A PUZZLE ABOUT THE RATIONAL AUTHORITY OF MORALITY

David O. Brink
Massachusetts Institute of Technology

Commonsense morality recognizes various other-regarding duties to help, or forbear from harming, others. Most of us also regard moral obligations as important practical considerations that give agents reasons for action. But heeding these obligations may constrain the agent's pursuit of his own interest or aims. If we associate rationality with the agent's own point of view, we may wonder whether moral conduct and concern are always rationally justifiable. These thoughts reveal a tension in ordinary thinking about morality between living right and living well. That tension might be represented as a puzzle about the rational authority of morality that consists of a quartet of claims that can seem individually plausible but are mutually inconsistent.

I. Moral requirements—including other-regarding obligations—apply to agents independently of their aims or interests.
II. Moral requirements necessarily provide agents with reasons for action.
III. Reasons for action are dependent on the aims or interests of the agent who has them
IV. There is no necessary connection between other-regarding action and any aim or interest of the agent.

Each element of the puzzle articulates a familiar and initially plausible idea.

(I) is a statement of the objectivity of ethics, familiar from Kant, according to which moral requirements are categorical norms, they apply to us independently of our antecedent desires and interests. The agent's own interests and inclinations are not, in the relevant way, among the conditions of application for moral requirements. This is clearest in the case of other-regarding moral requirements. I do not defeat an ascription of obligation to me to help another by pointing out that doing so will serve no goal or purpose that I have

(II) captures the practical character of morality and moral deliberation. Agents typically engage in moral deliberation in order to decide what to do;
people give moral advice with the aim of guiding others' conduct; and most of us are quite sensitive to moral criticism. One explanation of these attitudes and expectations is that we think moral requirements give agents reasons for action. Such claims are sometimes defended as conceptual truths about morality. **Internalists** say that to be under a moral obligation to do something just is to have a reason to do it; they assume that norms of morality just are norms of rationality.¹

(III) articulates the common assumption that a fact must affect the agent in some way to be of rational significance to her. Call this assumption about rationality **agent-relative.²** By contrast, rationality is **agent-neutral** just in case another's interests give an agent reasons for action directly, independently of any connection with her own interests or desires.³ Agent-relative assumptions seem to underlie many formal and informal discussions of individual rationality in philosophy, economics, and politics. Moreover, an agent-relative theory provides a reliable link between reasons for action and motivation, we expect one who recognizes reasons for action to be motivated to act on them, and an agent seems more likely to be motivated by facts about his own interest or desires than by facts about the interest or desires of others. Also, when we explain an agent's behavior as an attempt to satisfy certain desires, given her beliefs, we are said to “rationalize” her behavior.⁴ This suggests that genuinely rational behavior is that which would promote the agent's desires or at least those desires that she would have if she met certain epistemic conditions.

(IV) reflects a common assumption about the independence of different people's interests and attitudes that I will call the **independence assumption**. Of course, agents often do care about the welfare of others and desire to do the morally correct thing, and there will often be connections between an agent's own interests and those of others. But neither connection holds either universally or necessarily. My aims could be largely self-confined, and my own good can be specified in terms that make no essential reference to the good of others, say, in terms of my own pleasure or the satisfaction of my desires.

Despite their individual appeal, not all four claims can be true; we must reject at least one element of the puzzle if we are to avoid inconsistency. Indeed, we can make sense of a number of familiar positions at the foundations of ethics as tacit responses to this puzzle that reject one element of the puzzle in order to preserve others. **Moral relativism** and **minimalist** moral theories, such as ethical egoism, reject the existence of categorical moral norms asserted in (I); they claim that moral requirements must further the agent's interests or desires in some way. **Externalists** reject the internalist assumption about the rationality of morality in (II). They distinguish between norms of rationality and norms of morality and recognize moral requirements such that failure to act on them is not
necessarily irrational. Kantians and others reject the agent-relative assumptions about reasons for action in (III); they claim that the interests of other people can directly and necessarily provide an agent with reason for action. Finally, metaphysical egoists reject the independence assumption in (IV) and resolve the puzzle by arguing that, properly understood, people’s interests are interdependent in ways such that acting on other-regarding moral requirements promotes the agent’s own interests.

I’ll sketch some of these solutions and their resources, limitations, and interrelations. My primary goal is not defend a particular solution to the puzzle, but to show that it provides a fruitful framework within which to address some foundational issues about ethics.

1. Relativist and Minimalist Solutions

Some writers begin with assumptions (II)-(IV) in place and conclude that we must reject the existence of categorical, other-regarding requirements.5

For example, in his article “Moral Relativism Defended” Gilbert Harman relies on Humean assumptions about the instrumental nature of rationality and internalist assumptions about the connection between moral obligation and reasons for action in order to defend a view that he calls moral relativism, according to which an agent’s moral obligations must be relativized to her pro-attitudes.

Formulating this as a logical thesis, I want to treat the moral “ought” as a four-place predicate (or “operator”), “Ought (A, D, C, M),” which relates an agent A, a type of act D, considerations C, and motivating attitudes M [1975 10]

Harman claims, for instance, that Hitler is someone to whom obligations of fairness, decency, or respect for human life could not have applied because he (Hitler) lacked the relevant attitudes necessary for him to have had reasons to be fair, to be decent, or to respect human life (1975: 7-11).

Harman’s argument has the following form

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

This is a version of relativism, because it relativizes moral obligations to people’s variable aims. It is a version of minimalism, because it holds the content of one’s obligations hostage to one’s interests and aims and so seems unlikely to recognize the normal range of other-regarding duties that common-sense morality does.

Moreover, this view can be represented as an agent-relative ethical theory if it claims that promoting an agent’s interests or desires is a sufficient as well
as a necessary condition for moral obligation (1975: 11). An agent-relative ethical theory claims that an agent is obligated to do something just insofar as it would promote his own interests or desires.

Another view that can be represented either as an agent-relative ethical theory or as a skeptical view is Callicles' position in Plato's Gorgias. Callicles resolves the apparent conflict between the demands of justice 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). And, precisely because the revision in ordinary moral views that he makes is so drastic—his account of virtue is so minimalist—Callicles is usually thought of as a moral skeptic, even though he accepts the rational authority of real virtue.

Libertarian views might be motivated in a similar way. They recognize negative other-regarding obligations to forbear from harming others and interfering with their liberty but no positive obligations of mutual aid. Robert Nozick, for example, defends a moral theory incorporating libertarian side-constraints against utilitarianism by appeal to the separateness of persons. He writes

Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm. Why not, similarly, hold that some persons have to bear some costs that benefit other persons more? But there is no social entity with a good that undergoes some sacrifice for its own good. To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person, that his is the only life he has. He does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him. [1974 32-3]

Nozick concludes that the separateness of persons grounds libertarian rights that protect individuals from other-regarding moral requirements that demand uncompensated sacrifices. His argument seems to have the following form.

1. It is unreasonable to demand uncompensated sacrifices.
2. Moral requirements must be reasonable.
3. There is no interpersonal compensation; benefits to another do not compensate me for my losses.
4. Hence moral requirements can include no other-regarding duties.

However, we might notice that libertarian minimalism is imperfectly supported by this argument. For, while she recognizes no positive duties of mutual aid that apply independently of the agent's own desires and choices, the libertarian does recognize negative duties to refrain from force and fraud that apply independently of the agent's own aims or interests. Even the libertarian thinks that negative duties of forbearance are other-regarding categorical norms.
Any agent-relative ethical theory that holds an agent’s obligations hostage to the promotion of her immediate or unreflective interests or desires seems committed to a highly revisionary form of minimalism. Some agents are concerned about others. But such other-regarding attitudes are not universal. Moreover, benefiting others consumes time and resources that might have been spent in more self-confined ways. As long as people are psychologically malleable, we may ask why those who do not have other-regarding attitudes should cultivate them and those who happen to have them should maintain them. Unless there are answers to these questions agent-relative ethical theories will be quite minimal.

A sophisticated agent-relative ethical theory might try to avoid minimalism by arguing that it is in the long-term interest of agents to develop, maintain, and act on other-regarding attitudes. Such a theory would be a version of ethical egoism; it would claim that an agent is morally obligated to do something just insofar as that course of action would promote his own interest overall.6 The ethical egoist’s explanation of other-regarding moral requirements would exploit the idea that morality is concerned with the appropriate terms for personal and social interaction and cooperation and requires people to restrain their pursuit of their own aims and interests and accept a fair division of goods and resources. Each individual has an interest in the fruits of such interaction and cooperation. Though it might be desirable from a self-interested point of view to reap the benefits of others’ forbearances 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 towards agents who are known to be noncompliant. So noncompliance secures short-term benefits that compliance does not, but compliance secures greater long-term benefits than noncompliance.

Because of the generally beneficial character of cooperative and restrained behavior, together with the cognitive and affective advantages of acting from fairly coarse-grained dispositions, people will have reason to develop and act on social sentiments and other-regarding attitudes. These attitudes will also receive external support. Because each has an interest in others’ cooperation and restraint, communities will tend to reinforce compliant behavior and discourage noncompliant behavior. Community pressure, therefore, will also foster the development of fairly coarse-grained compliant dispositions. It’s no accident, therefore, that people will have social sentiments and other-regarding attitudes, and these will give them agent-relative reason to act in other-regarding ways. And where they don’t yet have these attitudes, they will nonetheless typically have agent-relative reasons to act in other-regarding ways (if only) as part of a process of developing such attitudes.

The strategic egoist reasons as follows.
1. It is in the agent’s interest to receive the benefits of systems of cooperation and restraint.
2. The benefits of systems of cooperation and restraint are available only to those who maintain the appearance of cooperation and restraint towards strategic partners.
3. The least costly means of maintaining the appearance of cooperation and restraint is by being cooperative and restrained.
4. Hence it is in the agent’s interest to be cooperative and restrained towards strategic partners.
5. Hence there are other-regarding duties that have agent-relative foundations.

In this way, the strategic ethical egoist tries to explain why one should both cultivate and maintain other-regarding attitudes by appeal to one’s long-term interests. This form of ethical egoism attempts to justify other-regarding morality and avoid unacceptably minimalist conclusions while denying (I)’s claim that these moral requirements apply to agents independently of their interests and desires.7

But as long as ethical egoism recognizes the independence assumption, it must remain revisionary. In circumstances in which an agent would benefit from acting on selectively noncompliant dispositions, the ethical egoist cannot ascribe other-regarding moral obligations. First, (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 others’ compliance, though fairness typically seems to require compliance. Moreover, if the stakes are sufficiently high in a particular case, and one’s strategic partners have already complied, one may have no reason to maintain even the appearance of compliance. For the benefits of noncompliance in such a case can outweigh the costs of being excluded from future interaction. Second, (3) is sometimes false. Selective noncompliance may go undetected where it is difficult to monitor compliance and detect deception. But surely the moral obligation to comply does not cease just because successful deception is possible. Third, (4) doesn’t support a sufficiently strong version of (5). Noncompliance towards those with whom one does not strategically interact will be in one’s interest. So if the wealthy and talented have sufficient strength and resources so as to stand nothing to gain by participating with the weak and handicapped in a system of mutual cooperation and forbearance, then the strategic egoist can recognize no duties of mutual aid that the former have towards the latter. Finally, even if the strategic justification of other-regarding conduct were extensionally adequate, the independence assumption ensures that there are possible cases in which an agent has no strategic justification for compliant behavior, if only because the stakes are so high, her discriminatory capacities are so fine-grained, or she is such a successful deceiver. While our other-regarding obligations are presumably counterfactually stable, the strategic justification of other-
regarding conduct is not. These limitations in strategic ethical egoism are all limitations in the *scope of morality itself*, and not simply limitations in the rationality of other-regarding conduct and concern.

One aspect of viewing minimalist moral theories as, perhaps tacit, solutions to this puzzle about the rational authority of morality is that it establishes a link between metaethical and normative issues and so undermines the common claim that metaethics and normative ethics are completely independent of each other. The link is simply that if one accepts certain metaethical claims—viz. internalism, agent-relative assumptions about reasons for action, and the independence assumption—then one is committed to a particular, indeed, extremely controversial kind of moral theory, viz. some form of minimalism. Conversely, if one is to avoid these normative views, one must reject at least one of three metaethical claims.

2. Externalist Solutions

Some accept the existence of categorical moral norms, agent-relative assumptions about rationality, and the independence assumption and so reject the internalist assumption about the rationality of morality in (II). This externalist solution denies that it is a condition of the application of a moral requirement that it provide the agent to whom it applies with a reason for action. While moral requirements apply to us independently of our antecedent desires and interests, they give us reasons for action conditional on their promoting our interests or desires.

We might motivate this externalist position by noting an ambiguity within Kantian rationalism. Kant, of course, distinguishes between hypothetical and categorical imperatives (1785: 414). Hypothetical imperatives are conditional on whether the conduct enjoined promotes the agent’s antecedent interests or desires, while categorical imperatives are not. Kant claims that moral requirements express categorical, rather than hypothetical, imperatives (1785: 416). But we might identify two distinguishable claims here corresponding to two distinct senses in which an imperative can be categorical. In one sense, imperatives are categorical just in case they apply to people independently of their aims or interests. Imperatives are categorical in this sense insofar as they generate *categorical norms*. (I) asserts that moral requirements—including other-regarding moral requirements—are categorical norms. Imperatives are categorical in another sense just in case they provide those to whom they apply with *reasons for action* independently of their desires, aims, or interests. Imperatives are categorical in this sense just in case they generate *categorical reasons*. Categorical reasons are agent-neutral reasons, and other-regarding moral requirements could provide categorical reasons only if there are agent-neutral reasons for action.
8 / David O. Brink

Once we distinguish these two Kantian theses we may not find them equally plausible. We might agree that moral requirements are categorical in the first sense—they are categorical norms—but deny that they are categorical in the second sense—they do not generate categorical reasons.

However, the internalist might argue that we cannot separate categorical norms and categorical reasons.

1. Moral requirements apply to people categorically (i.e. independently of their contingent desires and interests).
2. If a moral requirement (categorical) tells me to do x, then I have a (categorical) moral reason to do x.
3. If I have a (categorical) moral reason to do x, then I have a (categorical) reason to do x.
4. Hence moral requirements generate (categorical) reasons for action.
5. Hence it would be pro tanto irrational of me to fail to act on moral requirements (regardless of my interests or desires).

But the externalist can reply that the argument trades on an equivocation between different senses of ‘reason’. In one sense (a) for there to be a reason for me to do something is simply for there to exist the relevant sort of behavioral standard or norm. In this sense, there are as many kinds of reasons as there are systems of norms; there are moral reasons, reasons of etiquette, reasons of state, etc. In this sense, the existence of categorical moral norms obviously does imply the existence of moral reasons and, hence, reasons. But when we call these reasons categorical reasons, we are just giving another name to what we had previously called categorical norms. The other sense of ‘reason’ (b) signifies more than the existence of a certain sort of behavioral norm; it indicates that there is a reason to behave in accordance with such a norm such that failure to behave in that way is ceteris paribus or pro tanto irrational. If there is reason, in this sense, to act on a norm, then practical reason endorses this norm. And it is this sense of reason and rationality that is at stake in the rational authority of morality and the debate about whether all reasons are agent-relative. Reasons, in this sense, do not follow from the mere existence of certain sorts of norms, as the case of etiquette seems to demonstrate. Norms of etiquette apply to my behavior independently of my interests or attitudes, but failure to observe them does not seem irrational unless this in some way undermines my interests or aims. So the externalist is justified in concluding that the argument trades on an equivocation between these two senses of reason: the argument for (4) is sound just in case it is understood as (4a), while (5) follows from (4) only if (4) is read as (4b).

In a similar way, the externalist may appeal to these two senses of ‘reason’ to dissolve the puzzle. On a univocal (b)-reading of ‘reason’, the externalist denies (II). On an (a)-reading of (II), (II) follows trivially from (I) and is not
an independent premise. The externalist won’t deny (IIa). But (III) clearly requires the (b)-reading of ‘reason’. But, of course, the four claims do not form a genuine paradox if we read (II) as (IIa) and (III) as (IIlb).

We can better see this challenge to Kantian rationalism by considering Philippa Foot’s useful analogy between morality and etiquette (Foot 1972). At some points it may be difficult to distinguish rules of etiquette and rules of morality; for instance, etiquette often enjoins the same sort of consideration and respect for others that morality does. Because we want to examine morality’s relation to something agreed to be fairly unimportant, the focus on etiquette must be on those rules of etiquette that do not overlap with moral requirements, for instance, rules requiring that invitations addressed in the third person be answered in the third person. So we must compare morality and mere etiquette (those rules of etiquette that are not also moral rules).

Rules of (mere) etiquette, like moral requirements, are categorical norms. The moral duty to help others in distress, when you can do so at little cost to yourself, does not fail to apply to you—we do not withdraw our ascription of obligation to you—just because you are indifferent to your neighbor’s suffering and in a hurry to read your mail, as would be the case if it was a hypothetical norm. In the same way, rules against replying to a third-person invitation in the first person don’t fail to apply to you—we don’t take back our ascriptions of duties of etiquette to you—just because you think etiquette is silly or you have a desire to annoy your host, as would be the case if rules of etiquette stated hypothetical norms.

But rules of etiquette seem to lack rational authority; they generate hypothetical, not categorical reasons. Rules of etiquette may state categorical norms, but failure to observe these norms does not seem irrational unless this in some way undermines the agent’s interests or aims. Here too moral requirements may seem on a par with requirements of etiquette. If the independence assumption is correct, obligations of forbearance, mutual aid, and justice will sometimes further no aims or interests of the agent. Though we need not withdraw the ascription of obligation in such cases, perhaps we should allow that immoral conduct in such a case is not irrational. We can reproach such a person for immoral behavior, but not for irrationality.

But the analogy between morals and manners provides no explanation of the common belief that morality has a special authority. Now the externalist might conclude that the special authority of morality is just an illusion—an artifact of moral education and the internalization of moral norms (as Foot may seem to). But he might claim that morality and etiquette are imperfectly analogous. While alike in generating categorical norms whose rationality is hypothetical, not categorical, they need not be equally authoritative systems of norms. This is because of the content of the two types of norms. Different systems of norms make different sorts of requirements and have different points or organizing principles. The externalist might claim that the moral
point of view has a more intimate and regular relationship to people's important interests and aims than the point of view of etiquette does.

The basic idea rests on familiar claims about what the moral point of view is (and what the point of view of etiquette is). One version appeals to claims about morality's other-regarding concerns with fairness, equality, and impartiality and the sort of strategic reasoning that we saw an ethical egoist might invoke. However, this sort of strategic reasoning figures here within a rational egoist attempt to defend the rationality of other-regarding moral considerations, not within an ethical egoist account of the content of morality. Rational egoism is the view that an action is rational just insofar as it would promote the agent's own interests, whereas ethical egoism is the view that an action is morally obligatory just insofar as it would promote the agent's own interests. The rational egoist, who is not an ethical egoist, allows that we can identify the content of moral requirements independently of the agent's own interest and then thinks that moral requirements can be shown to be rational just to the extent that they promote the agent's own interest. The rational egoist holds the rationality of conduct hostage to the agent's own interests and desires; but, unlike the ethical egoist, he does not hold the morality of conduct hostage in this way. Nonetheless, the same sort of strategic reasoning that the sophisticated ethical egoist employs can be invoked by the rational egoist. That reasoning purports to explain why agents have agent-relative reason to develop, maintain, and act on fairly coarse-grained other-regarding dispositions and attitudes. The rational egoist can employ these claims to provide a generally reliable contingent justification of the rational authority or other-regarding morality. We saw that there are actual or at least counterfactual limitations to the strategic justification of other-regarding conduct. But, unlike the ethical egoist, the rational egoist need not view these limitations as limitations in the scope of morality itself. They represent the limits of the rational authority of morality.

By contrast, I assume, the point of view of mere etiquette—whatever exactly it is—has no such intimate and regular relationship to people's important interests and aims. The rational authority of etiquette is not only contingent but weak and unreliable.

If so, the moral point of view is more centrally implicated in human interests and desires than is the point of view of etiquette. These constructive claims suggest how an externalist can try to show that morality has a special authority, not enjoyed by etiquette, while restricting herself to agent-relative reasons.

3. Agent-Neutral Solutions

Another solution to the puzzle is to give up the agent-relative assumptions about rationality in (III) in favor of agent-neutral claims. The trick, of course,
is to make agent-neutral assumptions about reasons for action plausible. We have to explain how I can have reason to do something simply because it benefits someone even if it would further no interest or desire of mine.

We may find agent-neutral assumptions prima facie mysterious. Whereas the link between the agent’s own interests and desires and his reasons for action seems intuitive, the link between the interests of others and his reasons for action may seem less clear or compelling. Moreover, agent-relative assumptions explain better why recognition of reasons for action should motivate the agent, whereas it seems possible to be unmoved by agent-neutral considerations. Further, agent-relative assumptions about rationality seem to be a natural extension of the “rationalizing” mode of explaining behavior, and this allows us to link normative and explanatory senses of ‘reason for action’. By contrast, if an agent has not already formed desires for the welfare of other people, it’s hard to see how the welfare of others could explain his behavior in any way. But then agent-neutral assumptions about rationality will not forge any link between normative and explanatory reasons for action.

Presumably, the friend of agent-neutrality believes that facts about another’s welfare can and will be motivational when the agent understands that and why such facts provide reasons for action. If so, agent-neutrality can maintain a link between explanatory and normative reasons for action. Explanatory reasons consist of the agent’s pro-attitudes, given her actual beliefs. Normative reasons represent an idealization of explanatory reasons. Genuinely rational behavior is behavior that would satisfy the aims the agent would have in an idealized epistemic state in which her various beliefs and desires were in wide reflective equilibrium (cf. Brink 1989: 63-6). If there are convincing arguments for agent-neutral reasons, then this will affect the agent’s aims in reflective equilibrium. If so, there can be no a priori argument from the connection between rationality and motivation or between explanatory and normative reasons against the possibility of agent-neutral reasons.11 We must consider particular arguments for agent-neutrality on a case by case basis. There are two main kinds of argument to consider: intuitive and theoretical.

Consider first an intuitive argument. Agent-relative theories of rationality assume that rational action must benefit the agent or further her aims. But surely, we might think, we have good reason to help others in distress or prevent harm to them if we can do so at little cost to ourselves, even if helping them won’t benefit us or further our aims. Suppose that with a little effort I could warn you, before our jump, that your parachute is jammed or that I could easily step around your gouty toes. Surely, I have reason to warn you and avoid your toes even if doing so doesn’t benefit me or further my aims.12

This claim has considerable force.13 But the externalist who accepts agent-relative assumptions about rationality can accept it on one reading. She can agree that there is a moral reason for me to warn you about your jammed
parachute and walk around your gouty toes, even if doing so does not benefit me or further my aims, and that I am immoral if I fail to do so. So she agrees that I have a reason to do these things in the sense that there is a categorical norm that enjoins such conduct. She denies only that there is a categorical reason for me to do these things. Assuming that I would receive no benefit from these actions, it would not be irrational of me to fail to warn you or avoid your gouty toes. What is intuitively clear is that it would be (grossly) immoral of me to let you jump or to tread on your gouty toes; it is not comparably clear that this would also be necessarily irrational.

Kant and neo-Kantians also offer theoretical arguments for recognizing agent-neutral reasons. One argument can be found in Thomas Nagel's *The Possibility of Altruism*. Nagel sees a parallel between intertemporal and interpersonal distribution of benefits and harms. He argues for agent-neutrality or altruism by analogy with prudence. Just as the interests of an agent's *future* self provide him with reasons for action now, so too, Nagel argues, others' interests provide him with reasons for action. Failure to recognize prudence involves temporal dissociation—failure to see the present as just one time among others—and failure to recognize altruism involves personal dissociation—failure to recognize oneself as just one person among others. Both kinds of dissociation are mistakes. The parity of time and person within rationality, therefore, requires accepting agent-neutrality.

But the rational egoist will not be very impressed by this analogy with prudence. Rational egoism assumes that *sacrifice requires compensation*, that is, that an agent has reason to make a sacrifice, say to benefit another, if and only if the agent receives some sufficient benefit in return. Nagel himself accepts the "extremely strict position that there can be no interpersonal compensation for sacrifice" (1970: 142). But if sacrifice requires compensation, prudence and altruism must be importantly disanalogous. For, in the prudential case, I am compensated for a sacrifice of my present interests in favor of my greater future interests; these future interests are mine. Because benefactor and beneficiary are the same, diachronic, intrapersonal compensation is automatic. But interpersonal compensation is not automatic; benefactor and beneficiary are distinct. If the independence assumption is correct, the interests of other selves, however great, are not *ipso facto* interests of mine. Unless there is some connection between my interests and those of others (as, of course, there will in fact often be), I am not compensated when I sacrifice my interests (present or future) for those of others. But then justified concern for my own future does not itself establish justified concern for others.

A more promising strategy is to develop the Kantian attempt to link categorical norms and categorical reasons. Kant thinks that moral requirements are categorical norms, because they apply to us insofar as we are rational beings and independently of our contingent interests and inclinations (1785: 408, 411, 425-7, 432, 442). If so, moral requirements apply to us in
virtue of our rational features, and these are essential to our being agents who deliberate and possess reasons for action. If so, such requirements arguably provide agents with reasons for action independently of their contingent interests and inclinations. Any norms that apply to us in this way would generate categorical, agent-neutral reasons for action.

It remains to be seen whether there are any such moral requirements and whether they include familiar other-regarding duties. Kant thinks so. He understands the first formulation of the Categorical Imperative to require that one act on maxims that a rational being as such could will to be universal (1785: 421, 425-7). He thinks that the one thing that an agent would choose for its own sake insofar as she is rational, and independently of her contingent inclinations and interests, would be the realization of rational agency. If I choose rational agency solely insofar as I am a rational being, then I will choose to develop rational agency as such, and not the rational agency of this or that being, in particular, not just my rational agency. If so, then reason directs me to be concerned about other rational agents, as rational agents, for their own sakes. Kant concludes that insofar as we are rational beings we would will that all rational agents be treated as ends in themselves and never merely as means (1785: 429); this is his second main formulation of the Categorical Imperative.

The second formula imposes other-regarding duties. It prohibits treating anyone merely as a means. The negative requirement that no one be treated as a mere means requires that each be treated as an end, and this carries certain positive duties. For to treat other rational agents as ends requires treating them as agents whose deliberation and agency is valuable. This requires, ceteris paribus, not simply that we refrain from doing things that would harm the interests and agency of others but also that we do things to promote their rational agency. And this will involve a concern to promote or assist, where possible, others’ opportunities for deliberation and agency, the effectiveness of their deliberations, and the execution of their choices and commitments (cf. 1785: 430).

This brief sketch of the Kantian strategy suggests a way of understanding the categorical application of other-regarding moral norms such that these norms generate categorical, agent-neutral reasons for action.

1. Moral requirements apply to people insofar as they are rational agents.
2. Hence moral requirements supply agents with reasons for action.
3. Insofar as an agent is rational, she will be concerned with rational agency.
4. Insofar as an agent is concerned with rational agency, she will treat rational agents as end in themselves.
5. Hence moral requirements include other-regarding duties to treat others as ends.
6. Hence agents have reason to act on other-regarding duties.

This is not the place to defend the details of this strategy, but it is a solution to the puzzle worth exploring further.\textsuperscript{14}

4. Metaphysical Egoist Solutions

An agent-neutral solution would be unnecessary if the independence assumption were false. But that assumption looks very plausible as applied to the connection between other-regarding action and the agent's desires or pro-attitudes. We've already noted that other-regarding attitudes seem neither universal nor necessary. But there is a more abstract argument for denying that there is a necessary connection between any other-regarding action and any pro-attitude of the agent that appeals to the systematic plasticity of pro-attitudes. There are various connections between facts and people's pro-attitudes. Given some of my pro-attitudes, certain facts or their recognition may commit me to other pro-attitudes. If I want to hire a suitable person for the position, and you are the only suitable person available, then, if I realize this, I will want to hire you for the position. But this sort of connection is not a necessary connection between a fact or belief and a pro-attitude. For in response to the fact or belief, it is always possible to change the pro-attitude that had previously made the fact or belief relevant, rather than adopt the new desire that the fact or belief recommends. So in a context in which some background pro-attitudes are held fixed there are necessary connections between facts or beliefs and other pro-attitudes. But these connections are not necessary simpliciter, because systematic modification of one's pro-attitudes in light of the facts or one's beliefs is always possible. If so, this aspect of the independence assumption is secure.

But this does not rule out the possibility of necessary connections between other-regarding actions and the agent's interests, because we can construct non-conative conceptions of people's interests. If people's interests were interdependent in certain ways, this would undermine the independence assumption. We might call such a view \textit{metaphysical egoism}. The metaphysical egoist claims that when the agent's own happiness or interest is correctly understood we will see that the good of others is, in the appropriate way, \textit{part} of the agent's own good so that acting on other-regarding moral requirements is a way of promoting his own interests. If so, the rational egoist will be able to explain the rational authority of other-regarding moral requirements on agent-relative assumptions.

Metaphysical egoism is sometimes defended as part of absolute idealism, which claims that everything that there is exists as part of a single cosmic or divine consciousness. For then what we think of as distinct lives (distinct streams of consciousness) would stand to cosmic or divine intelligence as we
now believe that the stages of a single life stand to the person whose life it is. There would be a super-personal entity that is both benefactor and beneficiary in the interpersonal case in much the same way that we take the person to be both benefactor and beneficiary in diachronic, intrapersonal sacrifice (contrary to Nozick’s claim, quoted above). While this would make intradestiny compensation automatic, it would not yet establish interpersonal compensation, because the person who is benefactor is not also beneficiary. And without interpersonal compensation, rational egoism will not recognize the rational authority of other-regarding demands. If absolute idealism is to establish interpersonal compensation, it must hold not only that there is a super-personal entity but also that persons are essentially parts of this super-personal entity such that their welfare is to be understood in terms of its. This, I gather, is roughly the view of idealists such as Hegel, Green, and Bradley.15

But, stated baldly, these metaphysical assumptions are likely to seem implausible. The general strategy, however, is also familiar from Greek philosophy. In the Republic Plato attempts to show that justice is in the agent’s own interest, properly understood, because justice is part of having a well-ordered soul. In Nicomachean Ethics ix 4-12 Aristotle argues that friendship, which is the virtue appropriate to communities and includes the perfection of justice (1155a22-8, 1159b25-1160a8), is a virtue that promotes the agent’s own happiness, because the virtuous friend is “another-self” (1168b2-6, 1170b6-9). If the friend is another-self, then benefiting her presumably benefits me. And if political communities are associations of friendship, then perhaps I can take this attitude towards others as well. If so, perhaps I can expect interpersonal compensation and so will have agent-relative reason to comply with other-regarding moral requirements.

One attempt to unpack these claims relies on familiar, though not uncontroversial, metaphysical claims. Indeed, I think that the basic strategy is Aristotelian.16 Aristotle claims that we can justify concern for one’s (best or complete) friends and family members (e.g. children and siblings) as cases of, or on the model of, self-love (1161b15-1162a5, 1166a10).

The excellent person is related to his friend in the same way as he is related to himself, since a friend is another self, and therefore, just as his own being is choiceworthy for him, the friend’s being is choiceworthy for him in the same or a similar way [1170b6-9]

One way to understand these claims is as a proposal to model the relationship between “other-selves” (e.g. intimates) on the relationship between a self and its temporal parts. I have reason to regard my intimates as other-selves, because they bear approximately the same relationship to me as future stages of myself bear to me, and this fact provides me with reason to care about them.

According to rational egoism, concern for my own future is rational; concern for my own future is concern for me, and I am compensated for
sacrifices of my present self on behalf of my future self, because my future self is a part of me. But how must a future self be related to my present self in order for both to be parts of me? A common and plausible answer (which I shall employ but not defend) is psychological continuity. On this view, a particular person consists of a series of psychologically continuous person stages. A series of person stages is psychologically continuous just in case contiguous members in this series are psychologically well connected. And a pair of person stages is psychologically connected just in case they are psychologically similar (in terms of such things as beliefs, desires, and intentions) and the psychological features of the later stage are causally dependent upon the earlier stage. On this view, self-love would seem to imply that I should be concerned about selves that are psychologically continuous with my present self.

But I can be psychologically continuous with other selves with whom I share a mental life and interact causally. Interpersonal, as well as intrapersonal, psychological continuity is possible. There will be psychological continuity between any people who share beliefs, values, and goals as the result of their causal interaction. This will be true to a significant extent in the case of intimates, such as spouses and friends. It will also be true to an interesting, though lesser, extent among members of the same community, because they have common goals and aims and because these shared goals have been produced at least in part by mutual discussion and interaction. Interpersonal psychological continuity can also be indirect, because it can hold between people who are not themselves connected but who are each connected to others in common. It can also be indirect when it is mediated by social institutions and practices (e.g. media and legal and political institutions) that otherwise isolated people both participate in. In these ways, interpersonal psychological continuity can extend quite broadly, even if the degree of continuity often weakens as it extends further.

To the extent that distinct individuals are psychologically continuous, each can and should view the other as another-self who extends her own interests in much the same way that her own future self extends her interests. If so, there can be automatic interpersonal compensation among other-selves just as there is automatic diachronic, intrapersonal compensation. One will have agent-relative reasons to promote the welfare of another proportional to the degree of psychological continuity one bears to her.

Degree of continuity will presumably affect the amount of other-regarding conduct and concern that can be justified in this way. Presumably, I have less agent-relative reason to benefit comparative strangers than my intimates, precisely because the former are less psychologically continuous with me than the latter. In this way, metaphysical egoism seems committed to a discount rate of rational concern proportional to the degree of psychological continuity the agent bears to others. But this need not be a threat to morality,
because the extent of one's obligations to others is commonly thought to be a function not simply of the amount of benefit that one can confer but also of the nature of the relationship in which one stands to potential beneficiaries. Commonsense morality recognizes more stringent obligations towards those to whom one stands in special relationships (e.g. to family and friends and to partners in cooperative schemes) than towards others. Even impartial moral theories, such as utilitarianism or consequentialism, typically try to justify recognition of special obligations and the legitimacy of differential concern for those to whom the agent stands in special relationships. The scale of stringency among our moral obligations to those to whom we stand in different relationships forms a moral discount rate. It remains to be seen whether the moral discount rate and the metaphysical egoist's discount rate of rational concern are isomorphic, but it is not a defect per se of its justification of other-regarding concern that it embodies a discount rate of concern, because there is a moral discount rate.

Because it is within one's power to affect the degree of psychological continuity one shares with others, there is a question whether one should cultivate or maintain other-selves. The metaphysical egoist can claim that having another-self extends my interests in important ways. A plausible conception of welfare can reasonably claim that it is in my interest to exercise those capacities that are central to the sort of being I am and that these capacities include, importantly, deliberative capacities (cf. EN i 7). Having another-self provides unique opportunities to exercise my deliberative capacities. My other-self brings me new information. I can learn and benefit from the experiences that my other-self has by adding them to my own. Moreover, deliberation includes practical deliberation. Practical deliberation is exercised in the reflective formation, assessment, and pursuit of projects and plans. By drawing on the experience of my other-self and engaging her in discussion, I improve my own practical deliberations, not just by reaching better informed decisions but also by forming my decisions in a reflective manner. I can also exercise new deliberative capacities by engaging in more complex projects requiring mutual cooperation. In these ways, my other-selves expand my deliberative powers, activities, and control.

These aspects of the way in which another-self extends one's interests constrain the degree of psychological similarity one has reason to seek in such a relationship. My activities will be more diverse and more complex if my other-self is no mere clone of me. Clearly, I will diversify my experience more by interacting with someone who has somewhat different interests and experiences. My deliberations will be aided by input and criticism from new perspectives. And cooperation in complex projects will often be enhanced when participants have different strengths and talents. These considerations provide reason to participate in larger, more diverse groups as well small intimate associations. My experiences will be enlarged and my own practical
deliberations will be enhanced by the input from people with different experiences, values, and perspectives, and larger groups with more diverse membership will typically make possible more complex forms of social cooperation and, hence, the exercise of new kinds of deliberative capacities.

Indeed, the arguments for cultivating another-self show that beneficial interaction with others is itself a way of extending one’s interests. If so, one will have agent-relative reason to benefit others, as a way of establishing another-self, even towards those with whom one is not already continuous.

These are reasons to think that there are agent-relative reasons to cultivate and maintain other-selves, and because there is automatic compensation for sacrifice among other-selves, there is reason to think that other-regarding action can be a necessary part of an agent’s good.

But we may wonder whether the egoist can justify the full range of moral demands, for morality seems to require not just that we perform the actions it demands of us but also that we fulfill its demands from the right sort of motives. Even if egoism can justify the moral demand that I benefit others, it may seem that it cannot account for the moral demand that I benefit others out of a concern for their own sakes. Because the metaphysical egoist justifies sacrifice on behalf of another by representing the other’s good as part of the agent’s own good, this seems to base the agent’s other-regarding action not on concern for the other’s own sake, but instead on self-love.

While the strategic egoist may find it difficult to defend other-regarding concern that is not, at bottom, instrumental, the metaphysical egoist can do better. The metaphysical egoist’s argument for cultivating other-selves appeals to a deliberative conception of individual welfare and claims that the right sort of other-regarding relationships extend my interests by giving me opportunities to exercise more diverse and complex deliberative capacities. This argument justifies acquiring other-selves by appeal to its effects on one’s capacities, but the capacities in question involve, among other things, one’s relation to other people in cooperative and beneficial ways. So if I’m justified in entering such relationships because they exercise these sorts of capacities, my justification does not obviously reflect a purely instrumental attitude towards those with whom I’m entering such relationships.

Moreover, once my relationship with others is under way, psychological continuity begins to get established, and I can begin to see their welfare as part of my own. But if this is true, then my concern for them will not be purely instrumental. Recall the intrapersonal parallel. When I undergo a present sacrifice for a future benefit I do so because the interests of my future self are interests of mine; in this way, I make present sacrifices out of a concern for the sake of my future self. Of course, the on-balance rationality of the sacrifice depends upon its promoting my overall good. But, because the good of my future self is part of this overall good, concern for my overall good requires, as a constituent part, a concern for the good of my future self.
In this way, concern for my future self for its own sake seems compatible with and, indeed, essential to self-love.

Now the metaphysical egoist wants to model interpersonal continuity and concern on intrapersonal continuity and concern. Just as the agent’s future self is a part of her, so too the interests of other-selves are part of her interests. And so just as egoism explains why the agent should be concerned about her future self for its own sake, insofar as it is continuous with her present self, so too it explains why she should be concerned about her other-selves for their own sakes, insofar as they are continuous with her.

My account of metaphysical egoism has appealed to both deliberative conceptions of welfare and psychological continuity accounts of personal identity. These may seem to be independent lines of argument. One could apparently justify cultivating and maintaining other-regarding relationships by appeal to deliberative conceptions of welfare without invoking the idea of another-self, and one could apparently appeal to interpersonal psychological continuity as a justification of other-regarding conduct even if one were, say, a hedonist. While these two lines of argument are somewhat independent, they are complementary and together strengthen metaphysical egoism. The deliberative conception of welfare plays a distinctive role in explaining why I should cultivate and maintain other-selves, while the parallel between interpersonal and intrapersonal psychological continuity provides a non-instrumental account of why I should be concerned about people to whom I am so related. Moreover, each line of argument arguably supports the other. We should expect to find interdependence between deliberative conceptions of welfare, deliberative conceptions of personhood, and psychological continuity accounts of personal identity in terms of reasoned control and modification of beliefs, desires, and intentions (psychological continuity, properly interpreted). Because personhood is itself a normative category and because psychological continuity is an account of what matters in personal identity, our views about these matters ought to affect our views about welfare or happiness. If what it is to be a person is to have certain deliberative capacities and what it is to be the same person over time is, roughly, for there to be reasoned continuity of intentional states, then we should expect one’s exercise of deliberative capacities to be an important ingredient of one’s welfare, in part because it will extend one’s interests. And to the extent that deliberative activities seem to be principal ingredients in a good life, we should expect reasoned continuity of intentional states to be a principal ingredient in what matters in personal identity and deliberative capacities to be principal ingredients in personhood. In this way, the metaphysical egoist’s dual appeal to deliberative conceptions of welfare and interpersonal psychological continuity promises to be a virtue.

The metaphysical egoist can provide agent-relative justification of other-regarding demands by showing how another’s good can be part of the agent’s
own good. Whether an agent will always have an overriding reason to fulfill every other-regarding moral requirement is another matter, requiring further articulation of both the nature of morality's other-regarding demands and this neo-Aristotelian version of metaphysical egoism.

5. Solutions

I have tried to show how some familiar views about the foundations of ethics can profitably be seen as, perhaps tacit, solutions to the puzzle. This perspective may help us better assess these views.

For instance, when we view certain kinds of relativist and minimalist theories as solutions to the puzzle, we can both see their rationale and locate their implausibility. These theories are motivated by an attempt to preserve the other intuitively plausible elements of the puzzle. But rejecting the existence of categorical other-regarding duties (norms), while maintaining the independence assumption seems the least plausible response to the puzzle. It would be reasonable only if the sole alternative was to insist that moral requirements have no more rational authority than requirements of mere etiquette. But this is not so.

Suppose we accept not only (I) but also (III)'s claim that all reasons for action are agent-relative and (IV)'s independence assumption; this would require us to be externalists and reject (II). But this would not show that moral requirements had no more authority than the requirements of etiquette. For a rational egoist can employ certain forms of strategic reasoning to show how agents have generally reliable, albeit contingent, reasons to fulfill other-regarding requirements of forbearance, cooperation, and mutual-aid. And in those actual or merely possible circumstances in which acting on other-regarding moral requirements would not further the interests or aims of the agent, we can maintain the immorality of failure to act on those requirements even if we cannot maintain its irrationality. So, even if we had to accept (III) and (IV), it would still be more reasonable to reject (II) than (I).

What's objectionable is the combination of an agent-relative ethical theory and the independence assumption. An agent-relative moral theory that denied the independence assumption need not be unacceptably minimalist. Metaphysical egoism tries to establish that people's interests, correctly understood, are interdependent in ways that ensure that other-regarding conduct and concern necessarily promote the agent's own interests. I presented metaphysical egoism as a version of rational, rather than ethical, egoism—that is, as an account of rationality and the rational authority of other-regarding moral requirements, rather than as an account of the content of morality. But the resources of metaphysical egoism are available to the ethical egoist who insists that moral requirements must themselves be agent-relative. Insofar
as metaphysical egoism can be articulated so as to justify other-regarding conduct and concern, an agent-relative moral theory incorporating metaphysical egoism can deny (I) while avoiding clearly unacceptable minimalist commitments. Indeed, we might conclude that the rejection of categorical other-regarding norms, asserted in (I), is plausible only if it is neither relativist nor minimalist; it will avoid relativism and minimalism only if it incorporates metaphysical egoism.

Of course, if we take (I) to be reasonably fixed, our views about the plausibility of any other element of the puzzle will vary inversely with our views about the plausibility of the other two. So, for instance, if we accept or hold reasonably fixed the claim that there are other-regarding requirements and the independence assumption, then we can see that the externalist denial of (II) and the agent-neutral denial of (III) will be inversely plausible. If, under these assumptions, we also think that all reasons for action are agent-relative, we must conclude that it is at least possible for there to be circumstances in which it would be immoral but not irrational for an agent to fail to fulfill her other-regarding moral requirements. We will, therefore, think (II) is false.

And, similarly, if, under these assumptions, we also assume that moral requirements must supply reasons for action, we should think that not all reasons for action can be agent-relative. This will lead us to reject (III). If we reject (IV)'s independence assumption, then, whether we accept agent-relative or agent-neutral assumptions about rationality, we should accept (II)'s claim that moral requirements do supply reasons for action (though we needn't think, as the internalist does, that this is a conceptual truth about morality).

Moreover, we may pool resources. The friend of agent-relativity can combine the resources of both strategic and metaphysical egoists to provide a strong rational egoist defense of the rational authority of other-regarding moral demands. Even the friend of agent-neutrality may wish to enlist the aid of strategic and metaphysical egoists. For even if we reject (III)'s claim that all reasons for action are agent-relative and accept the existence of agent-neutral reasons, we are unlikely to think that all reasons for action are agent-neutral. There will still be agent-relative reasons for action. But this raises the possibility of conflict between agent-neutral reasons that support other-regarding moral requirements and agent-relative reasons that may not. If we are to vindicate the importance of moral requirements, agent-neutral reasons must not be systematically overridden by countervailing agent-relative reasons.

This suggests the need to distinguish stronger and weaker rationalist theses. A weak rationalist thesis claims that there is always some reason to fulfill moral requirements such that failure to do so is to that extent irrational. The strongest rationalist thesis would claim that for every agent there is always overriding reason to fulfill moral requirements such that failure to do so is on-balance irrational. And, of course, there are intermediate rationalist theses.
In fact, the strength of the rationalist thesis will affect our formulation of the puzzle. We might revise our formulation of the puzzle so that (II) reflects a strong rationalist thesis

I. Moral requirements—including other-regarding obligations—apply to agents independently of their aims or interests.

II. Moral requirements necessarily provide agents with compelling or overriding reasons for action.

III. Reasons for action are dependent on the aims or interests of the agent who has them.

IV. There is no necessary connection between other-regarding action and any aim or interest of the agent.

A solution to the initial puzzle that establishes the existence of agent-neutral reasons to be moral may not solve the revised puzzle. The defense of agent-neutral reasons and the rejection of purely agent-relative assumptions about rationality, though significant claims, are not sufficient to vindicate the rationalist element of the revised puzzle. Unless agent-neutral reasons are necessarily superior reasons, the best solution would be to argue that agent-relative reasons, properly understood, support other-regarding moral requirements as well. So friends of agent-neutrality would do well to cultivate the resources of strategic and metaphysical egoists, even if they reject the rational egoist assumption that all reasons for action are agent-relative.

Further investigation of the puzzle would require more comprehensive investigations of (III) and (IV). Are there compelling arguments for agent-relativity? Can Kantian arguments for agent-neutrality be articulated plausibly? Are people’s interests independent, as the neo-Aristotelian view claims? If so, do other-regarding moral requirements and other-regarding aspects of an agent’s welfare dovetail appropriately? How strong is the authority of moral demands when the various resources for justifying other-regarding conduct are counted together? When we answer these questions, we will have the resources for a satisfying solution to the puzzle about the rational authority of morality.20

Notes

1. Because internalism assumes the norms of morality are norms of rationality, it implies, in effect, that (II) is an a priori constraint on moral theory and is not up for grabs. So understood, internalism is a ground for asserting (II). Internalism is sometimes understood as a claim about moral motivation, whereas the kind of internalism that supports (II) is a claim about the rational authority or morality. For a discussion of various forms of internalism, see Brink 1989 ch 3. Internalism about moral motivation is relevant to a puzzle about moral motivation.
A Puzzle About the Rational Authority of Morality / 23

1. Moral judgments express beliefs
2. Moral judgments imply motivation
3. Motivation involves pro-attitudes
4. There is no necessary connection between any belief and any pro-attitude

Noncognitivists (e.g., Stevenson and Hare) reject (I), externalists (e.g., Foot and Nagel and McDowell) deny either (III) or (IV). Though I do not agree with their solutions, this second puzzle is usefully discussed by McNaughton 1988 and Smith 1989.

2. In some contexts it is important to distinguish the egoist claim that I have reason to do something insofar as it would promote my interests or advance my overall good and the instrumentalist claim that I have reason to do something insofar as it would satisfy my present desires. Among other differences, the latter has an important temporal restriction that the former does not. But both are agent-relative views, and my present purposes do not in general require that they be sharply distinguished. I will mark the distinction only where it seems relevant to my argument.

3. As I am understanding it, agent-neutrality says only that other people’s interests can directly provide me with reasons for action, independently of any connection with my interests or aims, it says nothing about the scope or weight of various directly other-regarding reasons. So agent-neutral theories can be quite varied.

For instance, Sidgwick’s rational benevolence and Broad’s self-referential altruism are both agent-neutral (cf. Broad 1953 279-280).

6. My account of strategic egoism does not distinguish between act egoism and motive egoism. That’s a topic for another occasion.

7. Cf. Epicurus Kurau Doxa 31-38, Taylor 1987, Axelrod 1984, Gauthier 1986, and Frank 1988. Though each accepts a form of rational egoism incorporating such strategic reasoning, Epicurus and Gauthier are clearest about employing such reasoning as part of ethical egoism and its attempt to derive the content of morality from agent-relative rational choice.

There are various significant constraints on the models of social interaction in which cooperation and restraint are strategically rational. The model must be dynamic, actors must not discount future pay-offs too much, the environment must include a sufficiently large number of conditional compliers, and conditional compliers must be able to monitor reliably the compliance and noncompliance of others (this last condition tends to be more easily satisfied in small groups).

8. The Ring of Gyges exposes the counterfactual instability of strategic egoism, this is why Plato insists that justice be shown to be valuable for its intrinsic, and not simply extrinsic, consequences (Rep 359b-361d). However, Republic 1-11 appears to be concerned with a rational egoist challenge to the rational authority of justice, not with ethical egoism. If so, it differs from the Gorgias, in which Callicles presents an ethical egoist challenge to conventional justice.

9. Indeed, Gauthier is quite candid about these limitations in his argument. His justification of other-regarding conduct among actors in strategic situations depends on the assumption that such actors are “psychologically translucent” (1986 173-4), and he acknowledges that he cannot explain duties to those with whom one has no strategic interaction (1986 16, 17).


11. Any view, such as this, that treats rational action as action that would satisfy the aims the agent would have in an idealized epistemic state in which her beliefs...
and desires are in reflective equilibrium might seem to be an agent-relative view about rationality. If so, it may seem that no arguments for agent-neutrality within reflective equilibrium could unseat agent-relative assumptions about rationality, the rationality of other-regarding conduct would still depend on facts about the agent’s aims or desires (cf. Williams 1980). But this a priori defense of agent-relativity has two problems. First, this kind of agent-relativity is not a constituent element of the puzzle. If reflective equilibrium might be such that aims that are in reflective equilibrium must include direct concern for others, then there is no problem about the rational authority of morality on the kind of agent-relative assumptions characteristic of this reflective equilibrium account of rationality. If we understand the agent-relativity of (III) in this way, there is or need be no inconsistency in (I)-(IV). To put it another way, the kind of agent-neutrality sufficient to resolve the puzzle is not inconsistent with agent-relativity in this sense. Second, whether the reflective equilibrium account of rationality is agent-relative depends upon whether the arguments for agent-neutrality, if good, are good arguments because they’re in reflective equilibrium or whether they’re part of reflective equilibrium because they’re good arguments. If the former, the rationality of other-regarding conduct is agent-relative, if the latter, it is agent-neutral. The friend of agent-neutrality presumably advances arguments for agent-neutrality the merits of which are intended to explain why agent-neutrality should figure in reflective equilibrium. As such, these arguments, if successful, would vindicate agent-neutrality.

12 The example involving gouty toes is Hume’s (1751 ch V, pt ii), though his concern is with egotist theories of motivation, not rationality.
13 It will have even less force against agent-relative assumptions about rationality if we reject the independence assumption (cf. §4 below).
14 I explore this strategy more fully in Brink 1991.
16 What follows is a condensed version of claims that I defend at greater length in Brink 1990, cf. Irwin 1988 ch 18.

This argument is also Platonic. Plato claims that (1) the virtuous person will have love for others (Rep 402d-403c, 412d), and (2) love (a) involves concern for the beloved’s own sake, and (b) is, for the lover, the next best thing to immortality (Symp 206e-212c and Phdhr 243c-257b) (Cf. Irwin 1977 241-3). Plato’s grounds for accepting (b) seem very similar to Aristotle’s reasons for regarding one’s friend as “another-self” (see below) (1)-(2) allow Plato to argue that the virtuous agent will be concerned about others for their own sakes and will necessarily benefit from benefitting them.

Also, there is a reading of this project, which I won’t pursue, that bears resemblance to the idealist strategy mentioned above.

19 Psychological diversity is compatible with psychological continuity provided the diversity is the result of interaction in which the participants exchanged and discussed beliefs, goals, and values. For then differences result from a process in which there were common intentional states produced by causal interaction.
20 Work on this paper was done during a fellowship at the Center for Advanced Study in the Behavioral Sciences that was funded by an Old Dominion Fellowship from the Massachusetts Institute of Technology and by grants from the National...
A Puzzle About the Rational Authority of Morality

Endowment for the Humanities (RA-20037-88) and the Andrew W. Mellon Foundation I would like to thank these institutions for their support I would also like to thank Randy Calvert, Stephen Darwall, Thomas Hurka, and David McNaughton for helpful discussion of some of these issues.

References

Aristotle Nicomachean Ethics (EN), trs T. Irwin Indianapolis Hackett, 1985
Bradley, F H 1927 Ethical Studies Cambridge, MA MIT Press
Brink, D O 1991 “Categorical Imperatives and Kantian Rationalism” Unpublished
Brink, D O 1990 “Rational Egoism, Self, and Others” In O. Flanagan and A. Rorty (eds), Identity, Character, and Morality Cambridge, MA MIT Press
Frank, R 1988 Passions within Reason New York Norton
Green, T H 1883 Prolegomena to Ethics New York Crowell, 1969
Harman, G 1984 “Is There a Single True Morality?” In D. Copp and M. Zimmerman (eds), Morality, Reason, and Truth Totowa, NJ Roman and Littlefield
Hegel, G W F 1840 Introduction to the Philosophy of History, trs L. Rauch Indianapolis Hackett, 1988
Hegel, G W F 1821 The Philosophy of Right, trs T. M. Knox Oxford Clarendon Press, 1942
Hume, D 1751 An Enquiry Concerning the Principles of Morals Indianapolis Hackett, 1983
Kant, I 1785 Grounding for the Metaphysics of Morals, trs J. Ellington Indianapolis Hackett, 1981 (Prussian Academy pagination)
Mackie, J L 1977 Ethics Inventing Right and Wrong New York Penguin
McNaughton, D 1988 Moral Vision Oxford Blackwell
Nozick, R 1974 Anarchy, State, and Utopia New York Basic Books
Plato Republic (Rep), trs G. Grube Indianapolis Hackett, 1974
David O. Brink


