Environmental Ethics

Volume 13, Issue 3, Fall 1991

John N. Martin
Pages 215-234
DOI: 10.5840/enviroethics199113311

Order Theoretic Properties of Holistic Ethical Theories

Order Theoretic Properties of Holistic Ethical Theories

John N. Martin*

Using concepts from abstract algebra and type theory, I analyze the structural presuppositions of any holistic ethical theory. This study is motivated by such recent holistic theories in environmental ethics as Aldo Leopold's land ethic, James E. Lovelock's Gaia hypothesis, Arne Naess' deep ecology, and various aesthetic ethics of the sublime. I also discuss the holistic and type theoretic assumptions of such standard ethical theories as hedonism, natural rights theory, utilitarianism, Rawls' difference principle, and fascism. I argue that although there are several common senses of part-whole in ethical theory, the central sense of holism in ethics is that of a theory that defines its key moral idea as an emergent group property grounded in the relational properties of its individual constituents. Hedonism and Kantianism do not count as holistic in this sense. Natural rights theory does in a degenerate way. Utilitarianism and various environmental ethics are paradigm examples. I point out as a general structural weakness of environmental holistic theories that their first-order grounding in nonmoral vocabulary seems to preclude an explanation of many moral intuitions about human ethics.

I. INTRODUCTION

Aldo Leopold's land ethic, the Gaia hypothesis, and deep ecology are examples of approaches now common in environmental ethics that embrace holism as a key principle.[1] In this paper I propose to explore the general features of these and more familiar ethical theories that employ holistic concepts. Being essentially a concept of order, holism invites analysis in terms of universal algebra and type theory. By focusing on the rather abstract assumptions theories make about the structure of their subject matter, I sketch a classification scheme for moral theories that I think places traditional theories in some new light and at the same time lays bare some difficulties facing the newer forms of holism.

---

*Department of Philosophy, University of Cincinnati, Cincinnati, OH 45221-0374. Martin is a logician with a special interest in environmental ethics. In addition to articles on logic and environmental ethics, he is the author of *Elements of Formal Semantics* (New York: Academic Press, 1987) and was for many years the book review editor of this journal.

[1] Much has been written on these closely related ideas. For the original statement of the land ethic see Aldo Leopold, *A Sand County Almanac and Sketches Here and There* (New York: Oxford University Press, 1949). For the Gaia hypothesis, see James Lovelock, *Gaia: A New Look at Life on Earth* (New York: Oxford University Press, 1979), and "Gaia: The World as a Living Organism," *New Scientist* 112, no. 1539 (1986): 25–28. The term *deep ecology* comes from Arne Naess, "The Shallow and the Deep, Long-Range Ecology Movements: A Summary," *Inquiry* 16 (1973): 95–100. See also "A Defence of the Deep Ecology Movement," *Environmental Ethics* 6 (1984): 265–70.

Because the understanding I am seeking is what might be called structural or perhaps mathematical in a broad sense of the term, it is different from that sought in many discussions of ethical theory. Rather than refining a particular approach to deal with the subtleties of moral situations, I shall be rather grossly abstracting, simplifying theories so as to reveal their crudest assumptions about structure. Rather than trying to produce a single more or less airtight theory, I am seeking an overview of theory types that provides some understanding of the goals and difficulties typical of each.

II. PRELIMINARIES FROM TYPE THEORY

The background ideas from abstract algebra that I employ derive from rudimentary type theory. The motivation for studying "types" arises from the fact that theories often impose type restrictions on their concepts. It is not uncommon to find a theory explicitly restricting the meaningful application of predicates, constants, and functors to universes of a specified kind. In the various sciences, for example, numerical expression may be required to stand for numbers, and physical vocabulary may be restricted to physical objects of certain sorts. The restriction to types is more interesting when the types themselves exhibit structural properties among themselves. Sometimes, for example, they are organized by an ordering relation intended to capture some notion of type composition. Entities from one type "compose" entities of the next higher type. In addition, the entities on a given level—within a single type—may themselves fall into specified sets or stand in specified relations. The types discussed later exhibit both sorts of structure.

Let us begin with some of the basic model theoretic assumptions I employ about theories.[2] It suffices for my purposes to think of a *theory* as a linguistic entity, as a set of sentences in a prescribed vocabulary. These are the sentences that are intended to be true in the theory. I call them the theory's *laws*. Laws, in this usage, embrace a theory's theorems as well as its axioms. By a *model* for a theory I mean the usual notion of a domain and an assignment of interpretations or extensions on this domain to the descriptive vocabulary of the theory that make the sentences of the theory true according to the usual definition of truth. Constants are assigned individuals in the domain of the model, functors are assigned operations on the domain, and predicates are assigned sets and relations over the domain. The set of *acceptable* or *intended* models for the theory may be further constrained by postulating in the metalanguage a series of conditions on models. In particular, a theory is said to *posit* a *type structure* if there is some set

[2] Although self-contained and informal, the discussion assumes a familiarity with the basic concepts of first-order model theory.

of monadic predicates (which I refer to as the *type predicates* of the theory) and some binary predicate (the theory's *ranking predicate*) such that in all acceptable models for the theory the extensions of type predicates are required to be mutually disjointed and to collectively exhaust the domain. In addition, the ranking predicate is required to have as its extension a relation $<<$ that organizes the types into a linear order. I call the extension of any type predicate P the *type determined* by P. The choice of type and ranking predicates used in the theory and the pre-analytical conceptual content that these expressions are intended to capture varies from theory to theory, but we may understand "$x << y$" as saying in the metalanguage that, in the type structure under discussion, x and y are in neighboring types and x enters into the composition of y in the relevant manner. To say that $<<$ is a linear ordering in this context means that its transitive closure is reflexive, transitive, antisymmetric, and complete. By a *type theory* I mean any theory for which there is a given type structure to which the theory's set of acceptable models is required to conform.

Some of the type theories I discuss exhibit additional structure and require that in every acceptable model the extensions of one or more of the type predicates must have some internal structure. Although this structure may consist of any kind of prescribed classification and relational organization, two sorts of internal structure are important in this discussion. The first involves using a type as the "significance range" of a predicate in the sense that it is part of the intended interpretation of the theory that the predicate meaningfully apply only to entities of that type. This restriction is accomplished by building into the definition of an acceptable model the constraint that the extension of a descriptive term must fall within the extension of a given type predicate. Constraints of this sort allow the vocabulary of the theory to be partitioned into "subject matters," each of which takes its interpretations within a particular type. Indeed, in a language with descriptive terms constrained to types, the type predicates may be viewed as sortals and the language construed as a many-sorted logic.[3] A second type of internal type structure that is of some importance consists of ordering relations within the type. Unlike the composition relation $<<$ that provides a global ordering for entities from distinct "significance ranges," orderings within a type allow for the ordering of entities within a sort. For any type predicate P within a type theory, the type determined by P is *ramified* if there is some binary predicate of the theory (a *ramifying predicate* for P) which has the property such that, in all acceptable models for the theory, the ramifying predicate stands for an

[3] For the purposes here, we need not decide what happens to the truth value of simple sentences that apply a predicate to an entity falling outside its "significance range." In classical logic such sentences normally are treated as false and their negations as true. In non-classical semantics, however, there are various treatments consistent with the discussion here that allow both to be neither true nor false.

ordering relation $\leq$ on the extension of P. To count as an ordering, $\leq$ must exhibit, at a minimum, the properties of reflexivity and transitivity. Again, the choice of predicates for ramifying relations vary from theory to theory and even perhaps from type to type within a theory. Nevertheless, "$x \leq y$" can be understood abstractly as saying in the metalanguage, relative to some type and sense of order posited in the theory for that type, that x and y are in that type and that either x comes earlier than y or is of equal rank to y. In some cases it is also required that there exist a highest element in the ramifying ordering within the extension of a type predicate, and that the vocabulary of the theory contain a constant referring to it. Let the *supremum* of $\leq$ in type T be that element of T which all elements of T bear $\leq$ to it, but which does not itself bear $\leq$ to anything in T other than itself. Only some of the ordering relations discussed below have a supremum in this sense.

With just this amount of theory it is possible to identify several senses in which something may be a "whole" within a type structure. If $x << y$, then the theory posits an intended sense of whole in terms of which y is a whole that contains x. Similarly, if a type is ramified by some relation $\leq$, and if that relation itself is intended to capture the conceptual content of some part-whole relation, then $x \leq y$ means that y is a whole relative to x in that sense. Similarly, any supremum of $\leq$, if there is one, is an ultimate whole.

Because the theories I discuss are in part ethical, I also assume that we can distinguish ethical from other sorts of theories and that we can do so by reference to their vocabularies. I assume that there is a stock of traditional philosophical and scientific terms and that it is the incorporation of these traditional terms (or their formal proxies) into a theory that allows us to determine the subject matter of the theory. We can tell that a theory is ethical because it is formulated at least in part in traditional ethical vocabulary. Similarly, a physical theory is one that employs traditional terminology from physics, and an ecological theory is one that uses terms from ecology. A semantic constraint on any theory, then, is that it employs its terms in a manner guided by traditional usage. *Very roughly*, the intellectual goals of science include defining a vocabulary composed of the descriptive terms from its scientific tradition, singling out a set of sentences as the body of truths of the theory, and formulating conditions for a plausible set of acceptable models for these sentences. One conceptual constraint on successfully accomplishing these goals is that the propositions should be, at least in large measure, conceptually plausible. That is, the laws of the theory must be phrased in a way that is compatible with pre-analytic scientific usage. One descriptive constraint is that the set of acceptable models of the theory should be fair approximations of some part of the actual world and its possible variants. That is, its laws should have as their acceptable models a set which is a fair approximation of the set of natural possibilities for that part of nature under investigation. Though highly abstract, this picture of scientific theories is familiar and is sufficient for my purposes here.

In addition, this picture applies well enough to ethics. In my discussion below, I assume that it is correct to view ethics as a theoretical enterprise designed to codify our reflective moral intuitions. The goals of an ethics, on this view, are similar to those of any other science. They consist of an attempt to define, first, the set of sentences that make up the theory and, second, the set of acceptable models for this set. Ethical theories, especially those dealing with environmental issues, often mix moral concepts with those more appropriately studied in one of the special sciences. Viewed in this way, ethical theories can be understood as developments or extensions of scientific theories. The ethical part consists of those sentences composed at least in part from the moral terms introduced from traditional ethics (the theory's *moral laws*). The non-ethical part is the special theory before it is extended by the inclusion of moral terms. It consists of those sentences formulated solely in the terms of a non-moral vocabulary. The theory's non-ethical fragment can be evaluated as any special science. The combined theory, in addition, has to conform to moral intuition. (Of course, some intuitions themselves might be rejected if the theory is in other ways ideal.) The phrasing of the moral laws have to be conceptually plausible in that it should conform to pre-analytic traditional usage. Moreover, the set of acceptable models have to approximate what we intuitively accept as ethically possible. That is, all acceptable models for the theory have to assign truth values to ethical judgments and to contingent factual statements in a way that matches ethical intuition. It counts against a theory if there is an acceptable model that assigns truth values to sentences sketching background circumstances that also assigned truth values to ethical sentences in a way that conflicts with our moral intuitions about what moral judgments should be made in such circumstances.

Because it is customary to speak of ethical theory as analyzing moral properties, for the purposes of this paper, a theory *posits a property* P if P is a concept from traditional science or ethics and if the theory has a descriptive term that is intended to express P as its conceptual content. If the property is non-measurable, it is usually represented in the theory by a predicate. If it is measurable, it might be represented by a functor interpreted by an operation that assigns a numerical value to objects of the appropriate type.

III. SIMPLE TYPE THEORIES

It is typical of ethical theory to posit a moral universe, by which I mean a set of individuals, actions, or events within which moral expressions are interpreted, and among which moral comparisons are meaningful. From the perspective of type theory, the simplest moral theory is one that attributes the least type structure possible to the moral universe. Such a case is a degenerate one, from the perspective of its algebraic structure, in which the type structure consists of a single unramified type. A single unramified domain of this sort is appropriate for a moral theory that allows the moral terms to range over all entities in the

domain, and which does not posit any relevant moral ordering among them.

A good example of this sort of theory is an unadorned version of hedonism. Let us posit a single moral property of pleasure that applies to actions or events. In this theory an entity is morally good if and only if it gives pleasure. No pleasurable act is intrinsically preferable to any other, and the only structural feature of the theory consists of the binary division of the moral universe into the good and the bad. A stark version of eudaimonism might be similarly sketched.

This theory can be made more complex by introducing degrees or kinds of pleasure. Before doing so, however, there is an ethically relevant structural issue present even in the simplest theory: whether the moral property in question is freely available. Let us define a very weak notion of scarcity and say that, relative to a theory, a property ranging over type T is *scarce* if and only if the extension of the property does not exhaust T in any intended model for the theory. Thus, a scarce property is one that in principle can never be had by all. Some may even count in favor of a theory the fact that it avoids scarcity, that in some sense it allows for the possibility of universal virtue. However, given that some pleasures consist in giving pain, hedonism probably ought to posit a built-in moral scarcity: in a plausible version of the theory, the definition of the theory's acceptable type structure should incorporate a restriction of scarcity for the pleasure predicate. In this regard hedonism and structurally similar versions of eudaimonism contrast with theories that, though structurally simple, intentionally preclude scarcity. Such theories include classical liberalism and Kantianism.

In its barest form liberalism may be understood as positing a universe of actions, some of which are in accordance with natural rights and some of which are not. A classical postulate of the theory, one that is essentially structural, requires that it be possible for everyone to act in a manner so as not to violate anyone's rights. Similarly, a Kantian would probably grant that it should be possible for everyone to act so as not to treat anyone as a means rather than an end. Likewise, a Christian finds no conceptual problem in everyone loving one another.

In "lack of scarcity" we have found an example of the sort of thing I am seeking. It is a theoretical feature, interesting and perhaps desirable, that is abstractly definable in terms of the general structural properties of a theory's type structure. Moreover, it is definable in a type structure of the simplest sort. I do not mean to say that this particular property is particularly important in the overall evaluation of competing theories, but it does suggest that it is worthwhile to hunt for this sort of feature in richer type theoretic contexts.

IV. SINGLE TYPE RAMIFIED THEORIES

Some theories posit essentially a single moral universe within which moral terms are interpreted, but in a way that attributes to the entities within this do-

main some important relational structure. Perhaps the simplest way this is done in ethical theory is by introducing moral inequities. In some accounts, it is important to know not only whether an entity falls under a moral property, but also the moral rank of the entity.

Hedonistic theories that discriminate among pleasures are of this sort. Pleasurable acts are organized into kinds and a ranking is imposed upon them so that an act A from one kind may be inferior to an act B from another kind. For simplicity let us assume that a consistent scheme of this sort is possible and that, for example, an action is pleasant because it pleases Jones regardless of its effect on Smith. In the terminology of type theory, the moral theory posits as its moral universe a single type of action together with a distinguished set of pleasant actions and a ramifying ordering relation $\leq$ on the type as a whole. A is morally preferable to B, if A is pleasant and B is not, or in the case in which both are pleasant, if $B < A$.

An intuitively different but structurally very similar sort of theory is found in environmental ethics. The theories I have in mind are guided by the intuition that it is sometimes wrong to harm animals and plants even when doing so has no affect on humans. Typically the moral universe is extended to include nonhumans as well as humans and to spell out a concept of moral interest that applies to both. Any such account, however, is open to the objection that in cases of conflict humans often seem to count more than natural objects. This problem is sometimes dealt with by positing a ranking reflecting relative moral importance. Roughly, the more human-like an object is, the more its interests count. Such theories may be viewed as positing a single type of event that is structured by a subset and an ordering relation. The subset is the extension of a theoretical predicate intended to capture the property of "advancing something's interests." The relation serves as the denotation of a predicate expressing the relevant property of moral priority. (As in the case of pleasures, I am here "abstracting from" the fact that what is in the interest of one may be against the interest of another.)

Common to both ramified hedonistic and interest theories is the feature that the ramifying relation itself has moral content. It is intended to unpack a concept of moral ranking that enters into the definition of moral preference. There is, however, quite a different sort of ramification that is sometimes posited in environmental theories and which has counterparts in traditional ethics. This is the ordering of part to whole. Let us begin with some simple examples in which the part-whole relation is ethically irrelevant.

One sense of part-whole is set membership. In traditional ethics—or perhaps it is better to think of it as legal theory—there is a precedent for introducing set-like entities to the moral universe on a footing equal to individual humans, and for ranking these entities by the relation of set membership. It is not implausible from the point of view of algebra simply to identify corporations with sets. Let us think of a corporation as the set of entities that own stock in it, and in imitation of

legal practice let us expand the moral universe of individuals to include such sets. Corporations may of course own stock in other corporations. In such cases, moral terms would then be understood as ranging over both sets and individuals in a single type ordered by the membership relation. It would be odd, however, to attribute any moral content to the membership relation here. It certainly cannot be claimed to be a moral priority relation. In legal practice, for example, in a suit between A and B, it is generally irrelevant whether A owns stock in B.

A theory that is structurally the same is found in environmental ethics. It is sometimes claimed that species have moral rights and interests in a manner independent of the rights and interests of their members. It is sometimes argued, for example, that it is wrong to exterminate a species, such as the snail darter or the spotted owl, not because it is wrong to kill its members, nor because loss of the species would indirectly harm humans, but simply because the species as an entity has the right to exist. This view in effect adds species to the moral universe. By positing a relation $\leq$ among the levels of the taxonomic hierarchy, we can order this universe much as we did the one made up of individuals and corporations. Let $x \leq y$ if either x is an individual and y is the lowest taxonomic class into which x falls or x is a taxonomic class and y is the next class up in the hierarchy that embraces x. Presumably this ordering relation like the one for corporations is morally irrelevant.

The part-whole theories we have been discussing may be thought of as "holistic" if only in a weak sense. Although they incorporate some notion of "whole," the ordering does not bear on moral judgments. For more robust varieties of holism that posit part-whole relations with moral content, we must turn to theories that postulate fundamentally distinct orders of reality and a concept of "the whole" that unifies them.

V. DUAL TYPE THEORIES: DEGENERATE AND TRIVIAL CASES

Some theories go beyond the single type part-whole theories I have been discussing to postulate part-whole relations that fall into distinct categories, each of which is accorded its distinctive set of concepts and vocabulary. Such theories are sometimes said to posit different "orders of complexity." In natural science, for example, subatomic particles may be discussed as obeying laws written in terms of one vocabulary, atoms in a second vocabulary with its own laws, molecules in a third, cells and organ systems in a fourth, organisms in a fifth, and ecosystems or societies in a sixth. One variety of "reduction" consists then of showing that laws about wholes are in some sense equivalent to laws governing parts. Theories of this sort are very naturally treated as type theories with some ordering $<<$ holding between entities of neighboring types and with distinct vocabularies relativized to types.

In ethics there are a number of theories of this sort, but the most familiar are those that distinguish between the individual and the social good. Such theories share the idea that moral concepts apply to social groups. From the perspective of this study, groups may be viewed as "wholes" standing in a part-whole relation to the individuals that make them up. What makes this sort of theory different from the single type theories I sketched earlier is the postulation of a distinct set of moral concepts for groups. From the perspective of this study, theories of this sort may be said to posit two distinct types, a lower or *first-order* type and a higher or *second-order* type, a part-whole relation $<<$ between their elements, and a descriptive vocabulary for each type, including some moral terms which have their interpretations restricted to the higher type.

The simplest theory of this sort, in terms of its type structure, is one that restricts moral judgments to groups and makes no attempt to relate these laws to the properties of the individuals that make up the groups. This sort of account may be characterized as one that holds that social morality is conceptually independent of the properties of the individuals and individual actions that make up society. For example, it might be claimed that there is a notion of social interest and that what is socially good depends on what is in the interest of the society as a whole in a manner that is independent of the interests and other features of those that compose society. Some versions of fascism and social Darwinism take this form.

As moral theories, examples of this sort are in a sense degenerate cases of dual type theories. What they have to say about morality is independent of what they have to say about first-order entities inasmuch as the part of the theory that discusses first-order entities may be shaved off without altering its moral fragment. Indeed, the resulting single type theory of social groups has exactly the same moral laws as its dual type progenitor.

In contrast to cases of this sort there are also theories in which social ethics is at least partly "grounded" in features of the individual. The extent to which social moral judgments might be explicated in terms of first-order propositions varies widely, from theories that merely admit some relevance of one domain to the other to theories that fully define second-order moral concepts in first-order terms. Let us say that a theory's second-order moral fragment is *weakly grounded* in its first-order fragment if and only if at least one of them has some bearing on the truths of the other in the sense that the two are not logically independent: some contingent sentences (i.e., those not true in all acceptable models) written in the vocabulary of one fragment entail (in all acceptable models) some contingent sentences written in the vocabulary of the other fragment, or vice versa. At the other extreme, let us say a descriptive term in the second-order fragment is *grounded* if and only if the theory contains a sentence explicitly defining it in the vocabulary of the first-order fragment. We say, then, that the theory's second-order fragment is *totally grounded* if and only if each of

its descriptive moral terms is grounded. With grounding of either sort, let us say that a theory is *morally* grounded if the grounding first-order sentences contain moral terms.

Because it is part of the goal of moral theory to explain moral intuitions, a global moral theory should explain the intuitions we have about what is and is not morally acceptable for the state to do to the individual and what obligations the individual has to the state. That is, it is reasonable to expect that any adequate dual level moral theory will contain a large number of propositions relating moral ideas at the social level to the individuals, actions, or events that make up society.

Clearly, grounding alone is not a sufficient condition for the adequacy of a moral theory, but it does appear to be necessary. Grounding does not insure that the judgments of the theory match those of intuition; nevertheless, requiring weak grounding so that the grounding sentences might be matched to intuition seems to be a reasonable minimal condition of adequacy for any acceptable second-order moral theory. No doubt requiring this grounding to be total is too strong. Explicit definition of every term is a very strong requirement. On the other hand, it is hard to envisage an acceptable theory that does not contain a large number of propositions relating the second-order moral concepts to concepts of the first level in intuitively plausible ways.

It is perhaps too strong and somewhat question begging to require the minimal grounding to be moral and to explicate second-order moral terms using first-order moral terms. To do so essentially precludes the possibility of certain sorts of naturalistic ethics. This much we can say: a theory that is grounded is preferable to one that is not inasmuch as it has greater explanatory reach, making judgments testable against usage about the relation of second-order moral terms to first-order ideas. A theory that is not morally grounded, moreover, takes on the extra burden of arguing that first-order nonmoral terms are sufficient for characterizing second-order moral ideas. Without additional first-order moral theory, or arguments showing that first-order morality is superfluous, such naturalistic accounts leave unexplained issues relating personal to public morality.

What is implausible for a broad category of theories like fascism may in part be put in structural terms. Being degenerate and ungrounded, they fail to meet this minimal condition of adequacy. Their closed second-order moralities fail to articulate any obligation by the state to the individual, and thereby permit the individual to be sacrificed for the state.

There is also a limiting case on the opposite extreme that trivializes (in a logical sense) any obligation the individual may have to the state. The cases I have in mind are theories in which the group is said to have a moral property precisely when all its elements individually satisfy some set of first-order moral conditions. Hedonism, for example, may be trivially extended to a second-order theory by laying down the principle that a set of individuals is "good" if and only

if all the individuals in the group experience a minimal level of pleasure. Likewise, a good Christian regime might be defined as one in which all its citizens are good Christians. Indeed, the lack of a developed social ethic in relatively primitive moral theories like hedonism and simple-minded Christianity may be explained in part by saying that they do in fact seem to presuppose some such trivial extension of value judgments to society. Classical liberalism is another example. A society in which each individual respects the rights of others is a good society, and according to the classical theory, there is nothing more to be said about social justice or virtue.

In these sorts of theories second-order moral judgments are grounded. Indeed, the theories incorporate explicit definitions of the social virtues in terms of first-order properties of individuals. To tell whether the group as a whole is "good" it suffices to inspect each of the members of the group individually noting its first-order moral properties. Such theories trivialize social ethics in the sense that moral judgments about individuals can be straightforwardly extended to groups by a simple rule. Roughly, they posit some first-order condition C of moral acceptability and include among their laws a principle saying something to the effect that the group as a whole is "good" if and only if all elements of the group meet C. Whether such theories are adequate turn, in part, on whether all social moral judgments can be accounted for by some such trivial principle.

It should also be remarked that, *mutatis mutandis*, a single type theory of the social good can likewise be trivially extended to a dual type theory of individuals and groups by the rule that an individual is "good" if and only if it is a "part" of a whole that is good.

VI. DUAL TYPE THEORIES AND EMERGENT PROPERTIES

Emergent Properties

Between the two extremes of ungrounded theories, on the one hand, and trivial extensions of single type moral theories, on the other, there are potentially a wide variety of ways to ground second-order moral intuitions. However, one general method—or perhaps it should be called one class of approaches—seems to embrace most important theories of this sort. Included in this class are not only traditional utilitarianism, but also Rawls' difference principle, and such environmentalist theories as deep ecology and the ethics of the sublime. The method, moreover, is holistic in a nontrivial and structurally definable way.

In ethical theory nontrivial attempts at grounding may all be classified as varieties of emergence. Roughly, an emergent property is one that is defined for a whole in terms of the relational properties of its parts. Viewed epistemologically, you cannot tell whether an emergent property is possessed by a whole by looking piecemeal at the properties of its parts. You must look at how the parts stand to one another.

A good example is consistency. Consistency is grounded insofar as its application to a set of sentences is defined in terms of their elements. Moreover, the definition requires that the sentences stand in a specified relation. A set is consistent, under the semantic definition of the concept, if there is some world in which all its sentences are true. Alternatively, the set is consistent in the proof theoretic sense if there is no proof starting with sentences in the set and ending in a contradiction. On either definition, the same sentence is simultaneously an element of both consistent and inconsistent sets and the monadic properties of the sentences in a set do not determine whether it is consistent. You cannot tell whether a set is consistent without considering the exact group of sentences it contains.

Harmony is another example. The harmony of a painting is a function of the units of paint that make it up. On the other hand, you cannot tell from an inspection of the properties of these parts considered individually whether they form a harmonious whole. A given individual patch of paint with its nonrelational properties of color, shape, texture, etc. might be part of various paintings, some of which are harmonious, others not, depending on the properties of the pieces with which it is combined and the way in which these pieces are put together to form the whole.

For my purposes we may consider an emergent property to be any property of a whole defined in terms of the relational properties of its parts. A stricter definition relativizes the idea to a theory and employs a more careful formulation of *relational*. An expression is defined relationally in a theory if its definition cannot be expressed in that part of the theory's syntax that does not employ the theory's relational predicates or multiple quantifiers. We may say that a term E is *emergently defined* within a theory if the theory meets two conditions. First, the theory defines the application of E to a whole in terms of its parts in the sense that the theory posits a part-whole relation and contains sentences defining E with quantifiers restricted to the parts of wholes to which E applies. Second, the theory entails no set of necessary and sufficient conditions for the application of E written in the fragment of the theory restricted to the monadic predicate calculus. An *emergent property*, then, is one which, relative to a theory, is intended to be expressed by an emergently defined term.[4] Though the whole to

[4] This definition of emergent property is a development of that found in the logical literature intended to embrace phenomena like consistency and coherence. See, for example, John N. Martin, "The Emergent Nature of Coherence and the Possibility of Truth-Functional Logic," *Proceedings of the Tenth International Symposium on Multiple-valued Logic*, 1980, pp. 233–37. It is close to ideas found in aesthetics, as in, for example, Joseph Margolis, "Works of Art as Physically Embodied and Culturally Emergent Entities," *British Journal of Aesthetics* 14 (1974): 187–96. A similar issue is discussed in personal identity theory under the heading of methodological individualism. See James H. Fetzer, "Methodological Individualism: Singular Causal Systems and Their Population Manifestations," *Synthese* 68 (1986): 99–128. I am *not* using emergence as it is sometimes used in the philosophy of biology to mean a property of a higher-order of complexity that cannot be grounded in any way in lower-order properties.

which an emergent property applies need not, according to this definition, be of a different type from the parts, typically it will be. In the usual case an emergent property is defined as applying to a second-order whole in terms of specific relational properties holding among its grounding first-order parts.

UTILITARIANISM

Utilitarianism is a straightforward example of a theory that defines its key moral idea as a grounded second-order emergent property. I sketch how it does so in some detail here both to illustrate the type theoretic ideas at issue and to provide a model that can be extended in the subsequent discussion to less familiar and less developed theories.

What distinguishes utilitarianism from other sorts of grounded second-order moral theories is that it defines the social good as a maximization of some moral property of individuals in the group. In type theoretic terms, a utilitarian theory is one that defines the key second-order moral property as applying to a second-order object when its first-order constituents maximize some measurable first-order property. Generally, the second level type is societies or social groups and its key property is something like social goodness. In the characterization of the first-order type, however, theories in the utilitarian tradition vary widely, from narrow theories that admit only mature rational adult humans to those that admit children, the insane, fetuses, future generations, higher animals, lower animals, plants, inanimate objects, species, and ecosystems.[5] The relevant first-order comparative property also varies widely, from more traditional moral properties like happiness, to more or less naturalistic or scientific properties like pleasure, welfare, interest, and revealed preference.

In a simple version of the theory, the social good itself may be regarded as an on-off monadic property. The group is said to be good if some requisite quantity of the relevant first-order property is manifest jointly in the elements of the group. Let the relevant first-order measurable property be P and the minimal standard be N. Let us call the measurement of the degree of P its *utility*. The form of the key definition is something like: a society S of type T_2 is good if and only if the sum total of utility for P among the elements of type T_1 that compose S is at least N.[6]

It is perhaps more common, however, to understand the social good comparatively. If S and S' are second-order entities and P is the relevant first-order defining property that admits of measurable degrees, then the definition of the social good takes the following form: society S is morally better than society S' if

[5] For an extended discussion applying utilitarianism to environmental ethics, see Robin Attfield, *The Ethics of Environmental Concern* (New York: Columbia University Press, 1983).

[6] Some versions average the total. Doing so is irrelevant here and is consistent with the points made in the discussion.

and only if the sum total of the utility for P among the entities of type T_1 that make up S is greater than that for the entities that make up S'.

Since this is the first complex type theory we have met, it is appropriate to summarize the discussion in a careful definition. For the comparative version, a utilitarian moral theory for the moral property P is any set T of sentences and set M of models meeting the following conditions:

I. *Syntax*. The syntax in which the sentences of T are formulated must contain the usual logical vocabulary and the following special expressions: (1) type predicates T_1 and T_2, and a type composition predicate $<$; (2) a binary predicate "is better than" (briefly, B) intended to express the property of the social good; (3) a functor u that is intended to represent the operation measuring the property P; (5) a summation operator Σ; and (6) the predicate $\geq$ representing the greater than relation over numbers.

II. *Designated Sentences*. The set T must contain mathematical axioms sufficient to give the usual interpretation to the numerical expressions u, Σ and $\geq$ as well as the utilitarian moral law: for any second-order societies x and y, x is better than y, if and only if the sum of utility for first-order individuals in x is greater than the sum utility for individuals in y.[7]

III. *Semantics*. M is the set of acceptable models for T and is restricted to structures that are at least a dual type structure with type predicates T_1 and T_2, and composition relation $<<$ that links the two types and is referred to by a special predicate $<$. In addition, in any model in M, B is required to stand for a binary relation on T_2 and u to a function (usually called a *utility operation*) that is restricted in such a way that it assigns a numerical value to all elements of T_1.

It is because these definitions refer to the properties of first-order individuals that utilitarians are able to say what they do about the relation of the social good to the welfare (or whatever) of the citizens. Unlike the ungrounded fascist theories, utilitarianism can offer an explicit account of this relation. This account may not be intuitively satisfactory—whether it is is irrelevant here—but utilitarianism of any stripe meets the minimal structural requirement that it be able to offer some such account.

Rawls' Difference Principle

This framework can also be applied to another familiar example, John Rawls' difference principle. In Rawls' account the social good is identified with the maximization of the welfare of the least best off social group. A relevant feature of this theory from the perspective of our investigation is that it is open to a

---

[7] In logical notation, this law reads something like this:
$\forall x \forall y\{T_2x \& T_2y . \rightarrow . xBy \leftrightarrow \Sigma[u(z):T_1z \& z<x] \geq \Sigma[u(z):T_1z \& z<y]\}$

purely abstract characterization in terms of its type structure. Viewed this way, it is a straightforward and relatively natural variation on utilitarianism. Rawls' theory or, more correctly, family of theories may be seen as the result of adding to any utilitarian theory a moral priority relation over individuals. More precisely, a ramifying relation $\leq$ of moral priority is added to the first-order type in a utilitarian type structure. In this way, a Rawlsian can be more discriminating than the utilitarian in deciding whose welfare should be maximized. The theory intentionally builds in a moral bias. It is morally preferable to maximize the utility of those highest in the ranking. Although there are alternative ways in which the details might be formulated, it is sufficient for my purposes to sketch the particular way in which Rawls chooses to do so. He suggests that the best society is one that maximizes the welfare of those with the highest priority.

The set of the most important elements from this ranking, K, is the set of all first-order elements x such that there is no y other than x such that $x \leq y$. Relative to a measurable property P and a minimal standard N, society S can then be defined to be good if and only if the sum total of utility for P among the elements of type T_1 that are both in K and compose S is at least N. In the comparative case, society S can be defined to be morally better than society S' if and only if the sum total of degrees of utility for P among the entities of type T_1 in K that make up S is greater than that for the entities in K that make up S'. The result is an abstract version of Rawls' theory.[8] By a ramified utilitarian theory for a moral property P and a moral ordering R is meant the theory T and the set of models M that meet the conditions for being a utilitarian theory for P and that meet the following conditions as well:

I. *Syntax*. The syntax of T contains a dyadic predicate $\leq$ intended to express the idea of moral priority R.

II. *Designated Sentences*. The utilitarian moral law in T is replaced by: for any second-order societies x and y, x is better than y, if and only if the sum of utility for first-order individuals in x that are least well off under $\leq$ is greater than the sum of utility for individuals in y that are least well off under $\leq$.[9]

III. *Semantics*. The type structures in M have a ramifying relation $\leq$ on T_1 referred to by the special predicate $\leq$.

---

[8] For the difference principle see, John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971). The family of second-order emergent morally grounded moral theories is really quite large. It is a relatively straightforward exercise to give definitions which show that it embraces, for example, Buchanan and Tullock's choice principle—that society should choose only that on which all agree—and the Pareto preference principle—that the socially preferable is that which does not diminish the utility of any citizen. See James M. Buchanan and Gordon Tullock, *The Calculus of Consent* (Ann Arbor: University of Michigan Press, 1962) and Amartya K. Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970).

[9] In logical notation, $\forall x \forall y\{T_2x \ \& \ T_2y \rightarrow xBy \leftrightarrow \Sigma[u(z): T_1z \ \& \ z \prec x \ \& \ \exists w(z \leq w \ \& \ w \prec x)] \geq \Sigma[u(z): T_1z \ \& \ z \prec y \ \& \ \neg\exists w(z \leq w \ \& \ w \prec y)]\}$.

By further restricting the notion of moral priority, we can also capture an abstract version of Rawls' "difference principle" in which moral priority is attached to the "least well off." One way to do so is to require that the theory contain a "meaning postulate" that says in effect that $x \leq y$ if the measure of y's welfare is less than that of x. By a Rawlsian theory let us mean a ramified utilitarian theory T and M such that T contains the sentence: for any x and y in T_1, $x \trianglelefteq y$, if and only if, $u(y) \geq u(x)$.

Holism Defined

What is interesting about these theories from the point of view of the study of holism—and what their careful definitions show—is that in each example a higher-order moral good is grounded in the relational properties of a lower order. They each contain a moral law that characterizes the necessary and sufficient conditions for the application of the second-level moral predicate in terms of relational properties of first-level entities. The group forms a whole and the moral properties of the group are a function of the "communal" properties of individuals within it. Indeed, properties of this sort are quite fairly called *holistic* in a nontrivial and important sense, and it is this sense that I propose as the appropriate one for understanding the way in which an ethical theory is properly called holistic. Viewed this way, a descriptive term E is holistic within a theory if and only if the theory posits at least two types and E is grounded and emergently defined relative to them. A *holistic property* may then be identified as one that, relative to a theory, is expressed by a predicate that is holistically defined, and a *holistic theory* as one that posits a holistic property.

VII. HOLISM IN ENVIRONMENTAL ETHICS

The holistic ideas found in the nonstandard moral theories in environmental ethics seem to match this pattern. I consider two examples: deep ecology and the ethics of the sublime.

Deep Ecology. By deep ecology I mean any of various attempts to construct a moral theory from ecology broadly conceived. Although theories of this sort are not always carefully formulated, they all posit some idea of ecological balance or equilibrium applying to complex ecological systems. Even though *balance* is supposed to be defined purely in the terms of ecology understood as a natural science, it is intended also to be a key moral property. Roughly, "the good" is equilibrium in an ecosystem as defined in ecology. Because humans and their societies are parts of the environment, they fall within ecosystems, and deep ecology mandates that they contribute to the "good" to the extent that they help constitute a balanced or well-functioning ecosystem. Moreover, moral judgments are applied to nonhuman as well as human parts of nature. Any alter-

ation of the environment, such as the elimination of the spotted owl, is wrong, quite independent of its effect on humans, if it disturbs the ecological balance. In addition, such accounts often posit a moral preference in favor of larger ecosystems. On this view, if there is a choice between balance in two systems, the larger and more inclusive is morally preferable. Because for practical purposes the Earth's planetary isolation puts an upper limit on ecosystems, such theories sometimes identify the *summum bonum* with balance in the global environment.

The Ethics of the Sublime Less developed, but by no means rare, are attempts to explain environmental goodness in aesthetic terms. The hypothesis, crudely stated, is that right action consists of pursuing the beautiful, and applied on a large scale to natural objects, this equation provides a social and environmental ethic. The "good" is beauty, especially natural beauty. Moreover, proponents of such views often think that bigger is better. The beauty of larger bits of nature, often called *the sublime*, is morally preferable to that of smaller parts. Beauty of nature as a whole is the ultimate good.[10]

I have sketched these views in very broad strokes and do not intend here to try to refine them. Greater detail is unnecessary for my purposes, for it is possible at this level of precision to abstract some basic type-theoretic features of such accounts and to show in what ways they are holistic.

Both theories are essentially dual type. (1) Deep ecology posits a lower order of biota and their interactions, and a higher order of ecosystems. We may presume that there is a distinct ecological vocabulary for describing these two levels and a composition relation $<<$ in terms of which parts of the biotic community of the lower order help to form systems of the higher order. (2) The aesthetic approach clearly posits a domain of natural objects, some of which are beautiful. Presumably, the beauty of a natural object is explained in terms of some features or aspects into which it is decomposed on a more fundamental level, much as the harmony of a painting is explained in terms of the color, shape, and texture of the units of paint from which it is composed. We may think of such a theory as positing two types, one of object constituents and one of natural objects, and a compositional relation $<<$ joining the two and determining which constituents contribute to which objects.

Both theories also have as their key moral ideas concepts that are probably holistic in the substantive sense of being both grounded and emergent. In deep

[10] Various romantic attempts to draw moral inspiration from nature have been around for a long time. For an explicit recent attempt to design an ethical theory to explain moral intuition in terms of natural beauty as a *summum bonum*, see Norman H. Morse, "An Environmental Ethic—Its Formulation and Implication," in Philip P. Hanson, ed., *Environmental Ethics: Philosophical and Policy Perspectives* (Burnaby, British Columbia: Institute for the Humanities/SFU Publications, 1986). In a similar vein, J. Baird Callicott, in "The Land Aesthetic," *Companion to A Sand Country Almanac*, (Madison: University of Wisconsin Press, 1987) finds an aesthetic ethics within Leopold's land ethic.

ecology the good is ecosystem equilibrium, which is grounded in the properties of biota and defined, it is reasonable to assume, in terms of their relational properties. Likewise, if we continue to use harmony as a model, it is reasonable to think that natural beauty within the aesthetic theories is grounded in first-order constituents and defined in terms of the way these features stand to one another.[11]

Thus far the two accounts are structurally similar to utilitarianism, with all three approaches mandating dual type theories with second-order moral properties that are holistic in the sense of being grounded and emergent. In several ways, however, the environmental theories have structural features that distinguish them from utilitarianism and allow for some general comparisons.[12]

A relatively minor difference, due as much as anything to the undeveloped state of the environmental theories, is that utilitarianism is totally grounded whereas the environmental accounts, at least in the sketchy versions I have given, make no attempt to give an explicit definition of balance or beauty.

A more interesting difference from a structural perspective is the fact that both environmental theories posit morally relevant ramifying relations among second-order entities, whereas neither utilitarianism nor its Rawlsian variants do so. According to the utilitarian maxim, all that is relevant in appraising a society is its utility. In the environmental accounts, in contrast, size matters in addition to balance or beauty. Moreover, on both accounts it is natural to posit a supremum to the second-order ranking, in the Earth or nature as a whole. Environmental theories are thus in a sense doubly or even triply holistic. First, their key moral ideas are holistic in the sense of being grounded and emergent within a dual type framework; second, they posit a second-order morally relevant ramifying part-whole relation; and third, this relation has a single whole as its supremum.

The second-order part-whole relation, moreover, begets a series of ungrounded emergent properties definable entirely within the second-order type. Consider the case of a painting. Paintings or portions of paintings are said to possess harmony as an emergent property inasmuch as the harmony of the whole is determined by the monadic and relational properties of the units of paint that

---

[11] For a deep ecology that is explicitly emergent, see Rafal Serafin, "Noosphere, Gaia, and the Science of the Biosphere," *Environmental Ethics* 10 (1988): 121–37. Whether beauty is emergent will depend on one's theory of beauty. A good example of a theory in which beauty is clearly a second-order emergent property grounded in first-order physical features is that of Guy Sircello, *A New Theory of Beauty* (Princeton: Princeton University Press, 1975).

[12] For a discussion of the part-whole issues in environmental ethics see Christopher D. Stone, *Should Trees Have Standing?* (Los Altos, Calif.: William Kaufmann, 1972). Although Stone discusses some conceptual puzzles that seem to arise when the moral universe is extended to natural objects and ramified by a part-whole relation, he is essentially a utilitarian, holding moral concepts to be second-order. He attaches no special moral relevance to it. For more recent developments of his ideas see *Earth and Other Ethics: The Case for Moral Pluralism* (New York: Harper and Row, 1987).

constitute the painting and which are described in a vocabulary appropriate to a lower order of complexity. The units of paint and their properties are first-level phenomena, and the painting and its harmony are second-level. Now imagine cutting up the painting into several parts. These parts become miniature paintings that are composed of units of paint just like the original painting, and the painting parts are second-level entities like the original painting. For example, if someone asks whether a painting section is harmonious, the same considerations will be brought to bear on the section as were brought to bear on the whole. The section is harmonious if its component units of paint exhibit the requisite relational properties. Harmony for painting sections remains a second-level emergent property grounded in first-level entities.

We may now define a ramifying ordering among sections of potential paintings: $P \leq P'$ if P can result by cutting up P'. This ordering together with the grounded emergent property of harmony allows us to define a sense of harmony in which it is plausible to say that the harmony of a second-level whole "emerges" as a feature of the relational properties of its second-level parts. If P is harmonious and cut up into sections P' and P'', there is a perfectly good sense in which the harmony of P results from the relational properties of its second-order parts P' and P'': by inspecting P' and P'', as standing in the relation in which they do in P, it is possible to tell whether P is harmonious in a way that it is not possible if looking at P' and P'' in isolation.

The property P is *induced* by a grounded emergent property P' and a second-level ordering $\leq$ if P is defined as applying to a second-level entity X just when P' emerges from the relational properties of the entire group of first-level constituents that enter into any one of the $\leq$ parts of X. P induced in this way must be emergent if P' is, because it is indirectly a function of the relational parts of the parts of X, and hence directly a function of the relational parts of X.

On the other hand, an induced emergent property of this sort is trivially definable for any emergent property on a ramified type. It has no conceptual interest beyond that already possessed by the original emergent property. Drawing attention to this ancillary idea of emergence, however, does aid in clarifying which idea of emergence—relative to which sense of part—is the root phenomenon. In subjects like paintings, ecosystems, and natural objects, there are two notions of part-whole, one correctly identified with a type composition relation $<<$ and one as a ramifying ordering $\leq$ among second-order entities. The proper way to analyze the more basic notion of emergence and its resulting holism is as that which holds among entities of different types. Emergence within a type, though genuine, derives from the more basic idea. Although it is correct to observe, for example, that ecosystems fall under one another and that in a sense the equilibrium of the larger system is an outcome of the relational features of the systems it contains, the reason this is so is that the balance of the whole arises from the interaction at a lower level of complexity among the various biotic parts of the system, which also happen to make up the larger system's subsystems.

Similarly, though it is correct to say that the beauty of a natural object arises from the relational features of those natural objects that form it, the reason this is so is that its beauty arises from the relational features of its more basic parts as described at a lower level of complexity and these parts, by definition, constitute the natural objects that make the original whole.

The last way in which the environmental theories contrast with utilitarianism poses what is perhaps the most serious difficulty facing this sort of theory. Utilitarianism as well as the two environmental accounts meet the minimal condition on second-order moral theories discussed earlier. They all ground their higher-order moral concepts in some first-order propositions. Utilitarian theories, however, do so in a first-order vocabulary of concepts like happiness, pleasure, interests, or preference. Although on some accounts these ideas are intended to be naturalistic in the sense of being part of some natural science, all these ideas have a long history in ethics proper. A theory which explains the social good in terms of any of them has gone a long way toward what I have been calling moral grounding. The environmental theories, on the other hand, quite clearly intend to ground their second-order concepts in terms foreign to traditional ethics. The ecology of biota does not overlap traditional ethics at all, and it is hard to see how natural aesthetics can contribute in any major way to the understanding of traditional individualistic ethics.

Thus, I think it is fair to say that holistic environmental ethical theories suffer from a systematic disadvantage. In attempting to provide a naturalistic grounding for higher-order moral properties, they are employing theories with very little overlap with traditional ethics. It is accordingly very hard to see how they can possibly provide a *moral* grounding for their ideas.[13] They appear to be frustrated from the start in any attempt to generate the intuitively plausible principles relating first and second-level moral concepts which articulate the mutual obligations between the individual and the whole of which he or she is a part. At their worst, holistic environmental ethics might be like fascism in requiring the individual to sacrifice him or herself in the interest of ecological balance or natural beauty. At best, they leave personal ethics unexplained and face the difficult task of providing a convincing grounding of second-level moral concepts in a vocabulary new to the moral tradition.

---

[13] I am certainly not the first to criticize holistic environmental theories using what I am calling type theoretic ideas. Anthony Weston in "Forms of Gaian Ethics," *Environmental Ethics* 9 (1987): 217–30, makes the point that the Gaia hypothesis may be understood as adding the Earth itself to the moral universe of persons. Both the land ethic and deep ecology have been criticized for being, in my terms, ethically ungrounded, and failing to adequately explain moral intuitions about humans and their societies. See John Moline, "Aldo Leopold and the Moral Community," *Environmental Ethics* 8 (1986): 99–120; Eric Katz, "Organism, Community, and the 'Substitution Problem,' " *Environmental Ethics* (1985): 241–56; and Richard A. Watson, "A Note on Deep Ecology," *Environmental Ethics* 6 (1984): 377–80. I am making similar points in a more generalized form.