Coherence and Models for Moral Theorizing
by
Geoffrey Sayre-McCord

1. Introduction

Modern moral theory has been dominated by a reverence for coherence. Indeed, many have thought that coherence marks the difference between real theory, on the one hand, and mere collections of principles and lists of commands, on the other. Yet because methodological concerns are often left unstated, the role played by coherence has been largely covert. Commonly, its impact is manifested in an intolerance for conflicting principles and a demand for systemic unity. In what follows, I articulate the concept of coherence (so often left merely implicit) and propose a model for moral theorizing which is metaphysically conservative, yet which gives a central place to objective moral truth. Against this backdrop, I argue that the common reverence for coherence has been misplaced.

On the grounds that morality itself is coherent, ethicists have reasonably assumed that the ideal moral theory must be coherent. However, coherence has not reigned merely as a regulative ideal. Ethicists have often also taken the coherence of their favored theory as evidence for its correctness, and have held that our working theories -- the theories we develop and use to guide action -- should be coherent. As a result they have taken coherence as a constraint on the construction of acceptable moral theories.[1] Unfortunately, by requiring that our working theories be coherent, we will insure, I'll argue, that they are wrong.

The first step in making my argument, of course, is to spell out what it is for a moral theory to be coherent. This is especially important since the term "coherence" has had various applications in various contexts. As a first approximation: a moral theory which is coherent (in the relevant sense) systematically covers, without gaps or inconsistencies, all the relevant territory (just how much territory that is depends upon the ambitions of the theory). Put more precisely, to be coherent a moral theory must meet three necessary and jointly sufficient conditions: (i) the principles of the theory must be fully consistent (the "consistency requirement"), (ii) each principle must be connected by a path of justification to all other principles of the theory (the "connectedness requirement")[2], and (iii) the theory must provide an answer for every moral question it recognizes as legitimate (the "completeness requirement"). More or less explicitly each of these criteria has played a significant role in moral theory.

Coherence, so characterized, depends only on the relationships which hold between the principles of a theory; it has nothing to do with how well the theory "coheres" with moral intuitions. So this sort of coherence has nothing directly to do with methods of reflective equilibrium nor with the intuitionists' high regard for common sense moral judgments. A theory may satisfy all three of coherence's requirements and still fail to generate even a single intuitively attractive moral evaluation. Coherence (in this sense) is thus a matter of theoretical a architecture, not of psychological fit.

Traditional utilitarianism offers an especially clear example of a coherent moral theory. In fact, utilitarianism's coherence has proven one of its most attractive features.[3] Each of the three requirements of coherence is met in an admirably straightforward way. First, the utilitarian principle (roughly: "maximize the Good") resolves all conflicts within the theory; so the theory meets the consistency requirement. Second, since the single first principle justifies all the lower level principles and directives, every part of the theory either justifies (in the case of the first principle) or is justified by (in the case of the rest of the theory) other parts and no set of principles is free floating; so the theory meets the connectedness requirement. Third, since the first principle is used to answer all moral questions, the theory offers a complete (though perhaps unsatisfactory) account of morality; so the theory also meets the completeness requirement.[4]

---

1. Among those who accept coherence as an operative constraint on moral theorizing are Kant, Mill, Sidgwick, Donagan, Dworkin, Gewirth, Hare, Harsanyi, and Rawls. Sidgwick, though, is a special case: he defends utilitarianism as the correct moral theory in part by appealing to its coherence, but then finds no principled way to resolve the conflict between utilitarianism and rational egoism and is forced "to admit an ultimate and fundamental contradiction" in Practical Reason. Practical reason, he concludes, is not coherent even though both utilitarianism and rational egoism are coherent. Interestingly, Sidgwick's view that the lack of an over-arching principle established a contradiction is itself a reflection of the importance he places on coherence.(The Methods of Ethics, (Macmillan and Co.: London, 1907), seventh ed., p. 508.)

2. There is a path of justification connecting all the principles of the theory if and only if (i) each principle either justifies or is justified by some other principle of the theory, and (ii) there is no set of principles floating free from the rest.

3. As Amartya Sen and Bernard Williams comment, "the ambitiousness of utilitarianism has itself served as a source of its appeal. In promising to resolve all moral issues by relying on one uniform ultimate criterion, utilitarianism has appeared to be the 'rational' moral theory par excellence." (Utilitarianism and Beyond, "Introduction" (Cambridge University Press: Cambridge, 1982), p. 16.) Reflecting this appeal John Harsanyi defends utilitarianism saying that it "is the only ethical theory which consistently abides by the principle that moral issues must be decided by rational tests." See "Morality and the Theory of Rational Behavior," Utilitarianism and Beyond, p. 40; reprinted from Social Research, 44 (1977) pp. 623-656.

4. John Rawls admires utilitarianism's coherence and writes that "...one of the great attractions of the classical doctrine is the way it faces the priority problem." Having acknowledged this as a strength, Rawls goes on to solve the priority problem within his own theory by imposing

Taken together the three conditions for coherence impose specific and apparently severe constraints on the structure of a moral theory. The consistency requirement, for instance, demands of a theory that it exhibit both logical and practical consistency. It will be met if and only if the theory is structured in such a way that it will not, *even in principle*, generate conflicting directives.[5] As Mill argues, in order to secure consistency "there ought either to be some one fundamental principle or law at the root of all morality, or, if there be several, there should be a determinate order of precedence among them."[6] Usually this 'determinate order' is achieved in one of three ways: by ranking the principles of the theory, by giving them weights, or by introducing an explicit decision procedure. Regardless of how it is done, coherence requires that a theory have resources sufficient to resolve all conflicts between its principles .

Coherence also requires that theories be effectively monistic at the level of first principles; theories which have more than one first principle will fail to meet at least one of the conditions for coherence. Right off, pluralistic theories face at least the threat of inconsistency; in principle there may be situations (perhaps unforeseen) in which the multiple first principles will come into conflict. Coherence requires that this threat be overcome either by means of a non-justificatory priority rule, or by means of a justifying priority rule. If by the former, then the theory will violate the requirement of connectedness. If by the latter, then in effect the theory is made monistic, for the priority rule takes on the role of being the first principle by which the others are justified. A third possibility is that the plurality of principles are related in such a way that they are guaranteed never to conflict. For example, their explicit conditions of application might be mutually exclusive, one saying "When conditions C hold, do...", the other saying "When conditions C do not hold, do..." In these cases consistency would be guaranteed without the aid of an additional priority rule. Yet such theories would not be pluralistic in any interesting sense since they are logically and practically equivalent to monistic theories consisting of the conjunction of the same principles.[7] Recognizing the connection between

coherence and monism, both Kant and Mill are led by considerations of coherence to search for and adopt a single moral principle as the foundation of their theories. Both think that to be acceptable a moral theory must be coherent, so each advances a single first principle rather than a plurality.

Not all monistic theories will be coherent, however. After all, *any* pluralistic theory can be made monistic simply by conjoining its disparate principles into a single complex principle. Such "complex monistic" theories will automatically meet the connectedness requirement since they have only a single principle. Nonetheless, some will run afoul of coherence just by failing to be consistent. Gimmicky monistic theories aside, even simple monistic theories may fail to be coherent. For example a theory which demands only that one always keep one's promises (or always help those in need, or always uphold one's honor...) will generate conflicting directives under some circumstances.[8] Again, then, the consistency requirement will not be met. Thus, while monism at the level of first principles is a necessary condition for coherence, it is not a sufficient one.

The attraction of coherence comes, in part, from a belief that one fully understands only those things for which one has a coherent theory. Straightway, theories which fail to be coherent because they are logically inconsistent will be incomprehensible.[9] But even those (non-coherent) theories which satisfy the consistency requirement, by including a non-justificatory priority rule, do not satisfy our understanding; for such a priority rule does not tell why principles are ranked as they are, but only that they are so ranked. Complete understanding comes only with a fully coherent theory. If nothing else, then, a desire for understanding may account for the appeal of coherence.

Despite the popularity of coherence, its appeal has not quite been universal. Specifically, and infamously, Intuitionists have spurned coherence. As a result they have often been chastised for not providing an explicit principle to resolve conflicts between the plurality of first principles they allow. Indeed, the reverence for coherence has nowhere had a clearer impact than in arguments attacking intuitionism. The intuitionists' rejection of coherence is said to constitute a problem because rational discussion must end where reasons can no

---

coherence as a constraint on the concept of Right. See A Theory of Justice (Harvard University Press: Cambridge, MA, 1972) p. 41 and pp. 133-34.

[5]. Given standard deontic logic the requirement simply demands negation-consistency; but with suitable adjustments to deontic logic conflicting directives can be allowed without contradiction, though the consistency requirement will still be violated because of conflicting directives. For this reason the consistency requirement has been rejected by those who believe there are moral dilemmas. I discuss the relation between moral dilemmas, moral theory, and deontic logic in "Deontic Logic and the Priority of Moral Theory" in *Noûs*, (forthcoming, 1986).

[6]. John Stuart Mill, Utilitarianism (Hackett Publishing Co.: Indianapolis, 1979), p. 3. Ronald Dworkin makes the same point: "In a well-formed theory some consistent set of these [goals, rights, and duties], internally ranked or weighted, will be taken as fundamental or ultimate within the theory." Taking Rights Seriously (Harvard University Press: Cambridge, MA, 1977), p. 171.

[7]. This possibility was suggested to me by Tom Hill.

[8]. Of course, through the handy-work of God (for instance) the in-principle threat of conflict may never make an appearance in practice. God might have so arranged the world that we would never actually find ourselves in those situations where our principles would conflict. As David Braybrooke has pointed out to me, this possibility makes conceptual room for a sort of practical consistency, even where there is the possibility of in-principle conflict. Nonetheless, since the fortuitous elimination of the conflicts is external to the theory, reliance on the good will and good planning of God will leave the theory proper a failure in the eyes of the consistency requirement.

[9]. Comprehensibility of a sort can be brought to inconsistent theories, as The Logic of Inconsistency (Blackwell: Oxford, 1980) by Nicholas Rescher and Robert Brandom shows. But even this comprehensibility is achieved by segregating inconsistent claims. As they point out: "To say of reality *per se*... that it is inconsistent in some strong sense... would be to say of it something that, in the final analysis, is simply meaningless" (p. 25).

longer be offered -- and intuitionism seems to let the discussion end too early. As Rawls emphasizes:

> The assignment of weights is an essential and not a minor part of a conception of justice. If we cannot explain how these weights are to be determined by reasonable ethical criteria, the means of rational discussion have come to an end. An intuitionist conception of justice is, one might say, but half a conception.[10]

Only coherent moral theories allow justification to extend to the very foundation of our moral judgments, and this is clearly an advantage.

Reflecting the attractions of coherence, contemporary moral theory has given a prominent role to a particular model for moral theorizing that places a primacy on coherence. According to this model, the task of moral philosophy is to construct a fully articulated and coherent justificatory theory for our considered moral convictions. This model, which I will refer to as the "coherence model," has been vigorously defended by Ronald Dworkin[11] and it evidently underlies Rawls's use of the method of reflective equilibrium.

Recently, this coherence model has exercised a strikingly strong pull on the allegiance of moral philosophers. I will argue, however, (i) that the coherence model is unacceptable, (ii) that coherence should serve (at most) as a regulative ideal, not as an operative constraint, and (iii) that a theory's being coherent constitutes overwhelming evidence that it is wrong.

In what follows I concentrate on the coherence model and the general approach to moral theorizing -- the conviction ethics approach -- to which the model is wed. I limit myself to the conviction ethics approach in particular because it is the one most commonly taken by those captivated by coherence. Yet the problems which render the coherence model unacceptable within this approach arise for the model's analogues in other approaches as well.

2. Conviction Ethics and Models for Theorizing

Having moral convictions, and making moral judgments, are common features of experience. Many have taken the job of moral theory to be the justification of these convictions. Others, though, have argued that the intuitions are mere reflections of socialization and self-interest and of little help in developing an acceptable ethical theory. This raises a question: what connection, if any, should there be between our particular moral judgments, the

justifying rationales we find intuitively attractive, and moral theory? What role should moral intuitions play in moral theory?

The answer given by those who adopt the conviction ethics approach is that moral intuitions constitute the appropriate point of departure for moral theorizing; moral intuitions, they say, should serve as the initial imput for moral theorizing. Alternative starting points identify other prominent approaches to moral theorizing. Rational ethics, theological ethics, and empirical ethics, for instance, all advocate basing moral theory on something other than our moral intuitions. Rational ethics takes our nature as rational beings, or some conception of rationality, as its starting point; theological ethics finds the starting point for moral theory in religious doctrine; and empirical ethics holds that the proper point of departure is to be found in some non-normative fact about the world.

The rational, theological, and empirical approaches to ethics each relegate moral intuitions to a relatively minor role. Although these approaches do require some agreement between moral intuitions and moral theory, the requirement is often minimal and the appeals to such agreement serve as reassurance rather than as justification. Kant, for example, considers and then rejects intuitive moral judgments as support for his theory "because ease of use and apparent adequacy of a principle are not any sure proof of its correctness, but rather awaken a certain partiality which prevents a rigorous investigation and evaluation of it for itself without regard to consequences."[12]

Different approaches to theorizing may thus be distinguished by what they take as the appropriate starting point for moral theorizing. In spite of the different starting points, though, the competing approaches might converge on a single moral code. Different paths may have the same destination. As a consequence, one should not identify a theory's point of departure with its (substantive) first principle(s). The principles are the product of theorizing; they are distinct from the choice of how one is to go about theorizing, and of what starting point one will take.

Beginning with intuitions, the conviction ethics approach proceeds by constructing a theoretical structure to both justify and explain cherished convictions. Within this approach, moral intuitions and moral theory are seen as standing in a relation of mutual support.[13]

---

[10]. A Theory of Justice, op. cit., p. 41.

[11]. Dworkin refers to the model as the "constructivist" model. I hope that the change in terminology does not cause confusion; it does serve to highlight the characteristic of the model with which I am concerned. See Taking Rights Seriously, op. cit., pp. 160-168.

[12]. Foundations of the Metaphysics of Morals (Bobbs-Merrill: Indianapolis, 1959) L. W. Beck, trans., p. 8. Kant rejects an appeal to examples even more strongly later saying "Nor could one give poorer counsel to morality than to attempt to derive it from examples. For each example of morality which is exhibited to me must itself have been previously judged according to principles of morality to see whether it is worthy to serve as an original example..." (p. 25). In a similar vein Mill observes that "A test of right and wrong must be the means, one would think, of ascertaining what is right or wrong, and not a consequence of having already ascertained it." (Utilitarianism (Bobbs-Merrill: Indianapolis, 1957), p. 4).

[13]. This approach has a long history, going back at least to Aristotle's Nichomachean Ethics, Book 1,

While the approach begins by accepting moral intuitions, the search for principles will lead to critical examination of the principles we are initially tempted to embrace. No doubt we may arrive rather quickly at a preliminary set of principles which seem to justify a good number of our intuitions. It would be very surprising, however, if we didn't find on further investigation that some of the preliminary principles were inadequate or incompatible. We might discover that one of these principles conflicted with a conviction we hold dear. Or we might find that a principle used to justify a cluster of our intuitions was incompatible with another of the principles we found attractive. We would go back then, and re-examine our principles and our intuitions, in the hope of finding a reconciliation. Perhaps we would find that we had made a mistake either in proposing the principle or in thinking that the intuition was one of which we were confident. Or, we might find a principle which would eliminate the apparent conflict. In order to arrive at a moral theory we will move back and forth between our convictions and our principles. In the process we will often be led to abandon or amend one in light of others. As long as we feel comfortable with the amendments we make, as long as we don't have to compromise any of our convictions in the search for a justificatory theory, no problem arises. Yet it is unduly optimistic to think that we will not come across cases in which our intuitions and the principles supporting them are incompatible. Our moral intuitions will most probably recommend a set of inconsistent moral principles.

As a general approach to moral theorizing, conviction ethics does not specify what relation our moral principles should bear to one another. How, if at all, should the principles of a moral theory fit together? We have a choice, within the conviction ethics approach, between three different models which define the structure of the theory we are to construct.

The natural model, because it puts a premium on intuitions, places no *a priori* constraints on the structure of the theory our intuitions support. This model respects moral intuitions as insights into objective moral reality. Our intuitions are seen as providing the reference points from which theories of morality must take their bearings, just as observational data provide the reference points for scientific theories. If our intuitions are found to have rationales which bear no relation to one another, or rationales which are inconsistent, then the model would have us conclude that morality itself is fragmented or inconsistent.[14] Thomas Nagel characterizes this approach (his own) as one of trusting "problems over solutions, intuition over argument, and

pluralistic discord over systematic harmony."[15]

Unlike the natural model, the consistency model places a constraint on the structure of the theory we are to construct. Specifically, it requires that the set of principles we adopt be consistent. And it does so on the grounds that a theory which generates conflicting directives will be unacceptable, because certainly wrong. A conflict between principles, Sidgwick suggests,

> ...is absolute proof that at least one of the formulae need qualification: and suggests a doubt about whether the correctly qualified proposition will present itself with the same self-evidence as the simpler but inadequate one; and whether we have not mistaken for an ultimate and independent axiom one that is really derivative and subordinate.[16]

On this model, apparent inconsistencies force a choice: either abandon one of the conflicting principles, amend one or both of the principles so as to render them consistent, or find a higher-order principle which establishes which is to prevail (and under what circumstances) when a conflict arises. This is a model for a theory in transition; it treats our intuitions not as infallible insights into moral reality, but as the closest thing to insights we have. It counts on our theory to evolve and it imposes consistency as an interim requirement on the theories we construct.[17] Underlying the requirement is the view that the moral reality our theory is about is itself consistent (and perhaps even coherent).

The coherence model, which is embraced by Dworkin and Rawls, places an even stricter constraint on the sort of theory our intuitions can countenance. The coherence model instructs us to take our considered judgments, order them according to conviction, and then construct a *coherent* (not merely a consistent) justificatory theory for our strongly held convictions. Those convictions which we cannot fit within a coherent theory must (for the time being) be set aside.

The coherence model finds its support, Dworkin argues, in a doctrine of responsibility that condemns the adoption of a fragmented set of principles for action and that "condemns the practice of making decisions that seem right in isolation, but cannot be brought within some comprehensive theory of general principles and polices..." According to Dworkin "...men and women have a

---

chapters 3-4. See Joel Feinberg's discussion of what I call conviction ethics in "Justice, Fairness, and Rationality," Yale Law Journal 81, (1972) pp. 1004-1031.

[14]. Dworkin introduces this model when discussing Rawls' method of reflective equilibrium. See Taking Rights Seriously, op. cit., p. 160.

[15]. Mortal Questions, (Cambridge University Press: Cambridge, 1979) p. X. G.E. Moore endorses a similar view. He writes that "To search for 'unity' and 'system' at the expense of truth is not, I take it, the proper business of philosophy, however universally it may have been the practice of philosophers. And that all truths about the universe possess to one another all the various relations which may be meant by 'unity' can only be legitimately asserted when we have carefully distinguished those various relations and discovered what those truths are." (Principia Ethica (Cambridge University Press: Cambridge, 1971), p. 222.)

[16]. The Methods of Ethics, op. cit., p. 341, (BK. III, Ch. XI, sec. III).

[17]. Norman Daniels suggests the need for some such model in "On Some Methods of Ethics and Linguistics," in Philosophical Studies, 37 (1980), pp. 21-36.

responsibility to fit the particular judgments on which they act into a coherent program of action..." And they must "integrate their intuitions and subordinate some of these, when necessary, to that responsibility."[18]

Along the same lines, Richard Wasserstrom claims that rationality requires that "...the process of justification continue until the 'ultimate' premise upon which any decision stands and from which it draws its claim for acceptability is fully revealed." This requirement, Wasserstrom maintains, asks only "...that men be *rational* -- in the best sense of the word -- so that the conclusions they have reached may be as accurate as possible, and the conduct undertaken as beneficial as possible."[19]

Recommending the doctrine of responsibility is the concern, expressed by Peter Geach, that asking anything less will lead "to quite laxist consequences."[20] Prejudice, malice, and self-interest will be free (it seems) to masquerade as principled action unless we demand a coherent justification that applies principles consistently and explains their priority.

The doctrine of responsibility is not easily satisfied: it does not merely say that coherence is a desirable attribute in a theory, nor does it simply suggest that the perfect theory would be coherent; it holds that the only acceptable justification for action is a fully articulated coherent justification.[21] The doctrine requires that we act always and only in accordance with the best fully coherent theory we have on hand. It requires that we treat our working theories as complete, and it requires that we take the dictates of our theories as conclusive until we can develop a more attractive coherent theory. Stray principles must be brought within a coherent theory before they may legitimately be acted upon, regardless of how independently attractive they are. Thus the doctrine compels us to artificially suppress crises in the theory. We must treat the best coherent theory we have as the final arbitrator of moral disputes while we await the development of a more attractive theory.

Both the coherence and the consistency models demand that our intuitions be supported by principles. They agree that our actions ought to be backed by justifying rationales. Where they differ is in the constraints they place on the principles used in the justificatory theories. The coherence model demands the most; it requires that the principles belong to the best coherent theory the agent has on hand. Given that coherent moral theories are available, the coherence model would have us stick with the best of these -- even when its use led to unintuitive results -- until we developed a better coherent theory.[22] The consistency model, on the other hand, allows that we may justifiably act even when not capable of offering a coherent justification, as long as our justification is consistent and plausible.

A coherent theory leaves everyone with a clear decision procedure for resolving all moral issues. A (merely) consistent theory provides far less guidance. Most likely, such a theory will contain a plurality of important principles, and various arguments for assigning the principles more or less specific relative weights. Still, it will leave us with an open-ended theory and may not provide resources sufficient to settle all moral questions. Those armed with a merely consistent theory will have arguments to support the rankings of the principles in particular cases. Sometimes the arguments will be clear and decisive. At other times, though, they will be less clear and vigorously contested. In some cases a consistent theory may be unable to provide an explicit principle to justify the weights assigned. A theory may be consistent and still be incomplete. Of course, if the principles invoked really are consistent then there will be some principle or other which could (theoretically) resolve apparent conflicts and assign weights. The consistency model, however, acknowledges that at any given time some of these higher order principles may be unknown. Unlike the coherence model, this model refrains from demanding that we artificially impose and act upon an explicit decision procedure. As a result, the consistency model allows that decisions made in the name of morality may outstrip a person's ability to account for the decisions within a complete coherent theory. Even so, the model demands that these decisions be backed by a consistent and plausible set of principles. Thus, it too demands that people act on principle rather than merely react.

Within the conviction ethics approach a choice between three significantly different models for theorizing must be made. All three are models for reasoning from particular moral convictions to a general theory of morality. This common ground is what brings them all under the conviction ethics rubric. Where the three models differ is on whether (and which) constraints should be placed on the structure of the theory we construct. The natural model places a priority on our intuitions and lets their underlying principles fall where they

[18]. Taking Rights Seriously, op. cit., pp. 87, 160 & 162. That the "coherent program of action" demanded by the doctrine of responsibility must be coherent in the sense I specify is not be immediately clear. I believe, however, that no characterization of coherence weaker than that one will justify the structural constraints explicitly imposed by many, and praised by others, on the construction of acceptable moral theories.

[19]. The Judicial Decision (Stanford University Press: Stanford, 1965), p. 95. Though Wasserstrom offers this argument in the context of a discussion of legal decisions, he seems to be endorsing coherence as a general requirement of rational justification.

[20]. "Good and Evil" in Philippa Foot (ed.), Theories of Ethics (Oxford University Press: Oxford, 1967), p. 72; reprinted from Analysis 17 (1956), pp. 33-42.

[21]. When applied to public officials, the doctrine "...demands that decisions taken in the name of justice must never outstrip an official's ability to account for these decisions in a theory of justice, even when a theory must compromise some of his intuitions. It demands that we act on principle rather than on faith." (Taking Rights Seriously, op. cit., p. 162, my emphasis.)

[22]. Just what the model demands when the best coherent theory is terrible is unclear. Presumably, since the model is used within the conviction ethics approach, the doctrine of responsibility will lead its advocates to say that we are not in the position to act justifiably.

may. The consistency and coherence models, in contrast, both place constraints on how the principles of a theory are to be related. Clearly, the character of the moral theory one develops will depend greatly on which of these models is adopted.

3. Two Levels of Theorizing

Coherence may play either of two very different roles within the conviction ethics approach (and any approach) to moral theory. Coherence might simply serve as a regulative ideal. Or, it might serve as an operative constraint on actual theorizing. The significant difference between these two roles can best be made clear by developing briefly a general account of morality which falls within the conviction ethics approach and which I call *conventional objectivism*.[23] According to conventional objectivism, the correctness of a moral principle depends on its falling within the best coherent justificatory theory available (in principle) for the practices and principles we embrace.[24] So construed, conventional objectivism rests midway between independent objectivism and subjectivism.

The one extreme, independent objectivism, makes three characteristic claims: (i) there is an objective standard of morality which is prior to all conventions, and which may be used to evaluate these conventions, (ii) conventions and practices which conflict with this objective morality are immoral, and (iii) when objective morality and conventional practice conflict, it is a person's duty to break with convention. At the other extreme is subjectivism, which has as its central tenet the claim that there is no objective standard, whatsoever, for morality. Thus subjectivism rejects all three tenets of independent objectivism. In their place subjectivism acknowledges only opinion and preference as the source and measure of rightness and of obligation. Conventional objectivism holds, with independent objectivism, that there is an objective standard against which any particular convention or action must be measured. In this way it makes room for there being immoral conventions and morally unacceptable practices. Unlike independent objectivism, however, the objective standard is seen as itself being relative to the moral community in question, and it is taken to reflect the institutions and practices of that moral community. As a result, conventional objectivism allows that the objective standards in different communities may differ (sometimes dramatically).

Independent objectivists hold that the existence and substance of the standards of morality are independent of human conventions. Objective

morality, they say, completely transcends society. Conventional objectivism, in contrast, subordinates morality's existence and substance to society's practices on two levels.

First, conventional objectivism applies only to a certain kind of society. Roughly speaking, only those communities that have the appropriate formal-regulative structure, peopled by agents who are capable of principled action and who are in circumstances that call for and allow cooperation, are committed to treating morality as if it were defined by a coherent justificatory theory. There may be human communities or martian communities (as there are animal communities) which do not carry this commitment. Communities need not have the formal institutional structure, nor the method of reasoning, nor the social conditions, that generate this commitment. The first level of subordination of morality to social context, then, comes with the formal-regulative structure, material circumstances, and mental resources, of society; conventional objectivism applies only to a certain kind of community, a community capable of cooperation and social constraint -- a moral community. Second, even within those societies which have the appropriate structure, what morality demands will depend on their practices and conventions. Which theory will best justify the practices and conventions of a community obviously turns on what the practices and conventions are. Different communities will likely be committed to different justificatory theories and so to a different standard of morality.

According to conventional objectivism only communities which possess the appropriate formal-regulative structure and resources should be treated as if they have an objective standard against which to measure the particular practices and conventions found within the community. And, even among these, the nature of the objective standard found within any given community will reflect the substantive rules and history of the community. This double subordination to convention distinguishes conventional objectivism from independent objectivism.

Conventional objectivism, then, splits the difference between independent objectivism and subjectivism. It allows, as does independent objectivism, that there is an objective standard of morality. This standard serves as a measure against which we may evaluate particular practices and conventions. Yet it also allows, as does subjectivism, that both the nature of morality, and morality's specific demands, are products of human convention, and reflections of human interests.

In order to embellish the account of conventional objectivism, imagine that the coherence model for theorizing is used by an Ideal Theorist who is charged with developing the best coherent justification he can for the practices, principles, and institutions we embrace as right.[25] The Ideal Theorist, in different

---

[23]. The account of morality I sketch in what follows draws heavily on Dworkin's account of the nature of law.

[24]. Of course, if conventional objectivism is to avoid pulling itself out of a hat there must be some standard for evaluating justificatory theories that is not itself a moral standard.

[25]. When developing the justificatory theory our Ideal Theorist will probably discover that he cannot construct a theory which will justify all of our convictions and practices. He will have to treat some of them as mistakes. Allowing that some of our practices and convictions may be treated

developed communities, would presumably arrive at different theories (using the coherence model) which would reflect the differences in their convictions, practices, and interests. As the Ideal Theorist travels from one community to another, he will construct justificatory theories to accommodate the differing practices. For significantly different practices, his use of the coherence model may result in wildly different moral theories.

Given the conventional objectivist account of the nature of morality, the Ideal Theorist, and his use of the coherence model, serve as expository devices. They help to explicate a particular conception of the nature of morality. Though we can never arrive at the Ideal Theorist's theory with our more limited talents, we can confidently reject any theory which we know to be incompatible with his. The coherent theory he constructs serves as the regulative ideal for our own moral theorizing.

So far, the coherence model has only been given a role in the characterization of what an Ideal Theorist should do. In this role the model serves to establish a regulative ideal for our theorizing. However, some people (Dworkin and Rawls, for instance) seem to hold that the coherence model should also be used by *real* people (or at least real moral theorists) in constructing moral theories. And they apparently think that the doctrine of responsibility requires that decisions made by real people must be backed by a coherent theory which justifies them.[26]

With this in mind, we should mark off two distinct levels of theorizing. The higher level is occupied by the Ideal Theorist, a being of super-human faculties. He will employ the coherence model in constructing a justificatory theory for the general conventions and practices of the appropriate community. The theory the Ideal Theorist constructs will, in turn, define the objective standard of morality for that community (according to conventional objectivism). The second level of theorizing is occupied by real people (who, of course, lack the Ideal Theorist's super-human talents). Mimicking the Ideal Theorist, we too might use the coherence model to construct the best coherent justificatory theory we can for the conventions and practices we embrace.[27]

The doctrine of responsibility that supports the coherence model holds that an action is justified only if its rationale can be brought within a coherent justificatory theory. It is a doctrine that applies to real people. Yet it lends itself to two significantly different interpretations, depending on whether coherence is treated as a regulative ideal or as an operative constraint.

The stronger interpretation, the one that supports our use of the coherence model (and the one Dworkin and Rawls seem to advocate), maintains that real people have a responsibility to act only on those judgments which they are actually able to fit within a coherent justificatory theory they have on hand. This interpretation holds that we should have a coherent theory in our possession before we act. It requires that real people construct a theory using the coherence model, and that they act only on those principles which are supported by the theory constructed. Those actions which cannot be justified by a currently available coherent theory must not be performed. So interpreted, the doctrine places the burden of coherence squarely on the shoulders of real people; it imposes coherence as an operative constraint on the construction of moral theories.

The weaker interpretation of the doctrine holds that real people have a responsibility to make only those decisions which could in principle be brought within a coherent justificatory theory. It recognizes coherence as an ideal without imposing it as a constraint. Hence it does not require that people have a fully articulated theory on hand. On this interpretation, the doctrine permits decisions in the absence of a coherent moral theory, as long as the principles used to justify the decisions may (theoretically) be fit within the coherent ideal. Since no set of inconsistent principles could ever be brought with a single coherent theory, the justificatory principles must be consistent. For this reason, apparent inconsistencies are aberrations which must be resolved.

Thus under the weaker interpretation, the doctrine requires consistency, though not coherence, of our working theories, on the grounds that the correct theory (the Ideal Theorist's theory) is coherent. The strong requirement of *articulated* coherence is abandoned, and with it the demand that people work out and adopt coherent theories. The doctrine of responsibility, on the weaker

---

as mistaken raises a problem, though. For there is likely to be more than one theory which can justify a large number of our practices and convictions, especially when some may be ignored. A choice will have to be made between competing coherent theories, and this choice will have to turn on their moral as well as their explanatory appeal. As Dworkin notes, "...justification must be plausible and not sham. If the justification...makes distinctions that are arbitrary and deploys principles that are unappealing, then it cannot count as a justification at all."(Taking Rights Seriously, op. cit., p. 118.)

[26]. Dworkin introduces the doctrine, originally, as a constraint only on political officials, but the conception of justification he invokes seems to require application of the doctrine to all who act in ways which might affect others. (See Taking Rights Seriously, op. cit., pp. 87-88.)

[27]. Failure to distinguish the two levels can lead one, as it appears to have led Rawls, into thinking

that because ideal theorists view themselves as constructing a theory in a moral vacuum, we can too. While it is true (given Kantian Constructivism) that "...the parties in the original position do not recognize any principles of justice as true or correct and so as antecedently given," it seems a mistake to say (as Rawls does) that "the idea of approximating the moral truth has no place in a constructivist doctrine." The claim is plausible only so long as we ignore our own role as moral theorists, rather than as parties in the original position. What those in the original position would agree to sets the standard (in the way the ideal theorist's theory does, according to conventional objectivism) against which our proposals are to be evaluated. Our claims of justice are correct (according to Kantian Constructivism) just to the extent we are right about what would be agreed to. So the idea of approximating moral truth will have a central place in constructivism, for approximating the truth is what we aim to do in figuring out what those in the original position would choose. (See John Rawls' "Kantian Constructivism in Moral Theory," The Journal of Philosophy, 77, 9 (1980), pp. 515-572, p.564. See also A Theory of Justice, op. cit., p. 45.)

interpretation, both supports and requires the consistency but not the coherence model for moral theorizing.

The two interpretations of the doctrine clearly require different things of a person. The stronger interpretation requires that a person *have a coherent theory on hand* which justifies practices and convictions. The weaker interpretation requires only that the person have a consistent, but not necessarily a complete or coherent, theory on hand which justifies these practices and convictions. The choice of which interpretation we should adopt, if either, is obviously one of significance, for on it will turn which model for theorizing (and justification) we will take as appropriate for less than Ideal Theorists like ourselves.

Ought we to place a primium on coherence, and thus require of ourselves and others the use of the coherence model for justifying and guiding our actions? Or should we require the consistency model and so demand of people that they justify their actions with a consistent (though not necessarily a coherent) moral theory? Or should we abandon constraints on the structure of the theory altogether and ask only that people act as their convictions would have them? I shall argue that, if we assume that the ideal moral theory is coherent, and if we grant that our theories will fall short of the ideal, then we ought to reject the coherence model and accept instead the consistency model for moral theorizing.

4. Against Coherence

Almost all moral theorists have assumed that the ideal moral theory is coherent. They have, in other words, accepted coherence as a regulative ideal. This is reflected in the fact that few seriously entertained moral theories have allowed the possibility of truly unresolvable moral dilemmas.[28] Almost as many ethicists have also assumed (mistakenly) that our working theories should likewise be fully coherent.

The argument against imposing coherence as an operative constraint on our theory building contains two premises. The first, the *premise of imperfection*, simply acknowledges that any moral theory we develop will fall short of the fully correct theory. The second, the *incompatibility premise*, asserts that any two (non-identical) coherent moral theories covering the same domain are incompatible. If we accept coherence as a regulative ideal -- if we assume that the fully correct theory of morality is coherent -- then the two premises together entail that any coherent moral theory we develop in practice will be

incompatible with the correct theory.[29] As a result, the real issue centers on the acceptability of the premises.

The plausibility of the premise of imperfection depends, clearly, on what the theory is about. The more ambitious the theory, the more plausible the premise, for the less likely we are to get everything right; the hold of the premise bears an inverse relationship to our confidence in a theory. In the case of moral theorizing the premise appears amply justified; by almost all accounts, the chances for developing a perfect (fully correct) theory are exceedingly slim. Of course, if we do become confident that we've got everything right, then the premise of imperfection would have to be abandoned. Most moral theorists, though, are quick to acknowledge that their theories are less than perfect. This legitimate bit of modesty is all that the premise embodies.

The incompatibility premise sets two conditions on its application: the moral theories in question must be different, and they must both be coherent. The premise holds only for different moral theories because, obviously, theories which are identical will not be incompatible. In determining whether this first condition is met, the following criteria for identity will serve our purposes: Two moral theories are identical if and only if they justify and require all and only the same actions for the same reasons. Identical theories, then, will agree both on which actions should be performed *and* on why they should be performed. So we may think of moral theories as functions which take situations into sets of pairs, each pair consisting of an action and its rationale.[30] This means that moral theories which are not identical will either (i) justify or require different actions, or (ii) offer different justifications for the actions they agree should be performed. Bentham's utilitarianism and Rawls' constructivism in fact require people to perform different actions. This is enough to guarantee that they are different theories. However, even if (by some fluke) they were to justify the same actions -- if doing what is right (according to Rawls) always maximized pleasure (as measured by Bentham's hedonic calculus) -- the theories would still be different, since they would require the actions they do for different, and opposed, reasons.[31]

---

[28]. Even intuitionists, who reject coherence as an operative constraint, studiously avoid allowing the possibility of true moral dilemmas.

[29]. That the argument is valid can be seen from its first order formalization:

(x)(y)[(Tx & Wy & Cy)--->x=y]
  (x)(y)[(Cx & Cy & x=y)--->Ixy]

  (x)(y)[(Tx & Cx & Wy & Cy)--->Ixy]

where Tx is "x is the true, or fully correct moral theory," Wy is "y is a moral theory we develop," Cy is "y is a coherent moral theory," and Ixy is "x and y are incompatible."

[30]. A moral theory will provide information like: "in situation x do y for reason r." Of course the theory will have to distinguish between mandatory, permitted, and forbidden acts. These latter two might be captured by substitutions of "may do" and "do not do" for "do" in the schema.

[31]. The opposition would presumably show up in a difference in the actions recommended and forbidden in some possible worlds (even if not in this world). Otherwise we'd have excellent

These criteria for identity reflect the fact that, in ethics, one's reasons for acting, as well as one's actions, are important. An acceptable moral theory must do more than enjoin us to action; it must supply good reasons for acting as it enjoins. As Kant emphasized, people may be counted among the moral only if they perform the right actions for the right reasons. A person who succumbs to temptation is not redeemed by the fact that morality happened to require the action performed: just as a judge who takes a bribe in exchange for finding someone guilty is not redeemed if the person happens to be guilty. Redemption for the selfish and the corrupt is found in reformed motivations and not merely acceptable actions.

Of course not all non-identical theories are incompatible. Different theories may deal with unrelated topics, or they may overlap and complement, rather than conflict with, one another (as a complete theory of morality would overlap and compliment a theory of justice). We may be sure, however, that two different coherent justificatory theories covering *all* and *only* the same territory will be incompatible. We know, from the fact that they are coherent, that they are complete and so offer answers to all and only the relevant issues. We know also, from the fact that they are different, that the actions they require, or the rationales they offer (when fully spelled-out), will be different. Since they will be giving different and conflicting answers to the same questions, they will be incompatible. So, as the incompatibility premise claims, any two different coherent theories which deal with the same subject matter (in this case the demands of morality) will be incompatible: the two justificatory theories, conjoined, will generate conflicting directives (at least in some conceivable circumstances). A similar argument holds for inconsistent theories, though the problems are even more obvious. Since one can derive conflicting directives from the inconsistent theory alone, it follows trivially that such a theory, combined with any other, will generate conflicting directives. An inconsistent theory is incompatible with all other theories. Hence, any inconsistent moral theory will be incompatible with the ideal theory.

Importantly, the incompatibility premise does not hold for merely consistent theories. A consistent theory, as long as it is merely consistent (and not coherent) will be compatible with many different theories covering the same actions -- even if these other theories are coherent. Hence if we accept coherence as a regulative ideal in moral theorizing then only the consistency model allows us to construct theories which may fall short of, and yet still be compatible with, the ideal. By demanding less than coherence we decrease our chances of going wrong.

Perhaps, though, the risk of going wrong is worth taking. In light of this possibility, we should go back to see what arguments can be offered in defense

___

grounds for thinking the apparently different theories are just terminological variations of one.

17

of the strong interpretation of the doctrine of responsibility and its imposition of coherence as an operative constraint. Dworkin has advanced two such arguments.[32] The first is moral, the second prudential; neither is effective.

The moral argument rests on the claim that fairness requires imposing coherence as a constraint on acceptable justification. As grounds for using the coherence model in our moral (as opposed, say, to legal) theorizing, however, an appeal to fairness is unacceptable; for such an argument either is circular or presupposes a more fundamental model for theorizing. If our conception of fairness is arrived at by using the coherence model (for instance, if we were to accept Rawls' theory) then fairness cannot be invoked as a defense of the coherence model without circularity. Still, if we could develop our conception of fairness without appeal to the coherence model then the circularity would be avoided. One might be tempted to argue, with this in mind, that our intuitive grasp of the concept of fairness is substantial enough to justify the use of the coherence model; intuition without moral theory would then justify the coherence model. The problem with this argument is that our intuitive view of fairness simply does not support the coherence model. We do not think that people are behaving unfairly if they act without possessing a fully articulated coherent theory. While we do often (and reasonably) demand justification for an action, and we do demand that the justification offered be consistent, we seldom if ever demand that a person have on hand a fully worked out justificatory theory. Providing such a theory goes far beyond anything a pre-reflective notion of fairness might require. To get off the ground, then, the fairness argument for the coherence model requires that some moral theorizing be done. However, if we arrive at our conception of fairness, not by intuition, and not by using the coherence model, but by using some other model for theorizing, then the argument from fairness will presuppose that other model. Either way, no moral argument for the coherence model emerges.[33]

The prudential argument rests on the claim that requiring a coherent justification helps to minimize the abuse of power by not allowing "appeals to unique intuitions that might mask prejudice or self-interest in particular cases."[34] Both the consistency and the coherence models go some way towards this: both demand that judgments be integrated into a theoretical structure. So both force people to subject their opinions and their intuitions to rational scrutiny. Yet neither model can guarantee protection from abuse. They each allow, as they

___

[32]. Taking Rights Seriously, op. cit., pp. 162-63.

[33]. Dworkin invokes the fairness argument to defend his use of the coherence model's analogue in legal theorizing; here he is able to avoid circularity. His argument will work even in this context, of course, only if fairness does require coherence of our legal justifications (which I doubt).

[34]. Taking Rights Seriously, op. cit., pp. 162-63. Dworkin also seems to think that, taking into account "the relative capacities of men and women who might occupy different roles," requiring the coherence model "may be expected to reduce the number of mistakes overall" made by such people (p. 130).

18

must, that past judgments may be abandoned in light of new considerations. They thus allow the unscrupulous to introduce disingenuous justifications (masquerading as honest changes of mind) which are developed to disguise the abuse of power. The justifications can always be surreptitiously manipulated.[35] In spite of this, the coherence and the consistency models share an advantage over the natural model; they at least require subordinating reactions to reason by forcing agents to offer justifications. To this extent prudence may recommend the adoption of one of these two models. Even so, all three models suffer the (inevitable) short-coming of not preventing the devious from abusing power.

While both the coherence and the consistency models may be preferable to the natural model (on prudential grounds), the consistency model alone meets the demands of prudence without unsavory consequences. For the coherence model, unlike the consistency model, protects against the abuse of power at the expense of insuring the falsity of the justifications developed. It requires embracing a coherent theory which we have good reason to believe will be incompatible with the correct theory; and it demands that we stick to that coherent theory until we develop a more attractive coherent theory, which, in turn, we will still have good reason to think is incompatible with the correct theory. A parallel problem, by the way, plagues the coherence model's analogue in science. Given evidence for both the wave and the particle theories of light, for instance, the coherence model's analogue in science would have us embrace one or the other (for a coherent theory must be in hand) and it would demand that we stick to that theory until some better coherent theory could be developed. It would demand this even in the face of the manifest inadequacy of either theory to account for all the phenomena. In contrast, science's analogue of the consistency model would require only that the claims of the two theories be sufficiently restricted so as to render them compatible (even if not unified).

Untouched by my arguments, of course, are heuristic considerations in favor of the coherence model. It is perfectly reasonable to pursue coherence hoping that the pursuit will lead to the construction of better theories -- it is perfectly reasonable to treat coherence as a regulative ideal. What is unreasonable, I've argued, is (i) thinking our theories are better *because* they are coherent or (ii) thinking that the only acceptable theories are those that are coherent.

The argument against the coherence model assumes there is some correct (ideal) theory with which our theories are in danger of being incompatible. Obviously, there is no need to claim *access* to the correct theory. Still, the argument may seem to presuppose an unacceptably strong objectivist conception

of morality. Importantly, though, this appearance is misleading, as the argument presupposes only a weak form of moral realism. The argument does assume that moral judgments may be (literally) true or false, and it assumes that the true judgments may, at least in principle, be brought under and justified by a coherent moral theory. Indeed, as conventional objectivism would have it, their fitting within such a theory is what makes them true. Yet the argument need not presuppose anything in particular concerning the conditions for the truth of moral judgments (except that such conditions, whatever they may be, sometimes hold). The argument does not assume, for example, that the conditions for the truth of moral judgments are either independent or universal. They may be, as conventional objectivism holds, relative to the institutions and practices of society, and thus be neither independent nor universal. So while the argument against coherence does assume that there is some sort of objective standard for moral judgments, the argument is compatible with this standard being a product of convention.[36]

The argument against the coherence model also assumes that moral theory ought to aim, to the extent possible, at developing the right principles for action. Many, though, appear to think that arriving at such principles is either unimportant or impossible. Moral theory, they argue, has an essentially social function; it should develop theories which are intuitively attractive, simple, and easy to use.[37] A coherent theory, they will say, is best able to meet these demands. Such a theory will be easy to grasp since it will only have a single first principle, and it will be easy to apply since it will include an explicit decision procedure for resolving moral issues. A merely consistent theory, on the other hand, can not adequately serve the social function: it will be resistant to mass consumption. A consistent theory, to be plausible, will include a multiplicity of principles, but no explicit decision rule. So people will be expected to develop sensitivities relevant to moral reasoning and to develop and be prepared to defend their own moral stance. A consistent theory will thus be tedious to learn, difficult to apply, and not readily accessible to public

[35]. As R.M. Hare points out: "In practice, especially when in haste or under stress, we may easily, being human, 'cook' our moral thinking to suit our own interest." Moral Thinking (Oxford University Press: Oxford, 1981), p. 38.

[36]. This weak form of moral realism is, I believe, accepted by all of the theorists discussed in this paper. One might evade the whole argument, of course, by denying the weak moral realist assumption. Clearly, if there is no objective standard at all then there is nothing for our theory to be incompatible with. One might then reintroduce coherence as a constraint on theorizing on grounds of either self-protection (from arbitrariness) or conceptual aesthetics. A major problem with this sort of move is that it reduces moral theory to mere rationalization (in a pejorative sense). It robs the theory of its object, and so of most of its interest. (Needless to say, though, this is not a knock-down argument against the anti-realist.) I try to characterize what is central to being a moral realist in "The Many Moral Realisms."

[37]. Considerations of this sort stand behind the publicity requirement Rawls introduces in A Theory of Justice, op. cit., p. 135. See also "Kantian Constructivism in Moral Theory," op. cit, p. 561 and "Justice as Fairness: Political not Metaphysical," Philosophy and Public Affairs (Winter, 1985), pp. 223-251. In the last, Rawls distinguishes political theory from moral theory and then advocates interpreting 'justice as fairness' along with the publicity requirement as part of political and not moral theory. See for example p. 225.

examination.[38]

As this defense of coherence emphasizes, we may have good reason for developing and promulgating a simple coherent theory. Such a theory might well provide the best available basis for political cooperation in the face of widespread moral disagreement. Indeed, it may be that our shared allegiance to democratic institutions call for some publically available and explicitly stated conception of social justice. Nonetheless, this is no reason for thinking that moral theory (as opposed to political theory) is essentially tied up with developing moral conceptions that are publically workable. And unless moral theory is so tied, coherence cannot reasonably be defended as a general constraint on moral theorizing just by pointing out that coherent theories are best suited to political needs. Moreover, since one point of developing a good moral theory is to help decide whether a political theory's compromises are morally acceptable, the social role of moral theorizing can't be its sole role. The moral theorist's concern (unlike the political theorist's) must also be with getting the moral facts right and not simply with getting a publically workable theory.

There remains one attractive argument in support of the coherence model. It rests on stressing the difference between having good reasons and having the right reasons. In ethics, I have maintained, one must not only act rightly, one must act for good reasons. However, it would be a mistake to equate -- as I seem to -- having good moral reasons with having true moral beliefs. Just as one may justifiably believe what turns out to be false, so too may one justifiably do what turns out to be wrong. (We needn't be perfect in order to be good.) A person will have good reasons for acting, it might be said, as long as (i) the person's reasons are appropriately moral (perhaps the reasons should be universalizable and not purely self-interested) and (ii) the person has good grounds for his or her moral beliefs. Then if a person's moral belief will have good grounds as long as it is supported by a coherent justification, the coherence model might find support in a coherence theory of justification. The one thing the model guarantees is that we will have a coherent justification for our moral beliefs.

Certainly we can have good reasons for doing what turns out to be wrong, but having a coherent justification does not insure that we have good reasons for doing what we do. Even a fully coherent justification will be acceptable only if we have no independent grounds for thinking it wrong. In ethics, though, we do have such grounds; for the only fully coherent justifications available to us are patently inadequate to the complexities of moral life. As a result, coherent justifications almost invariably constitute bad justifications.[39] Coherence, then, ought not be revered: rather it should be treated with caution, even suspicion; for the coherence of our theories belies their truth.[40]

---

[38]. Such an argument is implicit in Rawls' justification for limiting options in the original position to simple coherent theories.

[39]. I say "almost invariably" rather than "invariably" because the grip of the premise of imperfection can be escaped if the ambitions of the justification are severly limited.