Facts, Values, and Norms

ESSAYS TOWARD A MORALITY OF CONSEQUENCE

PETER RAILTON

University of Michigan

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2003

Printed in the United States of America

Typeface Bembo 10.5/13 pt. *System* LaTeX 2ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Railton, Peter Albert.
Facts, Values, and Norms : essays toward a morality
of consequence / Peter Railton.
p. cm. – (Cambridge studies in philosophy)
Includes bibliographical references and index.
ISBN 0-521-41697-3 – ISBN 0-521-42693-6 (pbk.)
1. Ethics. I. Title. II. Series.
BJ1012 .R33 2003
170–dc21 2002066522

ISBN 0 521 41697 3 hardback
ISBN 0 521 42693 6 paperback

Contents

1

Moral Realism

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.[1]

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 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.

1. 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 that holds that moral judgments can bear truth values in a fundamentally nonepistemic 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."[2]

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 countervailing, 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, 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.[3] 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.[4] 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.[5] 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 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.[6] 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 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.[7]

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."

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.[8]

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.[9]

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 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.[10] 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 grounds. 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*;[11] 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 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?[12] I will try to sketch one, filling in first a realist account of non-moral value – the notion of something being desirable for someone, or good for him.[13]

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 that, 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.[14] Give to an actual individual *A* unqualified cognitive and imaginative powers, and full factual and nomological information about his physical and psychological 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 nonidealized self *A* to want – or, more generally, to seek – were he to find himself in the actual condition and circumstances of *A*.[15] 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 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 Lonnie has a certain *objective interest*.[16]

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.[17]

Let us now say that X is *non-morally good for A* if and only if *X* would satisfy an objective interest of *A*.[18] 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*."[19] 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.[20]

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 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 phenomėna 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 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.[21] 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 self-conscious and un-self-conscious 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 drinks the milk. Soon afterward, 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 of 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.[22]

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 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 nonrelational entities as stones are: they do not depend for their existence or nature merely upon our conception of them.[23]

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 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.[24]

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, 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 × 8s at least, not 2 × 6s." 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.[25] 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.[26] 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 Ramsoe 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 Ramsoe 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.

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.[27] 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 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.[28] Instead, the

connection between the normative and 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 *nonindexical* 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 theorists 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.[29]