Philosophy and Phenomenological Research
Vol. LXVII, No. 1, July 2003

Anti-Consequentialism and the Transcendence of the Good

ROBERT MERRIHEW ADAMS
Yale University

Richard Boyd's "Finite Beings, Finite Goods"[1] is exactly the sort of response a philosopher hopes to evoke. It is perceptive and fair-minded in its reading and criticism of my work, illuminating the agreements and disagreements and the motivations on both sides, and showing points at which my position stands in need of more adequate development. At the same time it is much more than a response, offering a fuller and richer development, on several points, of what was already, in my opinion, the most plausible and promising version of naturalist moral realism. Here I will try to respond in a similar spirit, though on a smaller scale, taking up first issues about consequentialism, and helplessness, and then more metaethical issues, involving what I have called "the critical stance."

I should say something at the outset about my aims in engaging Boyd in debate, as I did already in my book (Adams 58-82).[2] I was and am arguing, of course, against Boyd's version of naturalism. But I was not and am not aiming at a definitive refutation of his view. His is one of relatively few constructive nontheistic theories in metaethics against which I argue at any length in *Finite and Infinite Goods* and that is precisely because his theory is so close to mine in some ways that seeing why I differ with it may be particularly helpful in understanding the motivation of my own view. I have not wanted to try to prove my own theory by refuting all the main competing types of metaethical theory and then saying, 'See now; my view wins by default.' The problem with that strategy is that *all* known metaethical theories are attended with grave problems. No doubt each theory that's still seriously in the running is commended in part by the difficulties of the others. But relatively little is gained by dwelling on those problems, because most of them are pretty well known already. The hard project, and the only

1 *Philosophy and Phenomenological Research*, 66 (2003): 505-53, and 67 (2003) 1-24. This work will be cited here by parenthetical references, in the text and notes, to "Boyd," with page numbers.

2 References in this form are to Robert Merrihew Adams, *Finite and Infinite Goods: A Framework for Ethics* (New York: Oxford University Press, 1999).

one likely to contribute very much to our understanding in these matters, is trying to make one's own metaethical approach work somewhat better than it did before. That's what I've been trying to do on behalf of supernaturalism, and that's what I take Boyd to have been trying to do, with more than a little success, on behalf of naturalism.

1. Consequentialism

Consequentialism is a more central and more focal topic in Boyd's essay than it is in *Finite and Infinite Goods*. There are a number of places in the book where, as I note in one of them, "I have not resisted the temptation to take a potshot at utilitarianism," but have not attempted a "refutation" of it (Adams 223n). Boyd is right, however, in observing that opposition to consequentialism is a fairly pervasive theme in the book, and that it can be linked to my rejection of naturalism. Indeed, I begin with it in the present context because I have come to agree with him that it sheds light on our metaethical disagreements. That this is so is one of the points on which Boyd has shown me something about my work of which I was not fully conscious in writing it.

Boyd observes, with becoming caution, that "some version or other of consequentialism has seemed to many to be the best candidate for a non-debunking naturalistic conception of morality" (Boyd 31). He does not claim (nor would I) that naturalism strictly entails consequentialism, but his consequentialism can be seen as lending support to his naturalism in at least two ways. It may help us to distinguish these ways if we use, as I did in my book, a broader, more abstractly structural conception of consequentialism than Boyd uses. He reserves the name 'consequentialism' for theories that take the well-being of persons, or human flourishing, as their primary concern (Boyd 31). But I take consequentialisms simply to be theories according to which the moral evaluation of some sort of action or state of a person or social system is to be determined by the value of its (actual or reasonably expected) consequences. Consequentialism as such, as I understand it, does not imply anything about how the value of the consequences is to be determined, except that the moral value of the type of object that is subjected to the test of consequences is not to be treated as an independent variable.

An ecocentric ethics, for instance, might be no less consequentialist for taking as its primary concern the integrity of the biosphere rather than the well-being of humans. Likewise a perfectionist ethics might be consequentialist in taking as its primary concern the physical, intellectual, and artistic prowess of the most talented humans rather than their or anyone else's well-being as such. Unlike such ecocentric or perfectionist forms of consequential-

ism, Boyd's is what we may call a *welfarist*[3] consequentialism, one in which the value of consequences is understood in terms of human well-being. The consequentialist and the welfarist character of Boyd's theory support his naturalism in distinguishable ways.

Its consequentialism supports his naturalism by virtue of the obvious fact that according to consequentialism the moral value of acts and states of human agents and social systems is determined by *causal* properties of some such acts or states—specifically by the value of consequences they *cause* or tend to cause. Boyd's naturalism aspires to identify the (natural) properties (or property clusters) to which normative and evaluative terms refer on the basis of *causal* roles of those properties. These causal roles are complex—too complex, in Boyd's opinion, to permit reductive definitions of the terms. Among the causal relationships involved in the roles are both those that help explain the "achievements" of the disciplines or belief-forming practices in which the ethical terms are used, and those that constitute the "homeostatic" or mutually supportive character of the property clusters to which, for the most part, the ethical terms refer. One main obstacle to reductive definition of ethical properties is that their roles prominently include relations to other ethical properties (Boyd 28). The vindication of the causal character of their roles still demands, however, that the main non-analytic relations among ethical properties be understood as at bottom causal. Consequentialism obviously helps in satisfying this demand, inasmuch as it maintains that one important family of normative and evaluative properties is to be understood on the basis of direct or indirect causal relations to another family of evaluative properties.

There remains the more fundamental problem of identifying, in terms of their causal role, the evaluative properties of the consequences, which are basic in this scheme. An important part of Boyd's strategy for doing this is his supposition that the terms in which the consequences are evaluated refer to a single homeostatic, and therefore causally interrelated, cluster of natural properties. He characterizes that cluster as constituting human well-being or human flourishing, and I think this welfarism helps him toward his goal of a fundamentally causal account. That is because human well-being is a matter of how human lives go, and one of the things that makes human lives natural units for ethical reflection is that virtually all their features, or all their ethically salient features, are bound together in a rich system of (mostly close) causal relations. The concept of human well-being therefore stands a rela-

---

[3] *Utilitarianism*, in any of its forms, is also a consequentialist view that is focused on human well-being, and one which weighs the well-being of each human person equally, as I assume Boyd would; but utilitarianism does not allow, as I assume Boyd might, that the optimization (perhaps equalization) of the *distribution* of well-being may be weighed independently, in competition with maximization of the sum or average of human well-being, in determining the value of consequences.

tively good chance of pointing us toward natural properties that cluster causally in the way that Boyd wants them to.

He might find himself in more difficulty at this point if he were a non-welfarist consequentialist who included among morally important consequences of fundamental value both human well-being and the preservation of biodiversity—each for its own sake and independently of the other, to some extent. For the ways that human lives go, on the one hand, and the biodiversity of (for example) arctic or desert ecosystems, on the other hand, do not form with each other such a strongly integrated system of causal relations as each manifests within itself. So if the most fundamental category of value (for a consequentialist, the value of good consequences) is shared by such diverse phenomena, there seems to be less likelihood that it will have the kind of *causal unity* that Boyd wants it to have.

1.1 Well-Being and Excellence

In developing my own views on these topics in response to what seems to me now to be Boyd's position, I will begin with issues connected with welfarism and then go on to issues connected with consequentialism in the broader sense. Boyd discusses at some length (Boyd 24-30) my chapter on well-being, which is in fact a chapter on "Well-Being and Excellence." In writing that chapter I was not primarily trying to raise problems for consequentialism or naturalism. Indeed, I think it had not occurred to me that I could be raising problems for naturalism there—although now that Boyd has made the connection, I can see that's a motive I could have had. My principal aim was to develop and vindicate the place of *excellence* in ethical theory, by arguing that a much more fashionable sort of value, human well-being, actually depends on excellence. Specifically I believe that human well-being is best understood as consisting in a life characterized principally by *enjoyment* of the *excellent*—the more excellent, and the more it is enjoyed, the better for the person; and I argue that this belief is supported by intuitions that can hardly be rejected from the point of view of one who enjoys life (Adams 93-101).

Boyd concurs with me in rejecting analyses of well-being in terms of counterfactual conditionals about preferences (Boyd 24). The sort of naturalist but non-reductionist account he proposes as a better alternative is interesting, illuminating, and beautifully integrated with his version of moral realism. Because he does not attempt "a full-blown analysis," it is not clear to me whether his account would allow excellence to have the kind of role in well-being that I think belongs to it. I think he has some reason not to allow it (as I'll explain presently), but he could consistently allow it. He says that he takes "terms used to describe aspects of well-being" to refer to natures that "could be characterized not only in terms of features of the psychology of the

individuals whose well-being is at issue, but also in terms of the relation between well-being and other morally relevant phenomena like goodness, fairness, sympathy, justice, ..." (Boyd 28). If he would insert excellence (including artistic and intellectual excellence, for example) into this open-ended list of values related to well-being, he might well end up with an account that, though naturalist, would be quite consistent with the argument of my chapter on "Well-Being and Excellence," which is not meant to entail my supernaturalism.

This is not to say that it is irrelevant to supernaturalism. The vindication of excellence as the primary form of value is important to the theocentrism of my ethical theory, for it is as intrinsically excellent, and not merely as *our* good, let alone as instrumentally good, that God is worshiped. I believe the vindication of excellence is similarly important for the ethics of Plato, for whom the forms of the Beautiful and the Good are to be loved, and imitated, for their intrinsic excellence, and of Aristotle, who conceives of human good or *eudaimonia* (happiness or flourishing) as activity of the soul in accordance with *arete* (virtue or excellence),[4] and by the same token as imitation of God. As suggested by the conceptions of primal or transcendent excellence in these theories, the objects to which excellence belongs are not chiefly states of affairs, and the best response to excellence is not necessarily one of trying to bring about or prevent a state of affairs, but may often be one of loving, admiring, or (in the supreme case) worshiping something that already exists, or perhaps sometimes, more sadly, mourning something that no longer exists. These facts pose obvious issues for anyone who would integrate excellence into a consequentialist framework, in which the consequences whose value is foundational are states of affairs (cf. Adams 17); but that is not my project, and I will set those issues aside.

Given what I have said about the dependence of well-being on excellence, I agree with Boyd, of course, that excellence, including artistic and intellectual excellence, can contribute to human well-being, and that is an important part of what we expect of excellence. I would emphasize that on my view, excellence's contribution is not merely causal but constitutive of well-being; I'm not sure whether Boyd will agree with that, but here I will simply present my own view rather than pursue this question about Boyd. The relations between excellence and well-being are rich and complex. The excellence that is enjoyed in well-being can be the excellence of one's own activity, as suggested by Aristotle's account of human good, but it can also be the excellence of something one knows, something that may even be beyond the reach of one's agency, as in the beatific vision of God (or of the form of the Good). Artistic and intellectual excellence can be enjoyed, as part of one's good, in both of

---

4 Aristotle, *Nichomachean Ethics*, I,7 (1098a16-17).

these ways. And the love of such impersonal excellences will not normally be neatly separable from love for persons.

Why do you want your child to enjoy great literature? Because you love your child and you love literature. Moral psychology shaped by consequentialism is apt to propose a single object for your motivation in this case: the state of affairs consisting of your child enjoying great literature; and I'm supposing that you do indeed want that state of affairs. But I don't suppose that you love it; I imagine rather that you want it for the sake of two other objects, of quite different types, that you do love, *each for its own sake*, your child and literature. Your love of literature will normally show itself in connection with your caring about some person (yourself if none other). You don't care that much about the great books just sitting there on the shelf, unread. But equally you don't just want your child to enjoy something or other; you want her to enjoy great books (or good music, as the case may be).

While my views about excellence as a fundamental type of value, more fundamental than human well-being, may not be strictly inconsistent with Boyd's naturalism, it should be clear at this point that they are unhelpful to it in precisely the way that his own welfarism is helpful to it. The things that I take to be excellent—great books, beautiful music, heroic deeds, the dynamic complexity of a living body or an ecosystem—are very diverse, and so are the ways in which they are excellent. I take beauty, moral virtue, and the aptness of a response or interpretation all to be types of excellence. Perhaps there are causal relations among all these things, but they will hardly form such an integrated causal system as Boyd expects of a homeostatic property cluster. Excellence seems much less likely than human flourishing to have a *causal* unity. I think the various types of excellence find their unity in all being ways of resembling or imaging God, but that solution is obviously not available to a naturalist.

1.2 Consequentialism and Helplessness

Forms of consequentialism differ, among other ways, in the objects they subject to consequentialist evaluation, which may be acts, social policies, systems of moral education, types of conscience, motives, or traits of character. The test of consequences can be applied directly or indirectly. In an "indirect consequentialism" about acts, for example, something else—a system of ethical practices, perhaps—is evaluated by the value of its consequences and determined to be the best, and acts are determined to be right if and only if they are acceptable according to that system of practices. Boyd and I agree that a plausible theory of the moral value of human acts can be at most indirectly consequentialist. We disagree about whether even an indirect consequentialist theory should be accepted.

I think some indirect consequentialisms are pretty good, from a practical point of view, yielding sensitive and approximately correct answers to substantive moral questions much of the time. Still I am not strictly a consequentialist of any sort. My most fundamental reason for not being even an indirect consequentialist is that I do not see why it is only the values of consequences that should be considered relevant (directly or indirectly) to the moral values of actions, attitudes, motives, and traits of character. In particular, I do not see why we should not suppose, as I believe in practice we do, that the character of actions, attitudes, and traits as ways of being *for* the good is morally valuable independently (to some extent) of their actual or probable effectiveness (individual or generic) in promoting the good.

This is connected with the disagreement between Boyd and me about the importance of moral advice to the helpless. I think this disagreement is real, and some of the things Boyd says about it are extremely interesting. He has given us some fascinating indications of what I would call his eschatology (cf. Adams 220-25); and I think his idea that improving our social systems can be seen as a way of integrating the good is a very good idea—though only up to a point, as I don't think that all important values can be integrated around the values of human flourishing, as I've already noted. But Boyd is right, in a way, to suggest that my view is more pessimistic than his.

Not that I regard myself as pessimistic on the whole. Theism, I think, is an optimistic view; and I am enthusiastic about many actually existing goods. I may actually be more optimistic than Boyd is about the possibilities (and indeed the historic actualities) of widespread enjoyment of natural, artistic, cultural, and spiritual values in social and economic circumstances that are far from ideal. Though there are certainly people who are deprived in various ways of cultural opportunities, one need not be rich or even prosperous to enjoy excellences of poetry, music, dance, and design of everyday objects; and that is a historically and morally important fact.

Where our prospects for improving things through social action are concerned, however, Boyd does seem to me more optimistic than I am. I am not very optimistic about the likely results of "social engineering." This is partly because ideas of persistent human sinfulness seem to me empirically plausible, and partly because I think we are not very good predictors of the consequences of our actions—not even very good predictors of the consequences of social policies, which turn out so often to have unexpected and undesirable side-effects.

But that sort of pessimism about the likelihood of social engineering achieving its aims is not what I want to stress here. For even if social action were to accomplish what we might reasonably want it to accomplish, this side of fantasy, I believe that helplessness would remain an important part of human life, as I have claimed that it is. In our social paradise, will death have

been eliminated?[5] Will all of us always agree about everything, or will there still be social conflicts, and will those on the losing side never feel helpless? Will people never take chances, and if they do will they never fail and feel defeated? In such ways as these—absent brainwashing that I wouldn't want to buy into—I think there is bound to remain plenty of helplessness.

Let's focus on *political* helplessness, which I suspect is the kind Boyd most hopes to reduce or eliminate. Am *I* more helpless politically in my actual social situation than I ought to be? I doubt that very much; in fact I suspect that I am less helpless politically than I would be in a more ideal social world. I am only one of about six billion people on the planet, among whom I am in many ways privileged. I live in a society in which I have extensive rights of political participation. I am well educated to make use of those rights, and I have more than an average share of my society's—not to mention the world's—material resources, from which I can, and do, make political contributions. There is some unfairness in that, and I am one of those whom it favors.

My political influence, nonetheless, seems pretty puny to me, and I see plenty of political helplessness in my own actual situation. Among my six billion human contemporaries, there are lots of diverse and in some ways conflicting views about how things ought to go, and I don't think that's bad. Even if some of the existing views are perverse, some such pluralism seems to me an inevitable result of freedom, and not in itself undesirable. Indeed, I regard it as positively related to the value of moral and spiritual quests, individual and communal, in human life. But given such a diversity of views, we can't all get our way. And absent abuses of power, none of us is likely to get our way entirely. As only one among so many, my personal influence on political outcomes, in a better world, would probably be quite small. Even in an ideal, if not utopian, social world, there would be plenty of political helplessness for practically everyone.

So I think helplessness is a permanent human problem, and a satisfactory complete ethics should offer some guidance for the helpless. No form of consequentialism seems to me to do this well. The non-consequentialist thought that at least in a symbolic way I can still be for, or even "stand for," the good that I care about can provide important meaning for my life, even when I cannot expect to bring about good consequences. And why shouldn't it, given the obvious connection between standing for and meaning? The indirect consequentialist thought that the social practice to which I am conforming is likely to have good consequences even though my individual action won't does not seem to me similarly inspiring—unless it is covertly, after all, the

---

5 Boyd seems to think not (Boyd 518).

thought that in conforming to such a valuable practice, I am at least standing for the good.[6]

These remarks express a concern for what I have called, not altogether happily, the "meaning" of our individual lives. This is a typically, though not exclusively, religious concern. I say *religious*, not just *theistic*, because I think something of the sort can be equally characteristic of nontheistic forms of religion. Such religious concerns have sometimes led to a political quietism that I would not endorse. Religion ought to support our moral duty to be concerned about political goods and to participate in the political life of our society to the extent that is appropriate in the light of our situations and vocations. It seems to me unreasonable and unrealistic, however, to expect most human beings to find the main moral significance of their lives in their political contributions. The personal political impact of almost all of us is too small and too uncertain for that. We must look elsewhere for much of the moral significance of our lives. This will be partly, no doubt, in the nearer and less uncertain consequences of our lives in the lives of individuals who are close to us; but I think we can ill afford to ignore the possible symbolic and non-consequential value of our simply being *for* the goods that are important to us.

2. Metaethical Issues

The metaethical disagreement between Boyd and me that is most directly related to our disagreement about the place of helplessness in human life and ethics is about the relation of ethics to the natural, and particularly the social, sciences. The latter disagreement is deeply implicated, in turn, in our disagreement about what I have called "the critical stance."

2.1 Ethics and the Natural Sciences

Boyd and I agree that we "differ about the extent to which moral reasoning and scientific theorizing should resemble each other." Boyd rightly adds that this "difference stems ... from different estimates about the sorts of problems we face" in trying to achieve goods ("in trying to care for each other," he says, but that reflects his welfarism, and my conception of the ethical aim is broader, as I have indicated above). Boyd is "inclined to believe that these problems stem largely from unfortunate features of social, economic, and political structures," and therefore "inclined to think that moral theory is in

---

[6] Boyd seems to take me as finding fault with "someone who refrained from symbolic opposition under circumstances in which she was morally certain that there would be no favorable consequences," and he proposes to be "more generous" (Boyd 38-39). But I do not mean to find fault with such a person, as I have tried to make clear with reference to Bonhoeffer (Adams 216, 225). My aim in my chapter on "Symbolic Value" was rather to give a rationale for admiring people who do choose the symbolic action, despite the obvious consequentialist reasons for not admiring them.

large measure a branch of political economy, *properly done*" (Boyd 531). Here our disagreement about the extent and ethical significance of human helplessness clearly enters in. Hoping for much less from social engineering, and hence from social science, in this area than Boyd does, I am correspondingly less ready to conceive of the role of ethical properties in terms of their role in the achievements of the social sciences.

I think this disagreement is not necessarily about the prospects for *epistemological* success of social science in general, but is probably about the specifically *ethical* payoff to be expected from social science, for moral understanding and moral life. I suspect further that both this disagreement and that about the place of helplessness in human life are rooted in broadly ethical disagreements about what are the significant goods and evils in human life, and perhaps especially what is the character of humanly (and therefore ethically) important aspirations for good. I believe that the evils (or at any rate the frustrations of good) are much more deeply rooted in the natural framework of human life, and the important aspirations push beyond that framework much more, than would fit Boyd's picture of the study of good and evil as a branch of political economy. This is no doubt a broadly *religious* as well as ethical disagreement; and to the extent that it is important to the motivation of my metaethical views, those views do not have religiously neutral grounds—though the view I have just expressed about goods and evils, and aspirations for good, in human life is not in itself *theistic*.

Central to Boyd's assimilation of ethics to natural science is his *causal* criterion for identifying the properties to which the terms of both disciplines refer. He holds that "the natural kind terms deployed within the discourse central to the inductive/explanatory successes" of what he calls a "disciplinary matrix" refer to families of properties if and only if two conditions are satisfied (Boyd 538). The first is what he calls the *epistemic access condition*. He gives it an interesting technical formulation that involves more of his philosophy of natural science than I am prepared to apply to ethics; but I take the core idea to be one with which I do agree, that "we cannot always or even usually be totally mistaken about goodness" (Adams 20), or that "On any plausible naturalistic conception of reference, the relation which use of the term 'good,' in ethical contexts, bears to the family of phenomena which are favorably evaluated by prevailing ethical standards within the relevant society must be *something like* the reference relation" (Boyd 533).

The second condition, which Boyd typically calls the *achievement explanation condition*,[7] is that "in order for [a term] t to refer to [a phenomenon] p, the epistemic access which uses of t affords speakers to the real properties of p must (help to) *explain* the theoretical and/or practical successes achieved in the domains of inquiry or of practice to which t-talk is central" (Boyd 515).

---

[7] Or occasionally the "accommodation condition" (Boyd 538).

The successes or achievements to be explained may be described as a "tendency ... to identify causally sustained generalizations, to obtain correct explanations, or to obtain successful solutions to practical problems" (Boyd 538). It is one of my main disagreements with Boyd's metaethics that I do not find it plausible to assign the achievement explanation condition such a central role.

Expecting less achievement of this sort than Boyd expects in the natural history of human ethics, I therefore find his *achievement* explanation criterion for the reference of ethical terms less plausible than a criterion that focuses on what would best satisfy (and likewise explain) the best of our *aspirations* for good. This is an argument for conceiving of excellence as approximation or resemblance to a transcendent standard; but it is not an argument from conceptual analysis. It is rather an argument from what appears to me—and has appeared to many others—to be the natural unsatisfiability of our aspirations for good. Alternatively, it may be construed as an argument from the fragmentary and elusive character of the goods and excellences that we find in life, and even of those that we might find in a life that benefited from the wonders of any future political economy. I think we can best do justice to those goods and their character—always limited but sometimes inspiring—by seeing them as images of a goodness that cannot be identified with any natural property or any cluster of natural properties. In all of this it's not obvious to me how far I'm more pessimistic than Boyd about our scientific and political capacities, and how much I'm more optimistic than he is about the possible boundlessness of good.

Even aside from issues of optimism or pessimism about the extent of ethical achievement to be expected from the social sciences, I am skeptical about the adequacy of the basis Boyd offers us for identifying such achievements. I agree with Boyd that moral terms can be, and sometimes are, used "to identify causally sustained generalizations [and] to obtain correct explanations," even broadly causal explanations. And to the extent that the achievements of ethical theory are of this sort, they can presumably be identified in ways characteristic of the natural sciences. But if these achievements are as far as ethical theory gets, I would count it a pretty dismal failure, from an ethical point of view. Such successes of causal reasoning are surely not the main function of moral terms, and I think it is not plausible to rely heavily on achievements of those sorts in assessing the correctness of an assignment of reference to moral terms. It may be more plausible to rely for that purpose on the achievement of "successful solutions to practical problems"; but I do not think that what counts as a (morally) *successful* solution to a practical problem can be determined on such a straightforwardly causal basis. It seems to me that a moral judgment must inevitably be involved in deciding what

counts, from a moral point of view, as a successful solution to a practical problem.

Boyd agrees with me that normative judgements are bound to be involved in evaluating semantic claims in natural science as well as in ethical theory; but he holds that "the only normative judgments involved in the evaluation of semantic claims about natural kind terms are *epistemic* judgments about the cogency of competing explanations for achievements within particular domains of practice" (Boyd 536), and that this is true about the terms of ethical theory as well as of natural science (Boyd 536). This seems to me one of the least plausible (and perhaps also most obscure) parts of Boyd's theory. How do we identify, in the first place, the achievements to be explained? And what distinguishes epistemic normativity from other sorts of normativity or evaluation?

It does not help very much, in my opinion, that Boyd also holds that "the only normative judgments which are implied by [the relevant] semantic claims" in either ethics or natural science "are hypothetical judgments about how to bring about [the relevant] achievements"—for instance, "classify that way if you want to figure out how we can effectively care about each other's well-being" (Boyd 536). Once again it seems to me that we are bound to fall back on *ethical* normativity (or perhaps more precisely, ethical evaluation) in determining what counts as a person's well-being. In accordance with what I have said above about the relation of well-being to excellence, I believe that some of the main evaluations here should be judgments of the *excellence* of activities, relationships, experiences, and forms of knowledge that might be enjoyed. Perhaps Boyd will suppose that if he rejects my views about excellence, he can rely on more purely epistemic forms of evaluation here, though I confess that I do not see how that would work.

One of the most striking but also (in my opinion) puzzling claims of Boyd's paper is that "referential hypotheses about moral terms and the natures of the entities to which they refer" can in principle be confirmed "by the hypothetical extraterrestrial anthropologist who, because of its psychological makeup, can have no commitment whatsoever to moral norms of practice" (Boyd 542). I have no quarrel with this if all that is meant by the lack of moral commitment of the imagined extraterrestrial is that it is incapable of caring about the well-being of us humans. I grant that such an inhumane being might be able to understand our moral language, but that's because I assume it might be able to make broadly ethical evaluations of its own, even if (in one of those evaluations) it considered us unworthy of its concern. My impression, however, is that Boyd supposes the hypothetical extraterrestrial to employ only epistemic norms or evaluations.

How would such a being identify the successes or achievements of our moral discourse? There is much in Boyd's paper to suggest that the uncom-

mitted observer will identify the achievements and successes of practices by reference to the *aims* of practitioners—that is, broadly speaking, by psychological criteria. He holds that "achievements have to have something or other to do with the aims, purposes, interests, intentions, etc. of those practitioners whose achievements they are" (Boyd 543)—which is plausible. What Boyd most wants to emphasize about this is that it's the aims of the practitioners that count, not the aims of those who are trying to discern what the practitioners have achieved. For this reason, Boyd thinks, "a pacifist can offer a referential hypothesis about terms employed in military engineering without endorsing any of the practices of military engineers. She will merely be offering (part of) an explanation for their success in achieving aims she does not endorse" (Boyd 542).

The aims of military engineers, however, are more simply deciphered than those which Boyd would ascribe to scientists and moralists. For he also holds that "that something-or-other" that achievements must have to do with the aims of practitioners "need not be a straightforward matter of bringing about what practitioners want or intend. ...we can achieve things we don't intend, even things we can't conceive of intending" (Boyd 543). Perhaps so; but then we don't have a straightforwardly psychological criterion for which among things that are done, produced, or caused by human actions count as achievements. With what will we supplement the psychological criterion, if not with a normative or evaluative criterion or judgment? And why should we suppose its normativity will be merely epistemic? Perhaps the evaluation of results of the practice of chemistry as being or not being real achievements is merely epistemic, but why should we suppose that is true of the evaluation of results of moral discourse as achievements or not? I suppose I do evaluate the quality of discussions in ethical theory in a way that, if not merely epistemic, does not bring the full force of my own moral commitments to bear. But I think that is not identifying an achievement in any way that might be decisive for reference or truth. If we are engaged in the latter sort of achievement identification, I will not allow my own moral evaluation, where it is strong and clear, to be overridden my merely epistemic evaluations. And that, I think, is as it should be, given our aims in ethical discourse.

Taking an example suggested by Boyd, suppose that among ends at which it seems psychologically possible for humans to aim, the one that a morally uncommitted anthropologist from another planet would find most spectacularly fulfilled by our moral discourse is the convenience of the ruling class (Boyd 548). If we were convinced of that, must we conclude that the good is whatever serves the convenience of the ruling class? I'd sooner give up moral realism altogether than accept that conclusion. But Boyd may be forced to accept it if he is strongly enough committed to as strictly causal a conception as he seems to hold of the role that moral properties are to fill.

Here we are already engaged with the main remaining subject I wish to discuss, that of the critical stance, as I call it.

2.2 The Critical Stance

In *Finite and Infinite Goods* I argue against Boyd's version of ethical naturalism on the ground that "allowing empirical reasoning of the causal explanatory sort to have the last word is incompatible with a stance that is essential[8] to ethical thinking." I call this a "critical stance," and claim that it

> is part of the general intentional framework in which we use evaluative and normative terms, at least where morality and excellence are concerned, and thus affects the semantically indicated role of such terms. The stance amounts to at least this. For any natural, empirically identifiable property or type of action that we or others may regard as good or bad, right or wrong, we are committed to leave it *always* open in principle to raise evaluative or normative questions by asking whether that property or action-type is *really* good or right, or to issue an evaluative or normative challenge by denying that it is *really* good or right. (Adams 77-78)

This, I have suggested, is the kernel of truth in G. E. Moore's famous "open question argument" against naturalism in ethics.

I argue that a theistic metaethics or more broadly, a metaethics based on belief in a transcendent Good, can accommodate the critical stance better than naturalism can, precisely by virtue of the thesis (which might otherwise seem a disadvantage of such a metaethics) that the transcendent Good can never be fully comprehended by any human mind, and thus may always be importantly different from what we think it is. Thus a theistic metaethics can agree with Iris Murdoch that "Good is indefinable not for the reasons offered by Moore's successors, but because of the infinite difficulty of the task of apprehending a magnetic but inexhaustible reality."[9]

In "Finite Beings, Finite Goods," Boyd gives a rich and extensive response to this argument of mine, further developing his own metaethical position in the process. My view of my own position as well as his has changed somewhat in the light of his response, and I want to articulate my revised position here. First I will offer a clarification about the *critical* character of the critical stance. Then I will note some main points on which I think Boyd and I agree, before going on to try to delineate the most important points of disagreement bearing on this argument.

---

8 I now think I probably overreached in claiming that the critical stance is *essential* to ethical thinking. I agree with Boyd that it is not very profitable to try to identify features of ethical thinking that are individually essential to it. What I should have claimed is just that the critical stance is sufficiently characteristic of ethical thinking, and sufficiently supported by typical aims of ethical thinking, that it is a great advantage in a metaethical theory to be able to sustain the critical stance.

9 Iris Murdoch, *Metaphysics as a Guide to Morals* (London: Penguin Books, 1992), p. 42, (quoted in Adams 82).

It is a serious problem, discussed in more than one place in *Finite and Infinite Goods*, how to hold firm moral convictions, on points on which one ought to have them, without failing idolatrously to acknowledge the distinction in principle between one's own convictions about the good, even on the most obvious points, and the good itself. In speaking about a persistently and globally *critical* stance as an antidote to idolatry, I certainly shouldn't mean, and hope I didn't mean, a stance of being more willing than most of us would be to abandon such "obvious" moral beliefs as that it is wicked to torture children.[10] What I think I should mean is something thinner, but still of moral and religious significance: a readiness in principle to think about possible challenges to even the most deeply entrenched of one's ethical beliefs, and to think about them as possible substantive disputes about the good or the right. As a realist in these matters, I believe we should think about these possible disputes as ones in which neither party would be merely expressing emotions, but both would be trying to conform their views to moral facts that are in principle independent of their own thoughts and feelings. A theistic metaphysics of morals supplies this desideratum easily: the standard to which we are trying to conform our views is in God, only imperfectly knowable by us, and thus distinct, more than in principle, from our opinions and feelings on the subject. My challenge to Boyd is whether naturalism can supply the desideratum as well.

Discussion of my challenge can be framed by noting relevant points of agreement between Boyd and me. The first four can be stated very briefly. I take it that Boyd and I agree (1) that evaluative terms such as 'good' are normally used with the assumption that they signify properties, but (2) that it is not a conceptual but a substantive metaphysical question, whether there is in fact a property signified by 'good' (in any important class or type of uses of the word), and if so, what it is—and likewise for other evaluative terms. We agree further (3) that experience is in principle relevant to this substantive metaphysical question, though perhaps we do not agree as to *how far* it is an empirical question. We agree also (4) that if 'good' does indeed signify a real property, there must be a tendency for members of the relevant community of discourse to be approximately right in (much of) what they say about the good, though this tendency may, and almost certainly does, fall hugely short of infallibility.[11]

Nonetheless, we also agree (5) that it is good to have a certain openness to consideration of the possible falsity of any ethical views, including our own; and here matters get more complicated. It is clear that our agreement on this

---

10 Boyd's example—shrewdly chosen, in view of the fact that I have said, and would say again, that it would be crazy to abandon that belief.

11 On this requirement of approximate truth of much evaluative discourse, which is connected with Boyd's "epistemic access condition" on the correctness of identifications of moral properties, see Adams 20, 360, and Boyd 515, 533, 538, 30.

subject is far from complete; we do not endorse just the same *kind* of openness. The openness that I endorse is what I call "the critical stance," and my aim is to explore how that differs from the openness that Boyd endorses; but we should see how far we may agree in this area.

We can certainly agree (5a) on what Boyd calls a "critical-stance-on-the-cheap." Like "scientific and everyday factual generalizations" ethical generalizations "do not follow deductively from the empirical evidence which supports them" (Boyd 514). As Boyd rightly observes, of course, this is much less than I meant in speaking of a "critical stance." But we can agree on more. (5b) Boyd does not think, nor do I, that the most fundamental ethical principles are *analytically* true; the most firmly established, he suggests, may be accepted as *quasi-analytic* truths. And he allows that "quasi-analyticity ... can be altered ... by pretty serious conceptual and theoretical 'revolutions,' whose directions are all but impossible to anticipate prior to the innovations or crises which precipitate them" (Boyd 520). We can agree, then, that no matter how well established an ethical theory may seem empirically (or theologically, as Boyd will add and I must grant), we must acknowledge not merely a logical but in some sense an epistemological possibility of an intellectual "revolution" that would sweep the theory away or (more likely) embed it in a new context that would transform it in unforeseeable ways. But this is fully compatible with Boyd's naturalism, as he has made clear, and is less than I meant by "the critical stance."

Our agreement goes farther, however. (5c) I believe that Boyd's closest counterpart to my critical stance comes to expression when he says,

> Although it is neither possible nor necessary for most of us to conduct morally unengaged investigations of the metaphysics of morals, it is possible—and arguably it is desirable—for us, at some points in our lives, to accept our own moral commitments somewhat tentatively, while still assigning some evidential weight to our moral judgments, until we satisfy ourselves that the referents of 'good' and similarly approbative moral terms are things we actually admire. It is this exercise to which we must turn if we are to have an adequate response to Marxist and other critiques of the social role of morality. (Boyd 545-46)

In the light of this passage I take it that Boyd thinks, as I do, that it is in some sense conceivable that we should find out very disturbing facts about the things to which our own respective theories look for definitive identification of the referents of 'good' (God in the case of my theory; the natural property-cluster to which our actual moral epistemic practice is causally accommodated in the case of Boyd's theory). And we both think that if such news were disturbing enough, a stance (dare I call it a critical stance?) we currently endorse would lead us to withhold or withdraw personal commitment from the objects that must define the good if anything does, according to our respective present metaethical theories. We even agree, I think, (6) that our withdrawal would have important grounds in affective or volitional atti-

tudes of our own. Boyd mentions admiration (or lack thereof); I would mention Eros (or lack thereof); they are close kin.

At this point, however, our disagreements begin. I think we disagree (A) about the form of this withdrawal of commitment. The withdrawal of commitment that Boyd seems to envisage takes the form of refusing any longer to be committed to being for the good, while still believing that there is a real property (or property-cluster) actually signified by 'good'. I find it very hard to imagine myself, or most people I know, responding in that way. In using the word 'good' in ethical contexts I do assume that it signifies a real property that things have or lack independently of my feelings and beliefs, but I also assume that it is a property that gives reason to favor things that have it, and that thus if properly identified would support a commitment to the good. The first of these assumptions is not more important than the second, to me or to the practices in which I use the word 'good'; and I would not cling to the first assumption at the expense of the second. Thus the withdrawal of commitment that I envisage as a live option in a crisis of metaethical belief takes the form of abandoning the belief that there is a real property actually signified by 'good'—or at least of provisionally abandoning it until a more plausible candidate is found. And that is the way in which I have in fact conceived of responding to such remote possibilities as that of my being convinced that God commands us to torture children (cf. Adams 46).[12]

This difference, I think, is at least partly grounded in another important difference (B) between our theories. In keeping with Boyd's insistence that the only sort of normativity that must be presupposed in identifying the properties to which moral terms refer is merely *epistemic* normativity of a sort operative in the natural sciences, causal relationships must ultimately be decisive for him in defining the role that moral properties must fill. I disagree. Following Plato, I take it to be a central feature of the reference-fixing role of excellence that *properly attuned Eros* is responsive to it (Adams 16, 19-28)—so that if I came (appallingly and, in a religious rather than an epistemological way, unthinkably) to think that God could not be an object of properly attuned Eros, that would be a reason for abandoning the view that likeness to God is a real property that can rightly be identified with excellence.

Boyd could attempt to domesticate this idea within his naturalism by maintaining that causal relationships must be ultimately decisive for the nature of proper attunement of Eros. But I do not find that a satisfying view of the matter. An important feature of the system of practices in which ethi-

---

[12] See also Robert Merrihew Adams, *The Virtue of Faith and Other Essays in Philosophical Theology* (New York: Oxford University Press, 1987), pp. 98-102, for a fuller discussion of this point in the context of an earlier and somewhat different version of my metaethical views.

cal terms have their meaning is that we have learned to be responsive (in ways we could not easily describe) to our own feelings and desires in judging of good and evil. This includes a measure of trust, mingled no doubt with caution or distrust, in the stirrings of our own Eros. Supporting or fitting this feature of our ethical practice is part, I believe, of the role to be filled by ethical properties. Just as a correct identification of the good cannot be one on which too few of our actual beliefs about the good are approximately true, as Boyd and I agree, so also the good cannot be something it would be crazy to expect us to love. To put the point in Boyd's terms, the good must not only be something to which we have epistemic access; we must have epistemic access to it in the right way, partly through our Eros and other affective and volitional responses.

Boyd may seem not so from agreeing with this, given what he says about admiration (Boyd 546, 548). But I think in fact the difference in metaethical theory is deep, and shows itself in what Boyd goes on to say about ideology and the possible partial denotation of ethical terms. He allows (indeed, maintains) that 'good' bears at least a *reference-like* relation to such properties as "convenience for the ruling class," in view of the way in which tracking of such properties by actual moral discourse contributes to "explaining certain achievements of those classes ... which occupy positions of disproportionate power and wealth," although I am sure that Boyd himself is not disposed to admire anything for having that property, nor would he admire such a disposition in others. Boyd equivocates as to whether the reference-like relation in this case is "partial denotation" or only "*very very* much like" it (Boyd 548-49); but convenience for the ruling class looks like the sort of property that Boyd supposes might turn out to be the good though he would not admire it (Boyd 545-46).

I grant there is some truth to Boyd's empirical claims about this sort of example—how much truth, there is no point in debating here. I agree with Boyd also that views he and I share about the semantics of morals do not allow us one obvious argument against the claim that 'good' partially denotes convenience for the ruling class; that is, we cannot reject it on the ground that "we do not (most of us anyway) intend to refer to *convenience for the ruling class* ... when we use the term 'good'" (Boyd 549). What properties our terms denote cannot be determined as simply as that by conceptual analysis.

Nevertheless I am not prepared to concede that 'good' denotes, even partially, any such property as convenience for the ruling class. That is in large part because I cannot love such a property or admire anything for having it, and those responses are supported by many particular ethical beliefs that I hold. And my conception of the role to be filled by the property of goodness encourages me to allow my affective reactions to join my actual ethical

beliefs in overruling any purely causal argument for an offensive identification of the property of goodness. I think I must therefore be committed (as I have already indicated that I am) to believing that *moral* normativity and *moral* evaluation need not and should not be avoided in favor of purely epistemic normativity and evaluation in identifying the properties to which moral terms refer, so that I can regard it as metaethically appropriate to reject as *morally* unacceptable the identification of convenience for the ruling class as a sort of goodness in moral contexts.

There is another difference (C) between Boyd and me that I think gets closer to the argument I was trying to make in *Finite and Infinite Goods*. Because the divine nature is conceived as surpassing our comprehension, my theory puts us much farther, epistemologically, than Boyd's does from reaching a place at which we could be forced, in the sort of way I've been discussing, to stop believing in a real property of goodness as a guide of life. Boyd will argue that this is merely a result of a prior metaphysical difference—of my believing, and his not believing, in God; and that is certainly not wholly mistaken. I believe, however, that one need not be a theist to think that it should be harder than it might turn out to be on Boyd's metaethical view to reach a situation in which the only available form of disagreement with a substantive ethical thesis (possibly even a thesis now rather controversial or unpopular[13]) is an amoralist or at least anti-realist position. And therein lies an advantage of supernaturalism over naturalism in ethics—an advantage, more precisely, of metaethical theories organized around a transcendent Good. That this seems an advantage to me doubtless reflects my sense that what we intend in seeking and thinking about the good goes beyond anything that could be defined or delimited by any cluster or set of properties of finite things. That is already a broadly religious sense, but I think it has intuitive "legs of its own" and is not merely parasitic on a framework of theistic belief.

[13] For instance, that moral goodness is the property of being convenient for the ruling class.