



## Philosophical Review

---

Moral Realism

Author(s): Peter Railton

Source: *The Philosophical Review*, Vol. 95, No. 2 (Apr., 1986), pp. 163-207

Published by: Duke University Press on behalf of Philosophical Review

Stable URL: <http://www.jstor.org/stable/2185589>

Accessed: 15/01/2009 15:35

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=duke>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



Duke University Press and *Philosophical Review* are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Review*.

<http://www.jstor.org>

## MORAL REALISM

Peter Railton

Among contemporary philosophers, even those who have not found skepticism about empirical science at all compelling have tended to find skepticism about morality irresistible. For various reasons, among them an understandable suspicion of moral absolutism, it has been thought a mark of good sense to explain away any appearance of objectivity in moral discourse. So common has it become in secular intellectual culture to treat morality as subjective or conventional that most of us now have difficulty imagining what it might be like for there to be facts to which moral judgments answer.

Undaunted, some philosophers have attempted to establish the objectivity of morality by arguing that reason, or science, affords a foundation for ethics. The history of such attempts hardly inspires confidence. Although rationalism in ethics has retained adherents long after other rationalisms have been abandoned, the powerful philosophical currents that have worn away at the idea that unaided reason might afford a standpoint from which to derive substantive conclusions show no signs of slackening. And ethical naturalism has yet to find a plausible synthesis of the empirical and the normative: the more it has given itself over to descriptive accounts of the origin of norms, the less has it retained recognizably moral force; the more it has undertaken to provide a recognizable basis for moral criticism or reconstruction, the less has it retained a firm connection with descriptive social or psychological theory.<sup>1</sup>

In what follows, I will present in a programmatic way a form of ethical naturalism that owes much to earlier theorists, but that seeks to effect a more satisfactory linkage of the normative to the empirical. The link cannot, I believe, be effected by proof. It is no more my aim to refute moral skepticism than it is the aim of contemporary epistemic naturalists to refute Cartesian skepticism. The

---

<sup>1</sup>Nineteenth-century evolutionary naturalism affords an example of the former, Dewey—and, on at least one reading, perhaps Mill as well—an example of the latter.

naturalist in either case has more modest aspirations. First, he seeks to provide an analysis of epistemology or ethics that permits us to see how the central evaluative functions of this domain could be carried out within existing (or prospective) empirical theories. Second, he attempts to show how traditional nonnaturalist accounts rely upon assumptions that are in some way incoherent, or that fit ill with existing science. And third, he presents to the skeptic a certain challenge, namely, to show how a skeptical account of our epistemic or moral practices could be as plausible, useful, or interesting as the account the naturalist offers, and how a skeptical reconstruction of such practices—should the skeptic, as often he does, attempt one—could succeed in preserving their distinctive place and function in human affairs. I will primarily be occupied with the first of these three aspirations.

One thing should be said at the outset. Some may be drawn to, or repelled by, moral realism out of a sense that it is the view of ethics that best expresses high moral earnestness. Yet one can be serious about morality, even to a fault, without being a moral realist. Indeed, a possible objection to the sort of moral realism I will defend here is that it may not make morality serious enough.

### I. SPECIES OF MORAL REALISM

Such diverse views have claimed to be—or have been accused of being—realist about morality, that an initial characterization of the position I will defend is needed before proceeding further. Claims—and accusations—of moral realism typically extend along some or all of the following dimensions. Roughly put: (1) Cognitivism—Are moral judgments capable of truth and falsity? (2) Theories of truth—If moral judgments do have truth values, in what sense? (3) Objectivity—In what ways, if any, does the existence of moral properties depend upon the actual or possible states of mind of intelligent beings? (4) Reductionism—Are moral properties reducible to, or do they in some weaker sense supervene upon, nonmoral properties? (5) Naturalism—Are moral properties natural properties? (6) Empiricism—Do we come to know moral facts in the same way we come to know the facts of empirical science, or are

they revealed by reason or by some special mode of apprehension? (7) Bivalence—Does the principle of the excluded middle apply to moral judgments? (8) Determinateness—Given whatever procedures we have for assessing moral judgments, how much of morality is likely to be determinable? (9) Categoricity—Do all rational agents necessarily have some reason to obey moral imperatives? (10) Universality—Are moral imperatives applicable to all rational agents, even (should such exist) those who lack a reason to comply with them? (11) Assessment of existing moralities—Are present moral beliefs approximately true, or do prevailing moral intuitions in some other sense constitute privileged data? (12) Relativism—Does the truth or warrant of moral judgments depend directly upon individually- or socially-adopted norms or practices? (13) Pluralism—Is there a uniquely good form of life or a uniquely right moral code, or could different forms of life or moral codes be appropriate in different circumstances?

Here, then, are the approximate coordinates of my own view in this multidimensional conceptual space. I will argue for a form of moral realism which holds that moral judgments can bear truth values in a fundamentally non-epistemic sense of truth; that moral properties are objective, though relational; that moral properties supervene upon natural properties, and may be reducible to them; that moral inquiry is of a piece with empirical inquiry; that it cannot be known *a priori* whether bivalence holds for moral judgments or how determinately such judgments can be assessed; that there is reason to think we know a fair amount about morality, but also reason to think that current moralities are wrong in certain ways and could be wrong in quite general ways; that a rational agent may fail to have a reason for obeying moral imperatives, although they may nonetheless be applicable to him; and that, while there are perfectly general criteria of moral assessment, nonetheless, by the nature of these criteria no one kind of life is likely to be appropriate for all individuals and no one set of norms appropriate for all societies and all times. The position thus described might well be called 'stark, raving moral realism', but for the sake of syntax, I will colorlessly call it 'moral realism'. This usage is not proprietary. Other positions, occupying more or less different coordinates, may have equal claim to either name.

## II. THE FACT/VALUE DISTINCTION

Any attempt to argue for a naturalistic moral realism runs headlong into the fact/value distinction. Philosophers have given various accounts of this distinction, and of the arguments for it, but for present purposes I will focus upon several issues concerning the epistemic and ontological status of judgments of value as opposed to judgments of fact.

Perhaps the most frequently heard argument for the fact/value distinction is epistemic: it is claimed that disputes over questions of value can persist even after all rational or scientific means of adjudication have been deployed, hence, value judgments cannot be cognitive in the sense that factual or logical judgments are. This claim is defended in part by appeal to the instrumental (hypothetical) character of reason, which prevents reason from dictating ultimate values. In principle, the argument runs, two individuals who differ in ultimate values could, without manifesting any rational defect, hold fast to their conflicting values in the face of any amount of argumentation or evidence. As Ayer puts it, "we find that argument is possible on moral questions only if some system of values is presupposed."<sup>2</sup>

One might attempt to block this conclusion by challenging the instrumental conception of rationality. But for all its faults and for all that it needs to be developed, the instrumental conception seems to me the clearest notion we have of what it is for an agent to have reasons to act. Moreover, it captures a central normative feature of reason-giving, since we can readily see the commending force for an agent of the claim that a given act would advance his ends. It would be hard to make much sense of someone who sincerely claimed to have certain ends and yet at the same time insisted that they could not provide him even *prima facie* grounds for action. (Of course, he might also believe that he has other, perhaps counter-vailing, grounds.)

Yet this version of the epistemic argument for the fact/value distinction is in difficulty even granting the instrumental conception of rationality. From the standpoint of instrumental reason,

---

<sup>2</sup>A. J. Ayer, *Language, Truth, and Logic* (New York: Dover, 1952), p. 111.

belief-formation is but one activity among others: to the extent that we have reasons for engaging in it, or for doing it one way rather than another, these are at bottom a matter of its contribution to our ends.<sup>3</sup> What it would be rational for an individual to believe on the basis of a given experience will vary not only with respect to his other beliefs, but also with respect to what he desires.<sup>4</sup> From this it follows that no amount of mere argumentation or experience could force one on pain of irrationality to accept even the factual claims of empirical science. The long-running debate over inductive logic well illustrates that rational choice among competing hypotheses requires much richer and more controversial criteria of theory choice than can be squeezed from instrumental reason alone. Unfortunately for the contrast Ayer wished to make, we find that argument is possible on scientific questions only if some system of values is presupposed.

However, Hume had much earlier found a way of marking the distinction between facts and values without appeal to the idea that induction—or even deduction—could require a rational agent to adopt certain beliefs rather than others when this would conflict with his contingent ends.<sup>5</sup> For Hume held the thesis that morality is practical, by which he meant that if moral facts existed, they would necessarily provide a reason (although perhaps not an overriding

---

<sup>3</sup>In saying this, I am insisting that questions about what it would be rational to believe belong to practical rather than theoretical reason. While results of theoretical reason—for example, conclusions of deductive inferences—are in general relevant to questions about rational belief, they are not determinative apart from the agent's practical reasons.

<sup>4</sup>Of course, individual belief-formation is not typically governed by explicit means-end reasoning, but rather by habits of belief-formation and tendencies to invest varying degrees of confidence in particular kinds of beliefs. If we accept an instrumental account of rationality, then we can call such habits rational from the standpoint of the individual to the extent that they fit into a constellation of attitudes and tendencies that promote his ends. This matter will arise again in Section IV.

<sup>5</sup>Neither these remarks, nor those in subsequent paragraphs, are meant to be a serious exegesis of Hume's arguments, which admit of interpretations other than the one suggested here. I mean only to capture certain features of what I take Hume's arguments to be, for example, in Book III, Part I, Section I of *A Treatise of Human Nature*, edited by L. A. Selby-Bigge (Oxford: Clarendon, 1973), esp. pp. 465–466, and in Appendix I of *An Inquiry Concerning the Principles of Morals*, edited by C. W. Hendel (Indianapolis: Bobbs-Merrill, 1957), esp. pp. 111–112.

reason) for moral action to all rational beings, regardless of their particular desires. Given this thesis as a premise, the instrumental conception of rationality can clinch the argument after all, for it excludes the possibility of categorical reasons of this kind. By contrast, Hume did not suppose it to be constitutive of logic or science that the facts revealed by these forms of inquiry have categorical force for rational agents, so the existence of logical and scientific facts, unlike the existence of moral facts, is compatible with the instrumental character of reason.

Yet this way of drawing the fact/value distinction is only as compelling as the claim that morality is essentially practical in Hume's sense.<sup>6</sup> Hume is surely right in claiming there to be an intrinsic connection, no doubt complex, between valuing something and having some sort of positive attitude toward it that provides one with an instrumental reason for action. We simply would disbelieve someone who claimed to value honesty and yet never showed the slightest urge to act honestly when given an easy opportunity. But this is a fact about the connection between the values *embraced* by an individual and his reasons for action, not a fact showing a connection between moral evaluation and rational motivation.

Suppose for example that we accept Hume's characterization of justice as an artificial virtue directed at the general welfare. This is in a recognizable sense an evaluative or normative notion—"a value" in the loose sense in which this term is used in such debates—yet it certainly does not follow from its definition that every rational being, no matter what his desires, who believes that some or other act is just in this sense will have an instrumental reason to perform it. A rational individual may fail to value justice for its own sake, and may have ends contrary to it. In Hume's discussion of our "interested obligation" to be just, he seems to recognize that in the end it may not be possible to show that a "sensible knave" has a reason to be just. Of course, Hume held that the rest of us—whose

---

<sup>6</sup>Philippa Foot has questioned this thesis, although her way of posing and arguing the question differs enough from mine that I cannot judge whether she would be in agreement with the argument that follows. See her *Virtues and Vices* (Berkeley: University of California Press, 1978), especially Essay XI. The presentation of the issues here owes its main inspiration to William K. Frankena's distinction between the rational and the moral points of view.

hearts rebel at Sensible Knave's attitude that he may break his word, cheat, or steal whenever it suits his purposes—have reason to be just, to deem Knave's attitude unjust, and to try to protect ourselves from his predations.<sup>7</sup>

Yet Knave himself could say, perhaps because he accepts Hume's analysis of justice, "Yes, my attitude is unjust." And by Hume's own account of the relation of reason and passion, Knave could add "But what is that to me?" without failing to grasp the content of his previous assertion. Knave, let us suppose, has no doubts about the intelligibility or reality of "the general welfare," and thinks it quite comprehensible that people attach great significance in public life to the associated notion of justice. He also realizes that for the bulk of mankind, whose passions differ from his, being just is a source and a condition of much that is most worthwhile in life. He thus understands that appeals to justice typically have motivating force. Moreover, he himself uses the category of justice in analyzing the social world, and he recognizes—indeed, his knavish calculations take into account—the distinction between those individuals and institutions that truly are just, and those that merely appear just or are commonly regarded as just. Knave does view a number of concepts with wide currency—religious ones, for example—as mere fictions that prey on weak minds, but he does not view justice in this way. Weak minds and moralists have, he thinks, surrounded justice with certain myths—that justice is its own reward, that once one sees what is just one will automatically have a reason to do it, and so on. But then, he thinks that weak minds and moralists have likewise surrounded wealth and power with myths—that the wealthy are not truly happy, that the powerful inevitably ride for a fall, and so on—and he does not on this account doubt whether there are such things as wealth and power. Knave is glad to be free of prevailing myths about wealth, power, and justice; glad, too, that he is free in his own mind to pay as much or as little attention to any of these attributes as his desires and circumstances warrant. He might, for example, find Mae West's advice convincing: diamonds are very much worth acquiring, and "goodness ha[s] nothing to do with it."

---

<sup>7</sup>See the *Inquiry Concerning the Principles of Morals*, Sec. IX, Pt. II, pp. 102–103.

We therefore must distinguish the business of saying what an individual values from the business of saying what it is for him to make measurements against the criteria of a species of evaluation that he recognizes to be genuine.<sup>8</sup>

To deny Hume's thesis of the practicality of moral judgment, and so remove the ground of his contrast between facts and values, is not to deny that morality has an action-guiding character. Morality surely can remain prescriptive within an instrumental framework, and can recommend itself to us in much the same way that, say, epistemology does: various significant and enduring—though perhaps not universal—human ends can be advanced if we apply certain evaluative criteria to our actions. That may be enough to justify to ourselves our abiding concern with the epistemic or moral status of what we do.<sup>9</sup>

By arguing that reason does not compel us to adopt particular beliefs or practices apart from our contingent, and variable, ends, I may seem to have failed to negotiate my way past epistemic relativism, and thus to have wrecked the argument for moral realism before it has even left port. Rationality does go relative when it goes instrumental, but epistemology need not follow. The epistemic

---

<sup>8</sup>The ancient criticism of non-cognitivism that it has difficulty accounting for the difference between moral value and other sorts of desirability (so that Hume can speak in one breath of our approval of a man's "good offices" and his "well-contrived apartment"), gains some vitality in the present context. To account for such differences it is necessary to have a contentful way of characterizing criteria of moral assessment so that moral approval does not reduce to "is valued by the agent." (Such a characterization will be offered in Section IV.) Value *sans phrase* is a generic, and not necessarily moral, notion. One sometimes hears it said that generic value becomes moral in character when we reach that which the agent prizes above all else. But this would invest pets and mementos with moral value, and have the peculiar effect of making amorality a virtual conceptual impossibility. It seems more plausible to say that not all value is moral value, and that the highest values for an individual need not be, nor need they even seem to him to be, moral values. Once we turn to questions of duty, the situation should be clearer still: moral theorists have proposed quite different relations among the categories of moral rightness, moral goodness, and non-moral goodness, and it seems implausible to say that deeming an act or class of actions morally right is necessarily equivalent to viewing it personally as valuable *sans phrase*.

<sup>9</sup>The character of moral imperatives receives further discussion in Section V.

warrant of an individual's belief may be disentangled from the rationality of his holding it, for epistemic warrant may be tied to an external criterion—as it is for example by causal or reliabilist theories of knowledge.<sup>10</sup> It is part of the naturalistic realism that informs this essay to adopt such a criterion of warrant. We should not confuse the obvious fact that in general our ends are well served by reliable causal mechanisms of belief-formation with an internalist claim to the effect that reason requires us to adopt such means. Reliable mechanisms have costs as well as benefits, and successful pursuit of some ends—Knave would point to religious ones, and to those of certain moralists—may in some respects be incompatible with adoption of reliable means of inquiry.

This rebuttal of the charge of relativism invites the defender of the fact/value distinction to shift to ontological ground. Perhaps facts and values cannot be placed on opposite sides of an epistemological divide marked off by what reason and experience can compel us to accept. Still, the idea of reliable causal mechanisms for moral learning, and of moral facts “in the world” upon which they operate, is arguably so bizarre that I may have done no more than increase my difficulties.

### III. VALUE REALISM

The idea of causal interaction with moral reality certainly would be intolerably odd if moral facts were held to be *sui generis*;<sup>11</sup> but there need be nothing odd about causal mechanisms for learning moral facts if these facts are constituted by natural facts, and that is the view under consideration. This response will remain unconvincing, however, until some positive argument for realism about moral facts is given. So let us turn to that task.

What might be called ‘the generic stratagem of naturalistic realism’ is to postulate a realm of facts in virtue of the contribution they

---

<sup>10</sup>Such theories are suitably externalist when, in characterizing the notions of *reliability* or *warrant-conferring causal process*, they employ an account of truth that does not resolve truth into that which we have reason to believe—for example, a nontrivial correspondence theory.

<sup>11</sup>Or if moral facts were supposed to be things of a kind to provide categorical reasons for action. However, this supposition is simply Hume's thesis of practicality in ontological garb.

would make to the *a posteriori* explanation of certain features of our experience. For example, an external world is posited to explain the coherence, stability, and intersubjectivity of sense-experience. A moral realist who would avail himself of this stratagem must show that the postulation of moral facts similarly can have an explanatory function. The stratagem can succeed in either case only if the reality postulated has these two characteristics:

- (1) *independence*: it exists and has certain determinate features independent of whether we think it exists or has those features, independent, even, of whether we have good reason to think this;
- (2) *feedback*: it is such—and we are such—that we are able to interact with it, and this interaction exerts the relevant sort of shaping influence or control upon our perceptions, thought, and action.

These two characteristics enable the realist's posit to play a role in the explanation of our experience that cannot be replaced without loss by our mere *conception* of ourselves or our world. For although our conceptual scheme mediates even our most basic perceptual experiences, an experience-transcendent reality has ways of making itself felt without the permission of our conceptual scheme—causally. The success or failure of our plans and projects famously is not determined by expectation alone. By resisting or yielding to our worldly efforts in ways not anticipated by our going conceptual scheme, an external reality that is never directly revealed in perception may nonetheless significantly influence the subsequent evolution of that scheme.

The realist's use of an external world to explain sensory experience has often been criticized as no more than a picture. But do we even have a picture of what a realist explanation might look like in the case of values?<sup>12</sup> I will try to sketch one, filling in first a realist

---

<sup>12</sup>J. L. Mackie, in *Ethics: Inventing Right and Wrong* (Harmondsworth, Middlesex: Penguin, 1977), and Gilbert Harman, in *The Nature of Morality: An Introduction to Ethics* (New York: Oxford University Press, 1977), both challenge moral realism in part by questioning its capacity to explain. Nicholas L. Sturgeon, in "Moral Explanations," David Copp and David Zimmerman, eds., *Morality, Reason and Truth: New Essays in the Foundations of Ethics* (Totowa, N.J.: Rowman and Allanhead, 1984), takes the opposite side, using arguments different from those offered below.

account of non-moral value—the notion of something being desirable for someone, or good for him.<sup>13</sup>

Consider first the notion of someone's *subjective interests*—his wants or desires, conscious or unconscious. Subjective interest can be seen as a secondary quality, akin to taste. For me to take a subjective interest in something is to say that it has a positive *valence* for me, that is, that in ordinary circumstances it excites a positive attitude or inclination (not necessarily conscious) in me. Similarly, for me to say that I find sugar sweet is to say that in ordinary circumstances sugar excites a certain gustatory sensation in me. As secondary qualities, subjective interest and perceived sweetness supervene upon primary qualities of the perceiver, the object (or other phenomenon) perceived, and the surrounding context: the perceiver is so constituted that this sort of object in this sort of context will excite that sort of sensation. Call this complex set of relational, dispositional, primary qualities the *reduction basis* of the secondary quality.

We have in this reduction basis an objective notion that corresponds to, and helps explain, subjective interests. But it is not a plausible foundation for the notion of non-moral goodness, since the subjective interests it grounds have insufficient normative force to capture the idea of desirableness. My subjective interests frequently reflect ignorance, confusion, or lack of consideration, as hindsight attests. The fact that I am now so constituted that I desire something which, had I better knowledge of it, I would wish I had never sought, does not seem to recommend it to me as part of my good.

To remedy this defect, let us introduce the notion of an *objectified subjective interest* for an individual *A*, as follows.<sup>14</sup> Give to an actual individual *A* unqualified cognitive and imaginative powers, and full factual and nomological information about his physical and psy-

---

<sup>13</sup>A full-scale theory of value would, I think, show the concept of someone's good to be slightly different from the concept of what is desirable for him. However, this difference will not affect the argument made here.

<sup>14</sup>It was some work by Richard C. Jeffrey on epistemic probability that originally suggested to me the idea of objectifying subjective interests. See note 17. I have since benefited from Richard B. Brandt's work on "rational desire," although I fear that what I will say contains much that he would regard as wrong-headed. See *A Theory of the Good and the Right* (Oxford: Clarendon, 1979), Part I.

chological constitution, capacities, circumstances, history, and so on. *A* will have become *A+*, who has complete and vivid knowledge of himself and his environment, and whose instrumental rationality is in no way defective. We now ask *A+* to tell us not what *he* currently wants, but what he would want his non-idealized self *A* to want—or, more generally, to seek—were he to find himself in the actual condition and circumstances of *A*.<sup>15</sup> Just as we assumed there to be a reduction basis for an individual *A*'s actual subjective interests, we may assume there to be a reduction basis for his objectified subjective interests, namely, those facts about *A* and his circumstances that *A+* would combine with his general knowledge in arriving at his views about what he would want to want were he to step into *A*'s shoes.

For example, Lonnie, a traveler in a foreign country, is feeling miserable. He very much wishes to overcome his malaise and to settle his stomach, and finds he has a craving for the familiar: a tall glass of milk. The milk is desired by Lonnie, but is it also desirable for him? Lonnie-Plus can see that what is wrong with Lonnie, in addition to homesickness, is dehydration, a common affliction of tourists, but one often not detectable from introspective evidence. The effect of drinking hard-to-digest milk would be to further

---

<sup>15</sup>We ask this question of *A+*, rather than what *A+* wants for himself, because we are seeking the objectified subjective interests of *A*, and the interests of *A+* might be quite different owing to the changes involved in the idealization of *A*. For example, *A+* presumably does not want any more information for himself—there is no more to be had and he knows this. Yet it might still be true that *A+* would want to want more knowledge were he to be put in the place of his less well-informed self, *A*. It may as a psychological matter be impossible for *A+* to set aside entirely his desires *in his present circumstances* with regard to himself or to *A* in considering what he would want to want were he to be put *in the place of* his less-than-ideal self. This reveals a measurement problem for objective interests: giving an individual the information and capacities necessary to “objectify” his interests may perturb his psychology in ways that alter the phenomenon we wish to observe. Such difficulties attend even the measurement of subjective interests, since instruments for sampling preferences (indeed, mere acts of reflection upon one's preferences) tend to affect the preferences expressed. For obvious reasons, interference effects come with the territory. Though not in themselves sufficient ground for skepticism about subjective or objective interests, these measurement problems show the need for a “perturbation theory,” and for caution about attributions of interests that are inattentive to interference effects.

unsettle Lonnie's stomach and worsen his dehydration. By contrast, Lonnie-Plus can see that abundant clear fluids would quickly improve Lonnie's physical condition—which, incidentally, would help with his homesickness as well. Lonnie-Plus can also see just how distasteful Lonnie would find it to drink clear liquids, just what would happen were Lonnie to continue to suffer dehydration, and so on. As a result of this information, Lonnie-Plus might then come to desire that were he to assume Lonnie's place, he would want to drink clear liquids rather than milk, or at least want to act in such a way that a want of this kind would be satisfied. The reduction basis of this objectified interest includes facts about Lonnie's circumstances and constitution, which determine, among other things, his existing tastes and his ability to acquire certain new tastes, the consequences of continued dehydration, the effects and availability of various sorts of liquids, and so on.

Let us say that this reduction basis is the constellation of primary qualities that make it be the case that the Lonnie has a certain *objective interest*.<sup>16</sup> That is, we will say that Lonnie has an objective interest in drinking clear liquids in virtue of this complex, relational, dispositional set of facts. Put another way, we can say that the reduction basis, not the fact that Lonnie-Plus would have certain wants, is the truth-maker for the claim that this is an objective interest of Lonnie's. The objective interest thus explains why there is a certain objectified interest, not the other way around.<sup>17</sup>

---

<sup>16</sup>'Interest' is not quite the word wanted here, for in ordinary language we may speak of a want where we would not speak of a corresponding interest. See Brian Barry, *Political Argument* (London: Routledge and Kegan Paul, 1965), especially Chapter X, for discussion. A more accurate, but overly cumbersome, expression would be 'positive-valence-making characteristic'.

<sup>17</sup>Suppose for a moment, contrary to what was urged above, that there is a workable notion of epistemic probability that determines rational degrees of belief independent of the contingent goals of the epistemic agent. Perhaps then the following analogy will be helpful. Consider a physically random process, such as alpha-decay. We can ask an individual what subjective probability he would assign to an event consisting in a certain rate of decay for a given sample of uranium; we can also ask what rational degree of belief the individual would assign to this event were he to become ideally informed about the laws of physics and the relevant initial conditions. Call the latter rational degree of belief the *objectified subjective probability* of the event, and suppose it to be equal to one fifth. (Compare Richard C.

Let us now say that  $X$  is *non-morally good for A* if and only if  $X$  would satisfy an objective interest of  $A$ .<sup>18</sup> We may think of  $A+$ 's views about what he would want to want were he in  $A$ 's place as generating a ranking of potential objective interests of  $A$ , a ranking that will reflect what is better or worse for  $A$  and will allow us to speak of  $A$ 's actual wants as better or worse approximations of what is best for him. We may also decompose  $A+$ 's views into *prima facie* as opposed to "on balance" objective interests of  $A$ , the former yielding the notion of 'a good for  $A$ ', the latter, of 'the good for  $A$ '.<sup>19</sup>

---

Jeffrey, *The Logic of Decision* (New York: McGraw-Hill, 1964), pp. 190–196.) But now consider the physical facts that, in conjunction with the laws of quantum mechanics, ground the idealized individual's judgment. Call these the *reduction basis* of that judgment. This reduction basis is a complex set of primary qualities that can be said to bring it about that the event in question has an *objective probability* of one fifth. (It should be said that it is not part of Jeffrey's approach to posit such objective probabilities.) The existence of this objective probability can explain why an ideally informed individual would select an objectified subjective probability equal to one fifth, but the probability judgment of an ideally informed individual cannot explain why the objective probability is one fifth—that is a matter of the laws of physics. Similarly, the existence of an individual's objective interest can explain why his ideally informed self would pick out for his less-informed self a given objectified subjective interest, but not *vice versa*.

<sup>18</sup>More precisely, we may say that  $X$  is non-morally good for  $A$  at time  $t$  if and only if  $X$  would satisfy an objective interest of  $A$  the reduction basis of which exists at  $t$ . Considerations about the evolution of interests over time raise a number of issues that cannot be entered into here.

<sup>19</sup> $A+$ , putting himself in  $A$ 's place, may find several different sets of wants equally appealing, so that several alternatives could be equal-best for  $A$  in this sense. This would not make the notion of 'the good for  $A$ ' problematic, just pluralistic. However, a more serious question looms. Is there sufficient determinacy in the specification of  $A+$ 's condition, or in the psychology of desire, to make the notion of objective interest definite enough for my purposes? Without trying to say how definite *that* might be, let me suggest two ways in which an answer to the worry about definiteness might begin. (1) It seems that we do think that there are rather definite answers to questions about how an individual  $A$ 's desires would change were his beliefs to change in certain limited ways. If Lonnie were to learn the consequences of drinking milk, he would no longer want his desire for milk to be effective. But a large change in belief can be accomplished piecemeal, by a sequence of limited changes in belief. Thus, if (admittedly, a big 'if') *order* of change is not in the end significant, then the facts and generalizations that support counterfactuals about limited changes might support an extrapolation all the way to  $A+$ . (2) Beliefs and desires appear to co-vary systematically. Typically, we find that individuals who differ markedly in their desires—for example, about careers or style of life—

This seems to me an intuitively plausible account of what someone's non-moral good consists in: roughly, what he would want himself to seek if he knew what he were doing.<sup>20</sup>

Moreover, this account preserves what seems to me an appropriate link between non-moral value and motivation. Suppose that one desires *X*, but wonders whether *X* really is part of one's good. This puzzlement typically arises because one feels that one knows too little about *X*, oneself, or one's world, or because one senses that one is not being adequately rational or reflective in assessing the information one has—perhaps one suspects that one has been captivated by a few salient features of *X* (or repelled by a few salient features of its alternatives). If one were to learn that one would still want oneself to want *X* in the circumstances were one to view things with full information and rationality, this presumably would reduce the force of the original worry. By contrast, were one to learn that when fully informed and rational one would want oneself *not* to want *X* in the circumstances, this presumably would add force to it. Desires being what they are, a reinforced worry might not be sufficient to remove the desire for *X*. But if one were to become

---

differ markedly, and characteristically, in their beliefs; as individuals become more similar in their beliefs, they tend to become more similar in their desires. This suggests that if (another big 'if') the characterization given of *A*+ fixes the entire content of his beliefs in a definite way (at least, given a choice of language), then his desires may be quite comprehensively fixed as well. If we had in hand a general theory of the co-variation of beliefs and desires, then we could appeal directly to this theory—plus facts about *A*—to ground the counterfactuals needed to characterize *A*'s objectified interests, eliminating any essential reference to the imaginary individual *A*+.

<sup>20</sup>The account may, however, yield some counterintuitive results. Depending upon the nature and circumstances of given individuals, they might have objective interests in things we find wrong or repulsive, and that do not seem to us part of a good life. We can explain a good deal of our objection to certain desires—for example, those involving cruelty—by saying that they are not *morally* good; others—for example, those of a philistine nature—by saying that they are not *aesthetically* valuable; and so on. It seems to me preferable to express our distaste for certain ends in terms of specific categories of value, rather than resort to the device of saying that such ends could under no circumstances be part of anyone's non-moral good. People, or at least some people, might be put together in a way that makes some not-very-appetizing things essential to their flourishing, and we do not want to be guilty of wishful thinking on this score. (There will be wishful thinking enough before we are through.)

genuinely and vividly convinced that one's desire for *X* is in this sense not supported by full reflection upon the facts, one presumably would feel this to be a count against acting upon the desire. This adjustment of desire to belief might not in a given case be required by reason or logic; it might be "merely psychological." But it is precisely such psychological phenomena that naturalistic theories of value take as basic.

In what follows, we will need the notion of intrinsic goodness, so let us say that *X* is *intrinsically non-morally good for A* just in case *X* is in *A*'s objective interest without reference to any other objective interest of *A*. We can in an obvious way use the notion of objective intrinsic interest to account for all other objective interests. Since individuals and their environments differ in many respects, we need not assume that everyone has the same objective intrinsic interests. *A fortiori*, we need not assume that they have the same objective instrumental interests. We should, however, expect that when personal and situational similarities exist across individuals—that is, when there are similarities in reduction bases—there will to that extent be corresponding similarities in their interests.

It is now possible to see how the notion of non-moral goodness can have explanatory uses. For a start, it can explain why one's actual desires have certain counterfactual features, for example, why one would have certain hypothetical desires rather than others were one to become fully informed and aware. Yet this sort of explanatory use—following as it does directly from the definition of objective interest—might well be thought unimpressive unless some other explanatory functions can be found.

Consider, then, the difference between Lonnie and Tad, another traveler in the same straits, but one who, unlike Lonnie, wants to drink clear liquids, and proceeds to do so. Tad will perk up while Lonnie remains listless. We can explain this difference by noting that although both Lonnie and Tad acted upon their wants, Tad's wants better reflected his interests. The congruence of Tad's wants with his interests may be fortuitous, or it may be that Tad knows he is dehydrated and knows the standard treatment. In the latter case we would ordinarily say that the explanation of the difference in their condition is that Tad, but not Lonnie, "knew what was good for him."

Generally, we can expect that what *A*+ would want to want were

he in *A*'s place will correlate well with what would permit *A* to experience physical or psychological well-being or to escape physical or psychological ill-being. Surely our well- or ill-being are among the things that matter to us most, and most reliably, even on reflection.<sup>21</sup> Appeal to degrees of congruence between *A*'s wants and his interests thus will often help to explain facts about how satisfactory he finds his life. Explanation would not be preserved were we to substitute 'believed to be congruent' for 'are (to such-and-such a degree) congruent', since, as cases like Lonnie's show, even if one were to convince oneself that one's wants accurately reflected one's interests, acting on these wants might fail to yield much satisfaction.

In virtue of the correlation to be expected between acting upon motives that congrue with one's interests and achieving a degree of satisfaction or avoiding a degree of distress, one's objective interests may also play an explanatory role in the *evolution* of one's desires. Consider what I will call the *wants/interests mechanism*, which permits individuals to achieve selfconscious and unselfconscious learning about their interests through experience. In the simplest sorts of cases, trial and error leads to the selective retention of wants that are satisfiable and lead to satisfactory results for the agent.

For example, suppose that Lonnie gives in to his craving and

---

<sup>21</sup>To put the matter in more strictly naturalistic terms, we can expect that evolution will have favored organisms so constituted that those behaviors requisite to their survival and flourishing are associated with positive internal states (such as pleasure) and those opposed to survival or flourishing with negative states (such as pain). 'Flourishing' here, even if understood as mere reproductive fitness, is not a narrow notion. In order for beings such as humans to be reproductively successful, they must as phenotypes have lives that are psychologically sustainable, internally motivating, and effectively social; lives, moreover, that normally would engage in a wide range of their peculiarly human capacities. Humankind could hardly have been a success story even at the reproductive level were not pursuit of the sorts of things that characteristically have moved humans to action associated with existences of this kind. However, it must be kept in mind that most human evolution occurred under circumstances different in important ways from the present. It therefore is quite possible that the interaction of evolved human motivational potentials with existing circumstances will produce incongruities between what we tend to aim at, or to be driven by, and what would produce the greatest pleasure for us. That is one reason for doubting hedonism as a theory of motivation.

drinks the milk. Soon afterwards, he feels much worse. Still unable to identify the source of his malaise and still in the grips of a desire for the familiar, his attention is caught by a green-and-red sign in the window of a small shop he is moping past: “7-Up,” it says. He rushes inside and buys a bottle. Although it is lukewarm, he drinks it eagerly. “Mmm,” he thinks, “I’ll have another.” He buys a second bottle, and drains it to the bottom. By now he has had his fill of tepid soda, and carries on. Within a few hours, his mood is improving. When he passes the store again on the way back to his hotel, his pleasant association with drinking 7-Up leads him to buy some more and carry it along with him. That night, in the dim solitude of his room, he finds the soda’s reassuringly familiar taste consoling, and so downs another few bottles before finally finding sleep. When he wakes up the next morning, he feels very much better. To make a dull story short: the next time Lonnie is laid low abroad, he may have some conscious or unconscious, reasoned or superstitious, tendency to seek out 7-Up. Unable to find that, he might seek something quite like it, say, a local lime-flavored soda, or perhaps even the *agua mineral con gaz* he had previously scorned. Over time, as Lonnie travels more and suffers similar malaise, he regularly drinks clearish liquids and regularly feels better, eventually developing an actual desire for such liquids—and an aversion to other drinks, such as milk—in such circumstances.

Thus have Lonnie’s desires evolved through experience to conform more closely to what is good for him, in the naturalistic sense intended here. The process was not one of an ideally rational response to the receipt of ideal information, but rather of largely unreflective experimentation, accompanied by positive and negative associations and reinforcements. There is no guarantee that the desires “learned” through such feedback will accurately or completely reflect an individual’s good. Still less is there any guarantee that, even when an appropriate adjustment in desire occurs, the agent will comprehend the origin of his new desires or be able to represent to himself the nature the interests they reflect. But then, it is a quite general feature of the various means by which we learn about the world that they may fail to provide accurate or comprehending representations of it. My ability to perceive and understand my surroundings coexists with, indeed draws upon the same mechanisms as, my liability to deception by illusion, expectation, or surface appearance.

There are some broad theoretical grounds for thinking that something like the wants/interests mechanism exists and has an important role in desire-formation. Humans are creatures motivated primarily by wants rather than instincts. If such creatures were unable through experience to conform their wants at all closely to their essential interests—perhaps because they were no more likely to experience positive internal states when their essential interests are met than when they are not—we could not expect long or fruitful futures for them. Thus, if humans in general did not come to want to eat the kinds of food necessary to maintain some degree of physical well-being, or to engage in the sorts of activities or relations necessary to maintain their sanity, we would not be around today to worry whether we can know what is good for us. Since creatures as sophisticated and complex as humans have evolved through encounters with a variety of environments, and indeed have made it their habit to modify their environments, we should expect considerable flexibility in our capacity through experience to adapt our wants to our interests. However, this very flexibility makes the mechanism unreliable: our wants may at any time differ arbitrarily much from our interests; moreover, we may fail to have experiences that would cause us to notice this, or to undergo sufficient feedback to have much chance of developing new wants that more nearly approximate our interests. It is entirely possible, and hardly infrequent, that an individual live out the course of a normal life without ever recognizing or adjusting to some of his most fundamental interests. Individual limitations are partly remedied by cultural want-acquiring mechanisms, which permit learning and even theorizing over multiple lives and life-spans, but these same mechanisms also create a vast potential for the inculcation of wants at variance with interests.

The argument for the wants/interests mechanism has about the same status, and the same breezy plausibility, as the more narrowly biological argument that we should expect the human eye to be capable of detecting objects the size and shape of our predators or prey. It is not necessary to assume anything approaching infallibility, only enough functional success to hold our own in an often inhospitable world.<sup>22</sup>

---

<sup>22</sup>'Functional success' rather than 'representational accuracy' for the following reason. Selection favors organisms that have some-or-other fea-

Thus far the argument has concerned only those objective interests that might be classified as needs, but the wants/interests mechanism can operate with respect to any interest—even interests related to an individual's particular aptitudes or social role—whose frustration is attended even indirectly by consciously or unconsciously unsatisfactory results for him. (To be sure, the more indirect the association the more unlikely that the mechanism will be reliable.) For example, the experience of taking courses in both mathematics and philosophy may lead an undergraduate who thought himself cut out to be a mathematician to come to prefer a career in philosophy, which would in fact better suit his aptitudes and attitudes. And a worker recently promoted to management from the shop floor may find himself less inclined to respond to employee grievances than he had previously wanted managers to be, while his former co-workers may find themselves less inclined to confide in him than before.

If a wants/interests mechanism is postulated, and if what is non-morally good for someone is a matter of what is in his objective interest, then we can say that objective value is able to play a role in the explanation of subjective value of the sort the naturalistic realist about value needs. These explanations even support some qualified predictions: for example, that, other things equal, individuals will ordinarily be better judges of their own interests than third parties; that knowledge of one's interests will tend to increase with increased experience and general knowledge; that people with similar personal and social characteristics will tend to have similar values; and that there will be greater general consensus upon what is desirable in those areas of life where individuals are most alike in other regards (for example, at the level of basic motives), and where trial-and-error mechanisms can be expected to work well (for example, where esoteric knowledge is not required). I am in no

---

ture that happens in their particular environment to contribute to getting their needs met. Whether that feature will be an accurate representational capacity cannot be settled by an argument of this kind. Of course, it would be a very great coincidence if beings who rely as heavily upon representations as we do were able to construct only grossly inaccurate representations while at the same time managing successfully in a range of environments over a long period of time. But such coincidences cannot be ruled out.

position to pronounce these predictions correct, but it may be to their credit that they accord with widely-held views.

It should perhaps be emphasized that although I speak of the objectivity of value, the value in question is human value, and exists only because humans do. In the sense of old-fashioned theory of value, this is a relational rather than absolute notion of goodness. Although relational, the relevant facts about humans and their world are objective in the same sense that such non-relational entities as stones are: they do not depend for their existence or nature merely upon our conception of them.<sup>23</sup>

Thus understood, objective interests are supervenient upon natural and social facts. Does this mean that they cannot contribute to explanation after all, since it should always be possible in principle to account for any particular fact that they purport to explain by reference to the supervenience basis alone? If mere supervenience were grounds for denying an explanatory role to a given set of concepts, then we would have to say that chemistry, biology, and electrical engineering, which clearly supervene upon physics, lack explanatory power. Indeed, even outright reducibility is no ground for doubting explanatoriness. To establish a relation of reduction between, for example, a chemical phenomenon such as valence and a physical model of the atom does nothing to suggest that there is no such thing as valence, or that generalizations involving valence cannot support explanations. There can be no issue here of ontological economy or eschewing unnecessary entities, as might be the case if valence were held to be something *sui generis*, over and above any constellation of physical properties. The facts described in principles of chemical valence are genuine, and permit a powerful and explanatory systematization of chemical combination; the

---

<sup>23</sup>Although some elements of their reduction basis depend upon our past choices, our objective interests are not therefore subjective in a sense damaging to the present argument. After all, such unproblematically objective facts about us as our weight, income, and spatial location depend in the same way upon past choices. The point is not that our subjective interests have no role in shaping the reduction basis of our objective interests, but rather that they can affect our objective interests only in virtue of their actual (rather than merely desired) effects upon this reduction basis, just as they can affect our weight, income, or spatial location only in virtue of actual (rather than merely desired) effects upon our displacement, employment, or movement.

existence of a successful reduction to atomic physics only bolsters these claims.

We are confident that the notion of chemical valence is explanatory because proffered explanations in terms of chemical valence insert explananda into a distinctive and well-articulated nomic nexus, in an obvious way increasing our understanding of them. But what comparably powerful and illuminating theory exists concerning the notion of objective interest to give us reason to think—whether or not strict reduction is possible—that proffered explanations using this notion are genuinely informative?

I would find the sort of value realism sketched here uninteresting if it seemed to me that no theory of any consequence could be developed using the category of objective value. But in describing the wants/interests mechanism I have already tried to indicate that such a theory may be possible. When we seek to explain why people act as they do, why they have certain values or desires, and why sometimes they are led into conflict and other times into cooperation, it comes naturally to common sense and social science alike to talk in terms of people's interests. Such explanations will be incomplete and superficial if we remain wholly at the level of subjective interests, since these, too, must be accounted for.<sup>24</sup>

#### IV. NORMATIVE REALISM

Suppose everything said thus far to have been granted generously. Still, I would as yet have no right to speak of *moral* realism,

---

<sup>24</sup>In a similar way, it would be incomplete and superficial to explain why, once large-scale production became possible, the world's consumption of refined sugar underwent such explosive development, by mentioning only the fact that people liked its taste. Why, despite wide differences in traditional diet and acquired tastes, has sugar made such inroads into human consumption? Why haven't the appearance and promotion of other equally cheap foodstuffs produced such remarkable shifts in consumption? Why, even in societies where sugar is recognized as a health hazard, does consumption of sugars, often in concealed forms, continue to climb? Facts about the way we are constituted, about the rather singular ways sugar therefore affects us, and about the ways forms of production and patterns of consumption co-evolved to generate both a growing demand and an expanding supply, must supplement a theory that stops at the level of subjective preferences. See Sidney W. Mintz, *Sweetness and Power: The Place of Sugar in Modern History* (New York: Viking, 1985) for relevant discussion.

for I have done no more than to exhibit the possibility of a kind of realism with regard to non-moral goodness, a notion that perfect moral skeptics can admit. To be entitled to speak of moral realism I would have to show realism to be possible about distinctively moral value, or moral norms. I will concentrate on moral norms—that is, matters of moral rightness and wrongness—although the argument I give may, by extension, be applied to moral value. In part, my reason is that normative realism seems much less plausible intuitively than value realism. It therefore is not surprising that many current proposals for moral realism focus essentially upon value—and sometimes only upon what is in effect non-moral value. Yet on virtually any conception of morality, a moral theory must yield an account of rightness.

Normative moral realism is implausible on various grounds, but within the framework of this essay, the most relevant is that it seems impossible to extend the generic strategy of naturalistic realism to moral norms. Where is the place in explanation for facts about what *ought* to be the case—don't facts about the way things *are* do all the explaining there is to be done? Of course they do. But then, my naturalistic moral realism commits me to the view that facts about what ought to be the case are facts of a special kind about the way things are. As a result, it may be possible for them to have a function within an explanatory theory. To see how this could be, let me first give some examples of explanations outside the realm of morality that involve naturalized norms.

“Why did the roof collapse?—For a house that gets the sort of snow loads that one did, the rafters ought to have been 2×8's at least, not 2×6's.” This explanation is quite acceptable, as far as it goes, yet it contains an ‘ought’. Of course, we can remove this ‘ought’ as follows: “If a roof of that design is to withstand the snow load that one bore, then it must be framed with rafters at least 2×8 in cross-section.” An architectural ‘ought’ is replaced by an engineering ‘if . . . then . . .’. This is possible because the ‘ought’ clearly is hypothetical, reflecting the universal architectural goal of making roofs strong enough not to collapse. Because the goal is contextually fixed, and because there are more or less definite answers to the question of how to meet it, and moreover because the explanandum phenomenon is the result of a process that selects against instances that do not attain that goal, the ‘ought’-

containing account conveys explanatory information.<sup>25</sup> I will call this sort of explanation *criterial*: we explain why something happened by reference to a relevant criterion, given the existence of a process that in effect selects for (or against) phenomena that more (or less) closely approximate this criterion. Although the criterion is defined naturalistically, it may at the same time be of a kind to have a regulative role in human practice—in this case, in house-building.

A more familiar sort of criterial explanation involves norms of individual rationality. Consider the use of an instrumental theory of rationality to explain an individual's behavior in light of his beliefs and desires, or to account for the way an individual's beliefs change with experience.<sup>26</sup> Bobby Shaftoe went to sea because he believed it was the best way to make his fortune, and he wanted above all to make his fortune. Crewmate Reuben Ramsloe came to believe that he wasn't liked by the other deckhands because he saw that they taunted him and greeted his frequent lashings at the hands of the First Mate with unconcealed pleasure. These explanations work because the action or belief in question was quite rational for the agent in the circumstances, and because we correctly suppose both Shaftoe and Ramsloe to have been quite rational.

Facts about degrees of instrumental rationality enter into explanations in other ways as well. First, consider the question why Bobby Shaftoe has had more success than most like-minded individuals in achieving his goals. We may lay his success to the fact that Shaftoe is more instrumentally rational than most—perhaps he has greater-than-average acumen in estimating the probabilities of outcomes, or is more-reliable-than-average at deductive inference, or is more-imaginative-than-average in surveying alternatives.

---

<sup>25</sup>For a discussion of how informally expressed accounts may nonetheless convey explanatory information, see Section II of my "Probability, Explanation, and Information," *Synthese* 48 (1981), pp. 233–256.

<sup>26</sup>Such explanation uses a naturalized criterion when rationality is defined in terms of relative efficiency given the agent's beliefs and desires. A (more or less) rational agent is thus someone disposed to act in (more or less) efficient ways. There is a deep difficulty about calling such explanation naturalistic, for the constraints placed upon attributions of beliefs and desires by a "principle of charity" may compromise the claim that rational-agent explanations are empirical. Although I believe this difficulty can be overcome, this is hardly the place to start *that* argument.

Second, although we are all imperfect deliberators, our behavior may come to embody habits or strategies that enable us to approximate optimal rationality more closely than our deliberative defects would lead one to expect. The mechanism is simple. Patterns of beliefs and behaviors that do not exhibit much instrumental rationality will tend to be to some degree self-defeating, an incentive to change them, whereas patterns that exhibit greater instrumental rationality will tend to be to some degree rewarding, an incentive to continue them. These incentives may affect our beliefs and behaviors even though the drawbacks or advantages of the patterns in question do not receive conscious deliberation. In such cases we may be said to acquire these habits or strategies because they *are* more rational, without the intermediation of any *belief* on our part that they are. Thus, cognitive psychologists have mapped some of the unconscious strategies or heuristics we employ to enable our limited intellects to sift more data and make quicker and more consistent judgments than would be possible using more standard forms of explicit reasoning.<sup>27</sup> We unwittingly come to rely upon heuristics in part because they are selectively reinforced as a result of their instrumental advantages over standard, explicit reasoning, that is, in part because of their greater rationality. Similarly, we may, without realizing it or even being able to admit it to ourselves, develop patterns of behavior that encourage or discourage specific behaviors in others, such as the unconscious means by which we cause those whose company we do not enjoy not to enjoy our company. Finally, as children we may have been virtually incapable of making rational assessments when a distant gain required a proximate loss. Yet somehow over time we managed in largely nondeliberative ways to acquire various interesting habits, such as putting certain vivid thoughts about the immediate future at the periphery of our attention, which enable us as adults to march ourselves off to the dentist without a push from behind. Criterial explanation in terms of individual rationality thus extends to behaviors beyond the realm of deliberate action. And, as with the

---

<sup>27</sup>For a survey of the literature, see Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs: Prentice-Hall, 1980), where one unsurprisingly finds greater attention paid to drawbacks than advantages.

wants/interests mechanism, it is possible to see in the emergence of such behaviors something we can without distortion call learning.

Indeed, our tendency through experience to develop rational habits and strategies may cooperate with the wants/interests mechanism to provide the basis for an *extended* form of criterial explanation, in which an individual's rationality is assessed not relative to his occurrent beliefs and desires, but relative to his objective interests. The examples considered earlier of the wants/interests mechanism in fact involved elements of this sort of explanation, for they showed not only wants being adjusted to interests, but also behavior being adjusted to newly adjusted wants. Without appropriate alteration of behavior to reflect changing wants, the feedback necessary for learning about wants would not occur. With such alteration, the behavior itself may become more rational in the extended sense. An individual who is instrumentally rational is disposed to adjust means to ends; but one result of his undertaking a means—electing a course of study, or accepting a new job—may be a more informed assessment, and perhaps a reconsideration, of his ends.

The theory of individual rationality—in either its simple or its extended form—thus affords an instance of the sort needed to provide an example of normative realism. Evaluations of degrees of instrumental rationality play a prominent role in our explanations of individual behavior, but they simultaneously have normative force for the agent. Whatever other concerns an agent might have, it surely counts for him as a positive feature of an action that it is efficient relative to his beliefs and desires or, in the extended sense, efficient relative to beliefs and desires that would appropriately reflect his condition and circumstances.

The normative force of these theories of individual rationality does not, however, merely derive from their explanatory use. One can employ a theory of instrumental rationality to explain behavior while rejecting it as a normative theory of reasons, just as one can explain an action as due to irrationality without thereby endorsing unreason.<sup>28</sup> Instead, the connection between the normative and

---

<sup>28</sup>To recall a point from Section II: one may make assessments relative to particular evaluative criteria without thereby valuing that which satisfies them.

explanatory roles of the instrumental conception of rationality is traceable to their common ground: the human motivational system. It is a fact about us that we have ends and have the capacity for both deliberate action relative to our ends and nondeliberate adjustment of behavior to our ends. As a result, we face options among pathways across a landscape of possibilities variously valenced for us. Both when we explain the reasons for people's choices and the causes of their behavior and when we appeal to their intuitions about what it would be rational to decide or to do, we work this territory, for we make what use we can of facts about what does-in-fact or can-in-principle motivate agents.

Thus emerges the possibility of saying that facts exist about what individuals have reason to do, facts that may be substantially independent of, and more normatively compelling than, an agent's occurrent conception of his reasons. The argument for such realism about individual rationality is no stronger than the arguments for the double claim that the relevant conception of instrumental individual rationality has both explanatory power and the sort of commendatory force a theory of *reasons* must possess, but (although I will not discuss them further here) these arguments seem to me quite strong.

\* \* \*

Passing now beyond the theory of individual rationality, let us ask what criterial explanations involving distinctively moral norms might look like. To ask this, we need to know what distinguishes moral norms from other criteria of assessment. Moral evaluation seems to be concerned most centrally with the assessment of conduct or character where the interests of more than one individual are at stake. Further, moral evaluation assesses actions or outcomes in a peculiar way: the interests of the strongest or most prestigious party do not always prevail, purely prudential reasons may be subordinated, and so on. More generally, moral resolutions are thought to be determined by criteria of choice that are *non-indexical* and in some sense *comprehensive*. This has led a number of philosophers to seek to capture the special character of moral evaluation by identifying a *moral point of view* that is impartial, but equally concerned with all those potentially affected. Other ethical theo-

rists have come to a similar conclusion by investigating the sorts of reasons we characteristically treat as relevant or irrelevant in moral discourse. Let us follow these leads. We thus may say that moral norms reflect a certain kind of rationality, rationality not from the point of view of any particular individual, but from what might be called a social point of view.<sup>29</sup>

By itself, the equation of moral rightness with rationality from a social point of view is not terribly restrictive, for, depending upon what one takes rationality to be, this equation could be made by a utilitarian, a Kantian, or even a non-cognitivist. That is as it should be, for if it is to capture what is distinctive about moral norms, it should be compatible with the broadest possible range of recognized moral theories. However, once one opts for a particular conception of rationality—such as the conception of rationality as efficient pursuit of the non-morally good, or as autonomous and universal self-legislation, or as a noncognitive expression of hypothetical endorsement—this schematic characterization begins to assume particular moral content. Here I have adopted an instrumentalist conception of rationality, and this—along with the account given of non-moral goodness—means that the argument for moral realism given below is an argument that presupposes and purports to defend a particular substantive moral theory.<sup>30</sup>

What is this theory? Let me introduce an idealization of the notion of social rationality by considering what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information.<sup>31</sup> Because of the assumption of full and vivid informa-

---

<sup>29</sup>I realize that it is misleading to call a point of view that is “impartial, but equally concerned with all those potentially affected” a *social* point of view—some of those potentially affected may lie on the other side of an intersocial boundary. This complication will be set aside until Section V.

<sup>30</sup>It also means that the relation of moral criteria to criteria of individual rationality has become problematic, since there can be no guarantee that what would be instrumentally rational from any given individual’s point of view will coincide with what would be instrumentally rational from a social point of view.

<sup>31</sup>A rather strong thesis of interpersonal comparison is needed here for purposes of social aggregation. I am not assuming the existence of some single good, such as happiness, underlying such comparisons. Thus the moral theory in question, although consequentialist, aggregative, and

tion, the interests in question will be objective interests. Given the account of goodness proposed in Section III, this idealization is equivalent to what is rational from a social point of view with regard to the realization of intrinsic non-moral goodness. This seems to me to be a recognizable and intuitively plausible—if hardly uncontroversial—criterion of moral rightness. Relative moral rightness is a matter of relative degree of approximation to this criterion.

The question that now arises is whether the notion of degrees of moral rightness could participate in explanations of behavior or in processes of moral learning that parallel explanatory uses of the notion of degrees of individual rationality—especially, in the extended sense. I will try to suggest several ways in which it might.

Just as an individual who significantly discounts some of his interests will be liable to certain sorts of dissatisfaction, so will a social arrangement—for example, a form of production, a social or political hierarchy, etc.—that departs from social rationality by significantly discounting the interests of a particular group have a potential for dissatisfaction and unrest. Whether or not this potential will be realized depends upon a great many circumstances. Owing to socialization, or to other limitations on the experience or knowledge of members of this group, the wants/interests mechanism may not have operated in such a way that the wants of its members reflect their interests. As a result they may experience no direct frustration of their desires despite the discounting of their interests. Or, the group may be too scattered or too weak to mobilize effectively. Or, it may face overawing repression. On the other hand, certain social and historical circumstances favor the realization of this potential for unrest, for example, by providing members of this group with experiences that make them more likely to develop interest-con-

---

maximizing, is not equivalent to classical utilitarianism. I *am* assuming that when a choice is faced between satisfying interest *X* of *A* vs. satisfying interest *Y* of *B*, answers to the question “All else equal, would it matter more to me if I were *A* to have *X* satisfied than if I were *B* to have *Y* satisfied?” will be relatively determinate and stable across individuals under conditions of full and vivid information. A similar, though somewhat weaker, form of comparability-across-difference is presupposed when we make choices from among alternative courses of action that would lead us to have different desires in the future.

gruent wants, by weakening the existing repressive apparatus, by giving them new access to resources or new opportunities for mobilization, or merely by dispelling the illusion that change is impossible. In such circumstances, one can expect the potential for unrest to manifest itself.

Just as explanations involving assessments of individual rationality were not always replaceable by explanations involving individual *beliefs about* what would be rational, so, too, explanations involving assessments of social rationality cannot be replaced by explanations involving *beliefs about* what would be morally right. For example, discontent may arise because a society departs from social rationality, but not as a result of a belief that this is the case. Suppose that a given society is believed by all constituents to be just. This belief may help to stabilize it, but if in fact the interests of certain groups are being discounted, there will be a potential for unrest that may manifest itself in various ways—in alienation, loss of morale, decline in the effectiveness of authority, and so on—well before any changes in belief about the society's justness occur, and that will help explain why members of certain groups come to believe it to be unjust, if in fact they do.

In addition to possessing a certain sort of potential for unrest, societies that fail to approximate social rationality may share other features as well: they may exhibit a tendency toward certain religious or ideological doctrines, or toward certain sorts of repressive apparatus; they may be less productive in some ways (for example, by failing to develop certain human resources) and more productive in others (for example, by extracting greater labor from some groups at less cost), and thus may be differentially economically successful depending upon the conditions of production they face, and so on.

If a notion of social rationality is to be a legitimate part of empirical explanations of such phenomena, an informative characterization of the circumstances under which departures from, or approximations to, social rationality could be expected to lead to particular social outcomes—especially, of the conditions under which groups whose interests are sacrificed could be expected to exhibit or mobilize discontent—must be available. Although it cannot be known *a priori* whether an account of this kind is possible, one can see emerging in some recent work in social history and

historical sociology various elements of a theory of when, and how, a persisting potential for social discontent due to persistently sacrificed interests comes to be manifested.<sup>32</sup>

An individual whose wants do not reflect his interests or who fails to be instrumentally rational may, I argued, experience feedback of a kind that promotes learning about his good and development of more rational strategies. Similarly, the discontent produced by departures from social rationality may produce feedback that, at a social level, promotes the development of norms that better approximate social rationality. The potential for unrest that exists when the interests of a group are discounted is potential for pressure from that group—and its allies—to accord fuller recognition to their interests in social decision-making and in the socially-instilled norms that govern individual decision-making. It therefore is pressure to push the resolution of conflicts further in the direction required by social rationality, since it is pressure to give fuller weight to the interests of more of those affected. Such pressure may of course be more or less forceful or coherent; it may find the most diverse ideological expression; and it may produce outcomes more or less advantageous in the end to those exerting it.<sup>33</sup>

---

<sup>32</sup>See, for example, Barrington Moore, Jr., *The Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World* (Boston: Beacon, 1966) and *Injustice: The Social Bases of Obedience and Revolt* (White Plains, N.Y.: M. E. Sharpe, 1978); E. P. Thompson, *The Making of the English Working Class* (New York: Pantheon, 1963); William B. Taylor, *Drinking, Homicide, and Rebellion in Colonial Mexican Villages* (Stanford: Stanford University Press, 1979); Charles Tilly, *From Mobilization to Revolution* (Reading, Mass.: Addison-Wesley, 1978); and Charles Tilly, et al., *The Rebellious Century, 1830–1930* (Cambridge: Harvard University Press, 1975).

<sup>33</sup>A common theme in the works cited in note 32 is that much social unrest is re-vindicative rather than revolutionary, since the discontent of long-suffering groups often is galvanized into action only when customary entitlements are threatened or denied. The overt ideologies of such groups thus frequently are particularistic and conservative, even as their unrest contributes to the emergence of new social forms that concede greater weight to previously discounted interests. In a similar way, individuals often fail to notice irrationalities in their customary behavior until they are led by it into uncustomary difficulties, which then arouse a sense that something has gone wrong. For familiar reasons, a typical initial individual response is to attempt to retrieve the *status quo ante*, although genuine change may result from these restorative efforts.

Striking historical examples of the mobilization of excluded groups to promote greater representation of their interests include the rebellions against the system of feudal estates, and more recent social movements against restrictions on religious practices, on suffrage and other civil rights, and on collective bargaining.<sup>34</sup>

Of course, other mechanisms have been at work influencing the evolution of social practices and norms at the same time, some with the reverse effect.<sup>35</sup> Whether mechanisms working on behalf of the inclusion of excluded interests will predominate depends upon a complex array of social and historical factors. It would be silly to think either that the norms of any actual society will at any given stage of history closely approximate social rationality, or that there will be a univocal trend toward greater social rationality. Like the mechanisms of biological evolution or market economics, the mechanisms described here operate in an "open system" alongside other mechanisms, and do not guarantee optimality or even a monotonic approach to equilibrium. Human societies do not appear to have begun at or near equilibrium in the relevant sense, and so the strongest available claim might be that in the long haul, barring certain exogenous effects, one could expect an uneven secular trend toward the inclusion of the interests of (or interests represented by)

---

<sup>34</sup>It should be emphasized that these mechanisms do not presuppose a background of democratic institutions. They have extracted concessions even within societies that remained very hierarchical. See, for example, Taylor, *Drinking, Homicide, and Rebellion*.

<sup>35</sup>Indeed, the mechanism just described may push in several directions at once: toward the inclusion of some previously excluded interests, and toward the exclusion of some previously included interests. To be sure, if interests come to be excluded even though their social and material basis remains more or less intact, a new potential for unrest is created. Some groups present a special problem, owing to their inherent inability to mobilize effectively, for example, children and future generations. To account for the pressures that have been exerted on behalf of these groups it is necessary to see how individuals come to include other individuals within their own interests. (Compare the way in which one's future selves, which can exert no pressure on their own behalf, come to be taken into account by one's present self in virtue of one's identification with them.) Unless one takes account of such processes of incorporation and identification, morality (or even prudence) will appear quite mysterious, but I will have little to say about them here. For some preliminary remarks, see Section IX of my "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13 (1984), pp. 134–171.

social groups that are capable of some degree of mobilization. But under other circumstances, even in the long run, one could expect the opposite. New World plantation slavery, surely one of the most brutally exclusionary social arrangements ever to have existed, emerged late in world history and lasted for hundreds of years. Other brutally exclusionary social arrangements of ancient or recent vintage persist yet.

One need not, therefore, embrace a theory of moral progress in order to see that the feedback mechanism just described can give an explanatory role to the notion of social rationality. Among the most puzzling, yet most common, objections to moral realism is that there has not been uniform historical progress toward worldwide consensus on moral norms. But it has not to my knowledge been advanced as an argument against *scientific* realism that, for example, some contemporary cultures and subcultures do not accept, and do not seem to be moving in the direction of accepting, the scientific world view. Surely realists are in both cases entitled to say that only certain practices in certain circumstances will tend to produce theories more congruent with reality, especially when the subject matter is so complex and so far removed from anything like direct inspection. They need not subscribe to the quaint idea that "the truth will out" come what may. The extended theory of individual rationality, for example, leads us to expect that in societies where there are large conflicts of interest people will develop large normative disagreements, and that, when (as they usually do) these large conflicts of interest parallel large differences in power, the dominant normative views are unlikely to embody social rationality. What is at issue here, and in criterial explanations generally, is the explanation of certain patterns among others, not necessarily the existence of a single overall trend. We may, however, point to the existence of the feedback mechanisms described here as grounds for belief that we can make qualified use of historical experience as something like experimental evidence about what kinds of practices in what ranges of circumstances might better satisfy a criterion of social rationality. That is, we may assign this mechanism a role in a qualified process of moral learning.

The mechanisms of learning about individual rationality, weak or extended, involved similar qualifications. For although we expect that, under favorable circumstances, individuals may become

better at acting in an instrumentally rational fashion as their experience grows, we are also painfully aware that there are powerful mechanisms promoting the opposite result. We certainly do not think that an individual must display exceptionless rationality, or even show ever-increasing rationality over his lifetime, in order to apply reason-giving explanations to many of his actions. Nor do we think that the inevitable persistence of areas of irrationality in individuals is grounds for denying that they can, through experience, acquire areas of greater rationality.

The comparison with individual rationality should not, however, be overdrawn. First, while the inclusion-generating mechanisms for social rationality operate through the behavior of individuals, interpersonal dynamics enter ineliminably in such a way that the criteria selected for are not reducible to those of disaggregated individual rationality. Both social and biological evolution involve selection mechanisms that favor behaviors satisfying criteria of relative optimality that are collective (as in prisoner's dilemma cases) or genotypic (which may also be collective, as in kin selection) as well as individual or phenotypic. Were this not so, it is hardly possible that moral norms could ever have emerged or come to have the hold upon us they do.

Second, there are rather extreme differences of degree between the individual and the social cases. Most strikingly, the mechanisms whereby individual wants and behaviors are brought into some congruence with individual interests and reasons operate in more direct and reliable ways than comparable mechanisms nudging social practices or norms in the direction of what is socially rational. Not only are the information demands less formidable in the individual case—that is the least of it, one might say—but the ways in which feedback is achieved are more likely in the individual case to serve as a prod for change and less likely to be distorted by social asymmetries.

Nonetheless, we do have the skeleton of an explanatory theory that uses the notion of what is more or less rational from a social point of view and that parallels in an obvious way uses of assessments of rationality from the agent's point of view in explanations of individual beliefs and behaviors. Like the individual theory, it suggests prediction- and counterfactual-supporting generalizations of the following kind: over time, and in some circumstances more

than others, we should expect pressure to be exerted on behalf of practices that more adequately satisfy a criterion of rationality.

Well, if this is a potentially predictive and explanatory theory, how good is it? That is a very large question, one beyond my competence to answer. But let me note briefly three patterns in the evolution of moral norms that seem to me to bear out the predictions of this theory, subject to the sorts of qualifications that the existence of imperfections and competing mechanisms would lead one to expect. I do so with trepidation, however, for although the patterns I will discuss are gross historical trends, it is not essential to the theory that history show such trends, and it certainly is not part of the theory to endorse a set of practices or norms merely because it is a result of them.

*Generality.* It is a commonplace of anthropology that tribal peoples often have only one word to name both their tribe and “the people” or “humanity.” Those beyond the tribe are not deemed full-fledged people, and the sorts of obligations one has toward people do not apply fully with regard to outsiders. Over the span of history, through processes that have involved numerous reversals, people have accumulated into larger social units—from the familial band to the tribe to the “people” to the nation-state—and the scope of moral categories has enlarged to follow these expanding boundaries. Needless to say, this has not been a matter of the contagious spread of enlightenment. Expanding social entities frequently subjugate those incorporated within their new boundaries, and the means by which those thus oppressed have secured greater recognition of their interests have been highly conflictual, and remain—perhaps, will always remain—incomplete. Nonetheless, contemporary moral theory, and to a surprising degree contemporary moral discourse, have come to reject any limitation short of the species.<sup>36</sup>

*Humanization.* Moral principles have been assigned various origins and natures: as commandments of supernatural origin,

---

<sup>36</sup>Here and elsewhere, I mean by ‘contemporary moral theory’ to refer to dominant views in the academies, and by ‘contemporary moral discourse’ to refer to widespread practices of public moral argumentation, in those societies that have achieved the highest levels of development of empirical science generally. Again, the moral realist, like the scientific realist, is not committed to worldwide consensus.

grounded in the will or character of a deity, to be interpreted by a priesthood; as formalistic demands of a caste-based code of honor; as cosmic principles of order; as dictates of reason or conscience that make no appeal to human inclinations or well-being; and so on. While vestiges of these views survive in contemporary moral theory, it is typical of almost the entire range of such theory, and of much of contemporary moral discourse, to make some sort of intrinsic connection between normative principles and effects on human interests. Indeed, the very emergence of morality as a distinctive subject matter apart from religion is an instance of this pattern.

*Patterns of variation.* In addition to seeing patterns that reflect some pressure toward the approximation of social rationality, we should expect to see greater approximation in those areas of normative regulation where the mechanisms postulated here work best, for example, in areas where almost everyone has importantly similar or mutually satisfiable interests, where almost everyone has some substantial potential to infringe upon the interests of others, where the advantages of certain forms of constraint or cooperation are highly salient even in the dynamics of small groups, and where individuals can significantly influence the likelihood of norm-following behavior on the part of others by themselves following norms. The clearest examples have to do with prohibitions of aggression and theft, and of the violation of promises.<sup>37</sup> By contrast, moral questions that concern matters where there are no solutions compatible with protecting the most basic interests of all, where there exist very large asymmetries in the capacity to infringe upon interests, where the gains or losses from particular forms of cooperation or constraint are difficult to perceive, and where individual compliance will little affect general compliance, are less likely to achieve early or stable approximation to social rationality. Clear examples here have to do with such matters as social hierarchy—for example, the permissibility of slavery, of authoritarian government, of caste or gender inequalities—and social responsibility—

---

<sup>37</sup>However, such prohibitions historically have shown limitations of scope that are no longer recognized as valid. The trend against such limitations is an instance of the first sort of pattern, toward increased generality.

for example, what is the nature of our individual or collective obligation to promote the well-being of unrelated others?

Given a suitable characterization of the conditions that prevailed during the processes of normative evolution described by these patterns, the present theory claims not only that these changes could have been expected, but that an essential part of the explanation of their occurrence is a mechanism whereby individuals whose interests are denied are led to form common values and make common cause along lines of shared interests, thereby placing pressure on social practices to approximate more closely to social rationality.

These descriptions and explanations of certain prominent features of the evolution of moral norms will no doubt strike some as naive at best, plainly—perhaps even dangerously—false at worst. I thoroughly understand this. I have given impossibly sketchy, one-sided, simple-minded accounts of a very complex reality.<sup>38</sup> I can only hope that these accounts will seem as believable as one could expect sketchy, one-sided, simple-minded accounts to be, and that this will make the story I have tried to tell about mechanisms and explanation more plausible.

Needless to say, the upshot is not a complacent functionalism or an overall endorsement of current moral practice or norms. Instead, the account of morality sketched here emphasizes conflict rather than equilibrium, and provides means for criticizing certain contemporary moral practices and intuitions by asking about their historical genesis. For example, if we come to think that the expla-

---

<sup>38</sup>Moreover, the accounts are highly general in character, operating at a level of description incapable of discriminating between hypotheses based upon the particular account of moral rightness proposed here and others rather close to it. (Roughly, those characterizing moral rightness in terms of instrumental rationality relative to the non-moral good of those affected, but differing on details regarding instrumental rationality—for example, is it straightforwardly maximizing or partly distributive?—or regarding non-moral goodness—for example, is it reducible to pleasure? For a discussion of not-very-close competitors, see Section VI.) If the method I have employed is to be used to make choices from among close competitors, the empirical analysis must be much more fine-grained. Similar remarks apply to the weak and extended theories of individual rationality appealed to above.

nation of a common moral intuition assigns no significant role to mechanisms that could be expected to exert pressure toward socially rational outcomes, then this is grounds for questioning the intuition, however firmly we may hold it. In the spirit of a naturalized moral epistemology, we may ask whether the explanation of why we make certain moral judgments is an example of a reliable process for discovering moral facts.

## V. LIMITATIONS

Thus far I have spoken of what is morally best as a matter of what is instrumentally rational from a social point of view. But I have also characterized a genuinely moral point of view as one impartial with respect to the interests of all potentially affected, and that is not a socially-bounded notion. In fact, I have claimed that a trend away from social specificity is among the patterns visible in the evolution of moral norms. Part of the explanation of this pattern—and part, therefore, of the explanatory role of degrees of impartial rationality—is that the mechanisms appealed to above are not socially-bounded, either. Societies, and individuals on opposite sides of social boundaries, constrain one another in various ways, much as groups and individuals constrain one another within societies: they can threaten aggression, mobilize resistance to external control, withhold cooperation, and obstruct one another's plans; and they are prone to resort to such constraining activities when their interests are denied or at risk. As with intrasocial morality, so in intersocial morality, the best-established and most nearly impartially rational elements are those where the mechanisms we have discussed work most reliably: prohibitions on aggression are stronger and more widely accepted than principles of equity or redistribution. Of course, many factors make intersocietal dynamics unlike intrasocietal ones. . . . But the reader will for once be spared more armchair social science. Still, what results is a form of moral realism that is essentially tied to a limited point of view, an impartial yet human one. Is this too limited for genuine moral realism?

A teacher of mine once remarked that the question of moral realism seemed to him to be the question whether the universe cares what we do. Since we have long since given up believing that

the cosmos pays us any mind, he thought we should long since have given up moral realism. I can only agree that if this were what moral realism involved, it should—with relief rather than sorrow—be let go. However, the account offered here gives us a way of understanding how moral values or imperatives might be objective without being cosmic. They need be grounded in nothing more transcendental than facts about man and his environment, facts about what sorts of things matter to us, and how the ways we live affect these things.

Yet the present account is limited in another way, which may be of greater concern from the standpoint of contemporary moral theory: it does not yield moral imperatives that are categorical in the sense of providing a reason for action to all rational agents regardless of their contingent desires. Although troubling, this limitation is not tantamount to relativism, since on the present account rational motivation is not a precondition of moral obligation. For example, it could truthfully be said that I ought to be more generous even though greater generosity would not help me to promote my existing ends, or even to satisfy my objective interests. This could be so because what it would be morally right for me to do depends upon what is rational from a point of view that includes, but is not exhausted by, my own.

In a similar way, it could be said that I logically ought not to believe both a proposition  $p$  and a proposition that implies not- $p$ . However, it may not be the case that every rational agent will have an instrumental reason to purge all logical contradictions from his thought. It would require vast amounts of cogitation for anyone to test all of his existing beliefs for consistency, and to insure that every newly acquired belief preserves it. Suppose someone to be so fortunate that the only contradictions among his beliefs lie deep in the much-sedimented swamp of factual trivia. Perhaps his memories of two past acquaintances have become confused in such a way that somewhere in the muck there are separate beliefs which, taken together, attribute to one individual logically incompatible properties. Until such a contradiction rears its head in practice, he may have no more reason to lay down his present concerns and wade in after it than he has to leave his home in suburban New Jersey to hunt alligators in the Okefenokee on the off chance that he might one day find himself stranded and unarmed in the backwaters of

southeast Georgia.<sup>39</sup> What an individual rationally ought to do thus may differ from what logic requires of him. Still, we may say that logical evaluation is not subjective or arbitrary, and that good grounds of a perfectly general kind are available for being logical, namely, that logical contradictions are necessarily false and logical inferences are truth-preserving. Since in public discourse and private reflection we are often concerned with whether our thinking is warranted in a sense that is more intimately connected with its truth-conduciveness than with its instrumentality to our peculiar personal goals, it therefore is far from arbitrary that we attach so much importance to logic as a standard of criticism and self-criticism.

By parallel, if we adopt the account of moral rightness proposed above we may say that moral evaluation is not subjective or arbitrary, and that good, general grounds are available for following moral 'ought's', namely, that moral conduct is rational from an impartial point of view. Since in public discourse and private reflection we are often concerned with whether our conduct is justifiable from a general rather than merely personal standpoint, it therefore is far from arbitrary that we attach so much importance to morality as a standard of criticism and self-criticism.

The existence of such phenomena as religion and ideology is evidence for the pervasiveness and seriousness of our concern for impartial justification. Throughout history individuals have sacrificed their interests, even their lives, to meet the demands of religions or ideologies that were compelling for them in part because they purported to express a universal—the universal—justificatory standpoint. La Rochefoucauld wrote that hypocrisy is the tribute vice pays to virtue,<sup>40</sup> but 'hypocrisy' suggests cynicism. We might

---

<sup>39</sup>It is of no importance whether we say that he has *no* reason to do this or simply a vanishingly small one. I suppose we could say that a person has a vanishingly small reason to do anything—even to expend enormous effort to purge minor contradictions from his beliefs or to purge alligators from distant swamps—that might *conceivably* turn out to be to his benefit. But then we would have no trouble guaranteeing the existence of vanishingly small reasons for moral conduct. This would allow naturalized moral rightness to satisfy a Humean thesis of practicality after all, but in a way that would rob the thesis of its interest.

<sup>40</sup>François (duc de) la Rochefoucauld, *Reflexions, ou sentences et maximes morales suivi de reflexions diverses*, ed., Jean Lafond (Paris: Gallimard, 1976),

better say that ideology is the respect partisans show to impartiality. Morality, then, is not ideology made sincere and general—ideology is intrinsically given to heart-felt generalization. Morality is ideology that has faced the facts.

I suspect the idea that moral evaluations must have categorical force for rational agents owes some of its support to a fear that were this to be denied, the authority of morality would be lost. That would be so if one held onto the claim that moral imperatives cannot exist for someone who would not have a reason to obey them, for then an individual could escape moral duties by the simple expedient of having knavish desires. But if we give up this claim about the applicability of moral judgment, then variations in personal desires cannot license exemption from moral obligation.<sup>41</sup>

Thus, while it certainly is a limitation of the argument made here that it does not yield a conception of moral imperatives as categorical, that may be a limitation we can live with and still accord morality the scope and dignity it traditionally has enjoyed. Moreover, it may be a limitation we must live with. For how many among us can convince ourselves that reason is other than hypothetical? Need it also be asked: How many of us would find our sense of the significance of morality or the importance of moral conduct enhanced by a demonstration that even a person with the most thoroughly repugnant ends would find that moral conduct advanced them?

---

p. 79. La Rochefoucauld apparently borrowed the phrase from the cleric Du Moulin. I am grateful to a remark of Barrington Moore, Jr. for reminding me of it. See his *Injustice*, p. 508.

<sup>41</sup>Contrast Harman's relativism about 'ought' in *The Nature of Morality*. Harman adopts the first of the two courses just mentioned, preserving the connection between an individual's moral obligations and what he has (instrumental) reason to do. He defends his approach in part by arguing that, if we suppose that Hitler was engaged in rational pursuit of his ends, an "internal" judgment like 'Hitler (morally) ought not to have killed six million Jews' would be "weak" and "odd" compared to an "external" judgment like 'Hitler was evil' (see pp. 107 ff). I would have thought the opposite, namely, that it is too "weak" and "odd" to give an account of morality such that Hitler can be judged to be consummately evil (which Harman claims, without explanation, his brand of relativism *can* do) but in which 'Hitler (morally) ought not to have acted as he did' is false.

One implication of what has been said is that if we want morality to be taken seriously and to have an important place in people's lives—and not merely as the result of illusion or the threat of repression—we should be vitally concerned with the ways in which social arrangements produce conflicts of interest and asymmetries of power that affect the nature and size of the gap between what is individually and socially rational. Rather than attempt to portray morality as something that it cannot be, as “rationally compelling no matter what one's ends,” we should ask how we might change the ways we live so that moral conduct would more regularly be rational given the ends we actually will have.

## VI. SUMMARY AND CONCLUSION

I have outlined a form of moral realism, and given some indication of how it might be defended against certain objections. Neither a full characterization of this view, nor full answers to the many objections it faces, can be given within the present essay. Perhaps then I should stop trying to say just a bit more, and close by indicating roughly what I have, and have not, attempted to show.

I have proposed what are in effect reforming naturalistic definitions of non-moral goodness and moral rightness. It is possible to respond: “Yes, I can see that such-and-such an end is an objective interest of the agent in your sense, or that such-and-such a practice is rational from an impartial point of view, but can't I still ask whether the end is good for him or the practice right?” Such “open questions” cannot by their nature be closed, since definitions are not subject to proof or disproof. But open questions may be more or less disturbing, for although definitional proposals cannot be demonstrated, they can fare better or worse at meeting various desiderata.

I have assumed throughout that the drawing up of definitions is part of theory-construction, and so is to be assessed by asking (1) whether the analyses given satisfy appropriate constraints of intelligibility and function, and (2) whether the terms as analyzed contribute to the formulation and testing of worthwhile theories. How do my proposals fit with these criteria?

(1) Beyond constraints of intelligibility, such as clarity and non-

circularity, specifically naturalistic definitions of evaluative terms should satisfy two further analytic constraints arising from their intended function. (a) They should insofar as possible capture the normative force of these terms by providing analyses that permit these terms to play their central evaluative roles. In the present setting, this involves showing that although the definitions proposed may not fit with all of our linguistic or moral intuitions, they nonetheless express recognizable notions of goodness and rightness. Further, it involves showing that the definitions permit plausible connections to be drawn between, on the one hand, what is good or right and, on the other, what characteristically would motivate individuals who are prepared to submit themselves to relevant sorts of scrutiny. (b) The naturalistic definitions should permit the evaluative concepts to participate in their own right in genuinely empirical theories. Part of this consists in showing that we have appropriate epistemic access to these concepts. Part, too, (and a related part) consists in showing that generalizations employing these concepts, among others, can figure in potentially explanatory accounts. I have tried to offer reasonably clear definitions and to show in a preliminary way how they might meet constraints (a) and (b).

(2) However, a good deal more must be done, for it remains to show that the empirical theories constructed with the help of these definitions are reasonably good theories, that is, theories for which we have substantial evidence and which provide plausible explanations. I have tried in the most preliminary way imaginable to suggest this. If I have been wholly unpersuasive on empirical matters, then I can expect that the definitions I have offered will be equally unpersuasive.

It is an attraction for me of naturalism in ethics and epistemology alike that it thus is constrained in several significant dimensions at once. One has such ample opportunities to be shown wrong or found unconvincing if one's account must be responsive to empirical demands as well as normative intuitions. Theorizing in general is more productive when suitably constrained; in ethics especially, constraints are needed if we are to have a clearer idea of how we might make progress toward the resolution of theoretical disputes. Of course, not just any constraints will do. A proposed set of constraints must present itself as both appropriate and useful. Let

me say something about (1) the utility of the constraints adopted here, and then a final word about (2) their appropriateness.

(1) Consider three classes of competitors to the substantive moral theory endorsed above, and notice how criticisms of them *naturally* intertwine concerns about normative justification and empirical explanation. *Kantian* conceptions of morality are widely viewed as having captured certain intuitively compelling normative characteristics of such notions as rationality and moral rightness, but it seems they have done so partly at the expense of affording a plausible way of integrating these notions into an empirical account of our reasons and motives in action. Moreover, this descriptive difficulty finds direct expression on the normative side. Not only must any normative 'ought' be within the scope of an empirical 'can', but a normatively compelling 'ought' must—as recent criticisms of Kantianism have stressed—reach to the real springs of human action and concern. *Intuitionist* moral theories also enjoyed some success in capturing normative features of morality, but they have largely been abandoned for want of a credible account of the nature or operation of a faculty of moral intuition. It is too easy for us to give a non-justifying psychological explanation of the existence in certain English gentlemen of something which they identified upon introspection as a faculty of moral insight, an explanation that ties this purported faculty more closely to the rigidity of prevailing social conventions than to anything that looks as if it could be a source of universal truth. *Social choice theories* that take occurrent subjective interests or revealed preferences as given fit more readily than Kantian or intuitionist theories with empirical accounts of behavior, and, unlike them, have found a place in contemporary social science. But they suffer well-known limitations as normative theories, some of which turn out to be bound up with their limitations as explanatory theories: they lack an account of the origin or evolution of preferences, and partly for that reason are unable to capture the ways in which we evaluate purportedly rational or moral conduct by criticizing ends as well as means.

(2) However, the issues at stake when we evaluate competing approaches to morality involve not only this sort of assessment of largish theories, but also questions about which criteria of assessment appropriately apply to definitions and theories in ethics, and about whether definitional systematization and largish theorizing

are even appropriate for ethics. I am drawn to the view that the development of theory in ethics is not an artificial contrivance of philosophers but an organic result of the personal and social uses of moral evaluation: time and again individuals and groups have faced difficult questions to which common sense gave conflicting or otherwise unsatisfactory answers, and so they have pressed their questions further and pursued their inquiry more systematically. The felt need for theory in ethics thus parallels the felt need for theory in natural or social science.<sup>42</sup> It does not follow from this alone that ethical theorizing must run parallel to or be integrable with theorizing in the natural and social sciences. Ethics might be deeply different. Although initially plausible and ultimately irrefutable, the view that ethics stands thus apart is one that in the end I reject. We are natural and social creatures, and I know of nowhere else to look for ethics than in this rich conjunction of facts. I have tried to suggest that we might indeed find it there.<sup>43</sup>

*The University of Michigan, Ann Arbor*

---

<sup>42</sup>This felt need is also reflected in the codification of laws, and in the development of legal theories. However contrived the law may at times seem, surely the general social conditions and needs that have driven its development are real enough. Indeed, the elaborate artifice of law and its language is in part an indication of how pressing the need to go beyond pretheoretic common sense has been.

<sup>43</sup>I am indebted to a great many people, including Peter Achinstein, Robert Audi, Annette Baier, Michael Bratman, Stephen Darwall, Allan Gibbard, Thomas Nagel, Samuel Scheffler, Rebecca Scott, Nicholas Sturgeon, Nicholas White, and the editors at *The Philosophical Review*, who have kindly provided comments on previous drafts or presentations of this paper.