7

R. M. Hare, Peter Singer, and the Distinction between Empirical and Normative Adequacy

> I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important. This sense, as Mackintosh remarks, "has a rightful supremacy over every other principle of human action"; it is summed up in that short but imperious word *ought,* so full of high significance. It is the most noble of all the attributes of man, leading him without a moment's hesitation to risk his life for that of a fellow-creature; or, after due deliberation, impelled simply by the deep feeling of right or duty, to sacrifice it in some great cause. Immanuel Kant exclaims, "Duty! Wondrous thought, that workest neither by fond insinuation, flattery, nor by any threat, but merely by holding up thy naked law in the soul, and so extorting for thyself always reverence, if not always obedience; before whom all appetites are dumb, however secretly they rebel; whence thy original?"
>
> – Charles Darwin, *The Descent of Man*

In Part One, I introduced Rawls' linguistic analogy and drew on the work of both Rawls and Chomsky to formulate a new analytic framework for the theory of moral cognition, modeled on aspects of Universal Grammar. In Part Two, I sought to clarify the empirical significance of the linguistic analogy by formulating, and stating a provisional solution to, the problem of descriptive adequacy with respect to a class of considered moral judgments, including the original trolley problems devised by Foot and Thomson. In Part Three, I turn my attention to some influential criticisms of Rawls' linguistic analogy and argue that they lack force against both Rawls' conception of moral theory and the research program outlined in Parts One and Two.

7.1 HARE'S AND SINGER'S CRITICISMS OF RAWLS' LINGUISTIC ANALOGY

A useful place to begin is by examining R. M. Hare's objections to Rawls' linguistic analogy. Hare begins his 1973 essay, "Rawls' Theory of Justice," by

distinguishing four topics that any moral philosopher is likely to discuss: (i) "philosophical methodology," (ii) "ethical analysis," (iii) "moral methodology," and (iv) "normative moral questions." According to Hare, philosophical methodology concerns "what philosophy is supposed to be doing and how it does it." Ethical analysis relates to "the meaning of moral words or the nature and logical properties of the moral concepts." Moral methodology addresses "how moral thinking ought to proceed, or how moral arguments or reasonings have to be conducted if they are to be cogent." Finally, normative moral questions are questions such as "what we ought or ought not to do, what is just or unjust, and so on" (Hare 1973: 81).

Hare's central claim is that "through misconceptions about (i), Rawls has not paid enough attention to (ii), and that therefore he has lacked the equipment necessary to handle (iii) effectively; so that what he says about (iv), however popular it may prove, is unsupported by any firm arguments" (1973: 81–82).[1] Hare organizes his review around these four topics. His explicit remarks on Rawls' linguistic analogy are contained in his discussion of (ii). I will examine Hare's specific objection to that analogy in a moment. But first some general remarks about Hare's review are in order.

In Part One, we saw that in Section 9 of *A Theory of Justice* Rawls organizes his conception of moral theory around the assumption that each normal human being possesses a moral sense or sense of justice, the existence of which he defends by the argument for moral grammar. Rawls takes describing this sense of justice to be the provisional aim of moral philosophy. He distinguishes different forms that such a description might take, and he suggests that a descriptively adequate moral grammar should take the form of a fully explicit scientific theory – a generative grammar, in approximately Chomsky's sense.

Rawls compares the problem of describing the sense of justice with the linguist's problem of describing linguistic competence. He emphasizes that both inquiries are difficult and require theoretical constructions that go beyond the familiar precepts of common sense. He distinguishes between a description of the sense of justice and a description of how the sense of justice is acquired – in other words, between descriptive adequacy and explanatory adequacy, in the linguist's technical sense. Finally, Rawls draws a distinction between moral performance and moral competence, taking his primary object of inquiry to be the moral competence of a single, idealized individual.

Hare's review is striking, in the first instance, for the number of these features of Rawls' stated conception of moral theory that he ignores or fails to recognize. Hare does not discuss Rawls' assumption that each person possesses a moral sense or sense of justice, nor Rawls' argument for moral grammar; on the

[1] This passage is a direct quotation, except that I have substituted "(i)–(iv)" for Hare's original "(1)–(4)."

contrary, the terms "moral sense" and "sense of justice" do not appear in his review. Hare does not appear to recognize the distinctions Rawls draws among different levels of empirical adequacy, or between operative and express principles. He does not acknowledge Rawls' use of a competence–performance distinction, nor how Rawls uses the linguistic analogy to clarify both the theory-dependence of that distinction and the importance of idealization.[2] Nor, finally, does Hare discuss Rawls' historical claim that Rawls' conception of moral theory is how the subject matter of moral philosophy was traditionally conceived by most classical British writers through Sidgwick.

Hare's most direct criticism of Rawls' linguistic analogy is contained in the following passage:

> There are significant passages where Rawls compares moral philosophy with mathematics and linguistics. The analogy with these sciences is vitiated by the fact that they do not yield substantial conclusions, as moral philosophy is supposed, on Rawls' view, to do, and in some sense clearly should. It is quite all right to test a linguistic theory (a grammar) against what people actually say when they are speaking carefully; people's *linguistic* 'intuitions' are indeed, in the end, authoritative for what is correct in their language. The kind of interplay between theory and data that occurs in all sciences can occur here, and it is perfectly proper for the data to be the utterances of native speakers. But the only 'moral' theories that can be checked against people's actual moral judgments are anthropological theories about what, in general, people *think* one ought to do, not moral principles about what one ought to do. That these latter can be so checked is not, indeed, what Rawls is suggesting in this passage; but do not the whole drift of his argument, and the passage quoted above, suggest it? (Hare 1973: 85–86, emphasis original, internal citations omitted)

According to Hare, Rawls' linguistic analogy is "vitiated" by the fact that linguistics does not yield the kind of "substantial" conclusions that it is reasonable to expect of moral philosophy. What Hare means by this remark is not entirely clear. Given what he says elsewhere in the passage and in his review

[2] For example, Hare writes:

> Rawls states quite explicitly how he thinks moral philosophy should be done: "There is a definite if limited class of facts against which conjectured principles can be checked, namely our considered judgments in reflective equilibrium" (p. 51). It is clear from the succeeding passage that Rawls does not conceive of moral philosophy as depending primarily on the analysis of concepts in order to establish their logical properties. Rather, he thinks of a theory of justice as analogous to a theory in empirical science. It has to square with what he calls the 'facts', just like, for example, physiological theories. But what are these facts? They are what people will *say* when they have been thinking carefully. (Hare 1973: 82, emphasis original)

The fact that Hare takes these observations to be cogent criticisms of Rawls suggests that he misunderstands Rawls' use of the competence–performance distinction and, in particular, the distinction between operative and express principles. In effect, what Rawls takes to be the facts against which conjectured principles may be checked are not what people *say* when they have been thinking carefully, but what they *cognize* when they have been thinking carefully. The difference between these two alternatives is crucial.

as a whole, however, Hare's point appears to be that a research program in moral theory modeled on linguistics as Rawls advocates would fail to be prescriptive or normative ("authoritative for what is correct"), as both Hare and, Hare suggests, Rawls think a moral theory clearly should be.

If this is an accurate interpretation of Hare's objection, then it seems that Hare has misunderstood the main point of Rawls' linguistic analogy. Moreover, it suggests that Hare has a surprisingly impoverished conception of the proper scope and limits of moral philosophy. The comparisons that Rawls draws in Section 9 of *A Theory of Justice* are not comparisons between linguistics and an undifferentiated set of questions called "moral philosophy," where the latter is interpreted to refer exclusively to normative ethics, as Hare understands this topic, and to exclude scientific questions about the nature and origin of commonsense moral knowledge. Rather, as we have seen, Rawls' conception of moral theory is more complex. In Section 9 and elsewhere, Rawls distinguishes at least five problems that moral philosophy seeks to solve, which I have labeled the problems of descriptive adequacy, explanatory adequacy, behavioral adequacy, normative adequacy, and metaethical adequacy, respectively (see Table 2.1). The analogy with linguistics is centered primarily on the comparison between the first two problems, descriptive and explanatory adequacy, and their counterparts in linguistics. Hare's criticism of the linguistic analogy is itself vitiated by his failure to grasp these points.

In my view, confusion over these issues has been widespread in the philosophical literature. Some version or other of Hare's criticism that Rawls' conception of moral theory is too empirical or insufficiently normative can be found in many leading philosophers' responses to Rawls. Peter Singer, for instance, objects to Rawls' linguistic analogy on the grounds that "the historical task of moral philosophy" is

> to develop theories that serve as guides to conduct. So long as there are grounds for hoping that discussion, argument, and the careful consideration of moral theories can help us to decide how to act, the importance of this historical task cannot be denied. Sidgwick, obviously, was engaged on this task. Can the same be said for Rawls? In a sense, yes. Rawls clearly intends his theory to be a guide to conduct; yet, at the same time, his use of the reflective equilibrium idea means that he is on the verge of slipping off into an altogether different activity, that of systematizing the considered judgments of some unspecified moral consensus. This latter task, while it may be of some interest, is, like the linguistic and scientific investigations on which it is modeled, a descriptive task from which, without supplementation from other sources, no normative or action-guiding consequences can be derived. We cannot test a normative theory by the extent to which it accords with the moral judgments people ordinarily make. Insofar as Rawls frequently does seem to be testing his own theory in this way, the theory fails to be normative and Rawls cannot be regarded as pursuing the same task as Sidgwick and most other moral philosophers. (Singer 1974: 515)

Similar concerns about the insufficiently normative or authoritative character of Rawls' conception of moral theory can be found in many other influential philosophers, including Dworkin, Daniels, Williams, and Brandt.

I believe this criticism is misguided, for at least the following reasons. First, those authors who have criticized Rawls on this basis appear to be operating with an unduly narrow and impoverished conception of moral philosophy and are simply refusing to join Rawls in recognizing the centrality of the problem of empirical adequacy in the history of their discipline. In *The Descent of Man,* for example, Darwin (1981/1871: 71) observes that the classical philosophical question – "whence thy original?" – that stimulates his own inquiry into the moral sense from the side of natural history "has been discussed by many writers of consummate ability." Darwin then explains this reference by referring to Alexander Bain's list in *Mental and Moral Science* (1868) of "twenty-six British authors who have written on this subject, and whose names are familiar to every reader" (Darwin 1981/1871: 71). Bain's list consists of Hobbes, Cumberland, Cudworth, Clarke, Wollaston, Locke, Butler, Hutcheson, Hume, Price, Smith, Hartley, Reid, Stewart, Brown, Paley, Bentham, Mackintosh, James Mill, Austin, Whewell, Ferrier, Mansel, J. S. Mill, Bailey, and Spencer (Bain 1868: 543–725). There seems little doubt that when Rawls refers to "the conception of the subject adopted by most classical British writers through Sidgwick" (1971: 51), it is this group of writers, with the addition of Sidgwick and perhaps a few other additions or subtractions, that Rawls has primarily in mind; and there is little question that most of these writers considered empirical questions about the nature and origin of moral knowledge, along with related topics concerning the existence, character, and origin of the human moral faculty, moral sense, or conscience, to be central problems – if not *the* central problems – in moral philosophy (see, e.g., Hume 1978/1739–1740, 1983/1751; Price 1948/1758; Smith 1976/1759; see generally Mackintosh 1830; Schneewind 1977, 1990, 1998; Sidgwick 1988/1902, 1981/1907; Stewart 1876/1793).

Moreover, many of the philosophers who have criticized Rawls on the ground that his conception of moral theory is too empirical or insufficiently normative have often underestimated or ignored the complexity of Rawls' approach to the problem of metaethical adequacy, that is, the problem of how moral principles can be justified (Section 2.2). In the first place, they have generally neglected Rawls' plausible contention in *Outline* (and *Grounds*) that there is a natural order of priority between the problems of descriptive and normative adequacy, and that a solution to the former problem constitutes at least a presumptive solution to the latter problem, given the properties of the considered judgments that a descriptively adequate moral theory purports to explicate.[3] In addition, they have also largely neglected Rawls' sensible observation that metaethical questions about meaning and justification may prove easier to answer once solutions to the problems of descriptive and explanatory

[3] Despite an extensive literature search, when I wrote *Rawls' Linguistic Analogy* I could not find a single critic who objected to Rawls' conception of moral theory on these grounds, yet who also discussed Rawls' claim in *Outline* that a solution to the problem of descriptive adequacy is presumptively normative in light of the properties of the judgments it purportedly explains. I put *Grounds* in parentheses in the text to signal that I am not

adequacy are achieved (1971: 50–51). More generally, they have not seriously engaged with the complexity and plausibility of Rawls' concept of reflective equilibrium and its implication, building on the work of Nelson Goodman (1983/1955) to which we return (Section 7.4.2), that the set of considered judgments in reflective equilibrium can, in principle, yield principles that are both descriptively and normatively adequate.

Finally, from a cognitive science perspective, what is perhaps most noteworthy about the writers who have pressed this criticism – what I call the *objection from insufficient normativity* – against Rawls is that they often beg the very empirical questions that Rawls' conception of moral theory, or the research program outlined in Parts One and Two at any rate, is ultimately designed to answer. In particular, these critics often simply take for granted that some version of behaviorism or social learning theory is the correct account of how commonsense moral knowledge is acquired (see, e.g., Brandt 1979, 1982; Hare 1952, 1973, 1981; Singer 1974; Williams 1985; see generally Mikhail 2007b).

Because this last point may be the most important, it seems appropriate to illustrate it with a specific example. In his influential commentary on Rawls' concept of reflective equilibrium, Singer (1974) questions Rawls' fundamental assumption that each person develops a moral sense or sense of justice under normal circumstances, suggesting Rawls is "led to" make this assumption because of a mistaken conception of the proper aims of moral philosophy. Singer (1974: 516) thinks moral philosophy should begin from what he calls "the opposite assumption," namely, that "all the particular moral judgments" human beings intuitively make are likely to result from their upbringing or other features of their environment. Here is how Singer puts the criticism:

> [F]rom the start [Rawls] thinks of moral philosophy in the wrong way. Thus he says: 'Now one may think of moral philosophy at first (and I stress the provisional nature of this view) as the attempt to describe our moral capacity' (p. 46). Even if this is only provisional, it is a misleading starting point. It leads to the assumption that we have a certain moral capacity, and that at least some of the moral judgments we make will, after consideration, remain as fixed points against which theories can be tested. Why should we not rather make the opposite assumption, that all the particular moral judgments we intuitively make are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past? In which case it would be best to forget all about our particular moral judgments, and start again from as near as we can to self-evident moral axioms. (Singer 1974: 516)

making the unreasonable suggestion that Rawls' critics should be faulted for failing to track down his dissertation. *Outline* is another matter, however, since Rawls repeatedly makes clear that the conception of moral theory on which he relies in *A Theory of Justice* follows the general point of view of this essay (see, e.g., 1971: 46, 579).

There are at least three interrelated points to make about this remarkable passage. First, as a description of Rawls, Singer gets things backwards: as we saw in Chapter 3, Rawls' characterization of moral philosophy *follows from* the assumption that human beings possess a moral capacity. Moreover, Rawls *defends* this assumption when he points to observable properties of moral judgment, including its spontaneity, novelty, and unbounded scope, from which the assumption is inferred. Like Chomsky, for whom the argument for linguistic grammar helps to frame the aims of linguistic theory, Rawls suggests "our natural readiness" to make moral judgments, and their "potentially infinite number and variety," point toward regarding "the sense of justice as a mental capacity," a capacity whose structure is "extraordinarily complex" and hence a topic for investigation (Rawls 1971: 46; see generally Section 3.1.1). In questioning Rawls' assumption that human beings possess a moral capacity, Singer simply fails to speak directly to those properties of moral judgment that Rawls actually invokes, however tentatively, in his own defense.

Second, Singer's criticism of Rawls is question-begging. Singer thinks that moral philosophy should begin from certain factual assumptions about the origin of moral judgments, while criticizing Rawls for making different empirical assumptions. But, like any theorist, Rawls is entitled to begin his enterprise from whichever assumptions he thinks are valid, so long as he is prepared to defend them with arguments and evidence. We will return to this topic in Chapter 8.

Third, Singer apparently thinks that the assumption that human beings possess a moral capacity is necessarily incompatible with the "opposite" claim that "all the particular judgments" they intuitively make "are likely to derive" from the various sources he cites, but this view is mistaken. The claim that human beings possess a moral capacity does not imply that their moral intuitions do not also derive from their environment in some manner, as I shall explain momentarily. As with language, the two issues are distinct.

Singer's tacit assumption that the *exclusive* origin of a child's moral knowledge is her surrounding environment, what one might call his *extreme environmentalism,* is a pervasive and influential one. The same basic premise, which is found elsewhere in G. H. Mead's claim that "minds and selves are essentially social products" (1956: 116) and in Emile Durkheim's claim that "society is the only source of morality" (Piaget 1965/1932: 327, citing Durkheim's *L'Education Morale*), informs many existing research programs in moral development, including those approaches inspired by the social constructivism of Piaget and Kohlberg. But the assumption, at least as Singer expresses it, appears badly confused, a point that becomes clear when one considers what a comparable assumption would imply in the case of language. Suppose Singer or someone skeptical of Universal Grammar asserted the following proposition:

> From the start Chomsky thinks of linguistics in the wrong way. Thus he thinks of linguistics as the attempt to describe our linguistic capacity. This is a misleading starting point. It leads to the assumption that we have a certain linguistic capacity,

and that at least some of the linguistic judgments we make will, after consideration, remain as fixed points against which theories can be tested. Why should we not rather make the opposite assumption, which is that all the particular linguistic judgments we intuitively make are likely to derive from the early training we receive from the environment? In which case it would be best to forget all about our particular linguistic intuitions, and start again from as near as we can to self-evident axioms.

There obviously are many problems with an argument like this. Here I simply will concentrate on what appears to be its main fallacy. The fact that a person's linguistic intuitions "derive" in some manner from her environment is obvious and uncontroversial. A child who grows up in England or the United States will learn English; if the same child had grown up in Egypt or Israel, she would speak fluent Arabic or Hebrew. This observation, however, is not usefully contrasted with an allegedly "opposite" assumption that the child's linguistic intuitions are a body of scientific evidence against which conjectured theories of her linguistic capacities can be checked. On the contrary, the two propositions are perfectly compatible. Indeed, it is precisely their compatibility, in effect, that motivates the study of Universal Grammar in the first place, whose main objective is to explain the phenomena of language acquisition within the boundary conditions set by experience and by the facts of linguistic diversity. The same logic presumably holds in the moral domain. The fact that people acquire moral beliefs that in some manner reflect the circumstances of their upbringing appears obvious and uncontroversial. The interesting question, from scientific point of view, concerns the character of the morality acquisition system (Universal Moral Grammar) that maps this early experience into the mature or steady state of each individual's moral competence. In implicitly denying that human beings possess such a system, Singer appears to be adopting an extreme version of epistemological empiricism, which, although popular among many midcentury Anglo-American moral philosophers (see generally Mikhail 2007b), seems both at variance with modern science and empirically untenable.

In the remainder of this chapter, I do not pursue all of the foregoing lines of inquiry. Instead, I focus my attention on the relationship between the problems of empirical and normative adequacy in Rawls' conception of moral theory, limiting the scope of my investigation as I have throughout this book to the early part of Rawls' career (1950–1975). I begin by examining how Rawls approaches this topic in *Grounds* and *Outline.* Then I turn my attention to *A Theory of Justice,* focusing in particular on how the concept of reflective equilibrium appears to function in the conception of moral theory that Rawls elaborates there. Finally, I examine Rawls' response to the objection from insufficient normativity in *Independence* (Rawls 1975). My primary aim is to show that Rawls' approach to the problem of metaethical adequacy is plausible and compelling, and that theorists who are inclined to pursue the research program outlined in Parts One and Two need not avoid doing so on the ground that it is insufficiently normative.