends just as far as self-interest. In those circumstances, whether actual or counterfactual, in which one can get the benefits of others' compliance without the costs of one's own, compliance is not morally required. This is a revisionary moral claim; commonsense morality condemns the noncompliance whose rationality the ring of Gyges makes possible.

Alternatively, we might understand morality to consist in compliance with mutually advantageous norms of interaction, such as norms of cooperation and restraint. On this view,

morality demands compliance even in those circumstances, actual or counterfactual, in which one can get the benefits of others' compliance without the burdens of one's own. In these circumstances, moral demands cease to be rationally authoritative. This is a kind of anti-rationalist position.

What is true on both views is that the strategic justification of compliance with the sort of other-regarding norms contained in commonsense morality is counterfactually fragile. Whereas the strategic justification of compliance must be unstable in this way, we, like Glaucon and Adeimantus, might demand that any justification of other-regarding morality be counterfactually stable.

E. Finally, this view must assume that the scope of other-regarding moral demands is limited to strategic contexts. Consider Gauthier's motive-egoism about morality.

Among beings, however rational, who may not hope to engage one another in a cooperative venture for mutual advantage, the proviso [prohibiting bettering one's bargaining position through interaction worsening the position of another] would have no force. Our theory denies any place to rational constraint, and so to morality, outside the context of mutual benefit [MbA 16].

Here, Gauthier displays the virtues of candor and consistency, rather than plausibility. Such a view about the scope of morality is morally revisionary. For instance, it's a common view that morality has wide scope; it imposes obligations of beneficence or aid where the agent stands to gain nothing from the cooperation or restraint of the beneficiary. So, for instance, on this view (a) there can apparently be no obligations to future generations, and (b) if the wealthy and talented have sufficient strength and resources so as to stand to gain nothing by participating with the weak and handicapped in a system of mutual cooperation and forbearance, the former can have no duties of aid, however modest, toward the latter. These are morally repugnant conclusions.