Chapter 8

Natural Kinds

W. V. O. Quine

What tends to confirm an induction? This question has been aggravated on the one hand by Hempel's puzzle of the non-black non-ravens,[1] and exacerbated on the other by Goodman's puzzle of the grue emeralds.[2] I shall begin my remarks by relating the one puzzle to the other, and the other to an innate flair that we have for natural kinds. Then I shall devote the rest of the chapter to reflections on the nature of this notion of natural kinds and its relation to science.

Hempel's puzzle is that just as each black raven tends to confirm the law that all ravens are black, so each green leaf, being a non-black non-raven, should tend to confirm the law that all non-black things are non-ravens, that is, again, that all ravens are black. What is paradoxical is that a green leaf should count toward the law that all ravens are black.

Goodman propounds his puzzle by requiring us to imagine that emeralds, having been identified by some criterion other than color, are now being examined one after another and all up to now are found to be green. Then he proposes to call anything *grue* that is examined today or earlier and found to be green or is not examined before tomorrow and is blue. Should we expect the first one examined tomorrow to be green, because all examined up to now were green? But all examined up to now were also grue; so why not expect the first one tomorrow to be grue, and therefore blue?

The predicate "green," Goodman says,[3] is *projectible*; "grue" is not. He says this by way of putting a name to the problem. His step toward solution is his doctrine of what he calls entrenchment,[4] which I shall touch on later. Meanwhile the terminological point is simply that projectible predicates are predicates ζ and η whose shared instances all do count, for whatever reason, toward confirmation of [All ζ are η].

Now I propose assimilating Hempel's puzzle to Goodman's by inferring from Hempel's that the complement of a projectible predicate need not be projectible. "Raven" and "black" are projectible; a black raven does count toward "All ravens are black." Hence a black raven counts also, indirectly, toward "All non-black things are non-ravens," since this says the same thing. But a green leaf does not count toward "All non-black things are non-ravens," nor, therefore, toward "All ravens are black"; "non-black" and "non-raven" are not projectible. "Green" and "leaf" are projectible, and the green leaf counts toward "All leaves are green" and "All green things are leaves"; but only a black raven can confirm "All ravens are black," the complements not being projectible.

If we see the matter in this way, we must guard against saying that a statement [All ζ are η] is lawlike only if ζ and η are projectible. "All non-black things are non-ravens" is a law despite its nonprojectible terms, since it is equivalent to "All ravens are black."

Reprinted from *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), pp. 114–138, by permission of the publisher and the author.

The contrast between properties and sets which I suggested just now must not be confused with the more basic and familiar contrast between properties, as intensional, and sets as extensional. Properties are intensional in that they may be counted as distinct properties even though wholly coinciding in respect of the things that have them. There is no call to reckon kinds as intensional. Kinds can be seen as sets, determined by their members. It is just that not all sets are kinds.

If similarity is taken simple-mindedly as a yes-or-no afair, with no degrees, then there is no containing of kinds within broader kinds. For, as remarked, similarity now simply means belonging to some one same kind. If all colored things comprise a kind, then all colored things count as similar, and the set of all red things is too narrow to count as a kind. If on the other hand the set of all red things counts as a kind, then colored things do not all count as similar, and the set of all colored things is too broad to count as a kind. We cannot have it both ways. Kinds can, however, overlap; the red things can comprise one kind, the round another.

When we move up from the simple dyadic relation of similarity to the more serious and useful triadic relation of comparative similarity, a correlative change takes place in the notion of kind. Kinds come to admit now not only of overlapping but also of containment one in another. The set of all red things and the set of all colored things can now both count as kinds; for all colored things can now be counted as resembling one another more than some things do, even though less, on the whole, than red ones do.

At this point, of course, our trivial definition of similarity as sameness of kind breaks down; for almost any two things could count now as common members of some broad kind or other, and anyway we now want to define comparative or triadic similarity. A definition that suggests itself is this: *a* is more similar to *b* than to *c* when *a* and *b* belong jointly to more kinds than *a* and *c* do. But even this works only for finite systems of kinds.

The notion of kind and the notion of similarity seemed to be substantially one notion. We observed further that they resist reduction to less dubious notions, as of logic or set theory. That they at any rate be definable each in terms of the other seems little enough to ask. We just saw a somewhat limping definition of comparative similarity in terms of kinds. What now of the converse project, definition of kind in terms of similarity?

One may be tempted to picture a kind, suitable to a comparative similarity relation, as any set which is "qualitatively spherical" in this sense: it takes in exactly the things that differ less than so-and-so much from some central norm. If without serious loss of accuracy we can assume that there are one or more actual things (*paradigm cases*) that nicely exemplify the desired norm, and one or more actual things (*foils*) that deviate just barely too much to be counted into the desired kind at all, then our definition is easy: *the kind with paradigm a and foil b* is the set of all the things to which *a* is more similar than *a* is to *b*. More generally, then, a set may be said to be a *kind* if and only if there are *a* and *b*, known or unknown, such that the set is the kind with paradigm *a* and foil *b*.

If we consider examples, however, we see that this definition does not give us what we want as kinds. Thus take red. Let us grant that a central shade of red can be picked as norm. The trouble is that the paradigm cases, objects in just that shade of red, can come in all sorts of shapes, weights, sizes, and smells. Mere degree of overall similarity to any one such paradigm case will afford little evidence of degree of redness, since it will depend also on shape, weight, and the rest. If our assumed relation of comparative similarity were just comparative chromatic similarity, then our paradigm-and-foil defiare a great help, and in our learning we often do depend on them. Still one would like to be able to show that a single general standard of similarity, but of course comparative similarity, is all we need, and that respects can be abstracted afterward. For instance, suppose the child has learned of a yellow ball and block that they count as yellow, and of a red ball and block that they do not, and now he has to decide about a yellow cloth. Presumably he will find the cloth more similar to the yellow ball and to the yellow block than to the red ball or red block; and he will not have needed any prior schooling in colors and respects. Carnap undertook to show long ago how some respects, such as color, could by an ingenious construction be derived from a general similarity notion;[8] however, this development is challenged, again, by Goodman's difficulty of imperfect community.

A standard of similarity is in some sense innate. This point is not against empiricism; it is a commonplace of behavioral psychology. A response to a red circle, if it is rewarded, will be elicited again by a pink ellipse more readily than by a blue triangle; the red circle resembles the pink ellipse more than the blue triangle. Without some such prior spacing of qualities, we could never acquire a habit; all stimuli would be equally alike and equally different. These spacings of qualities, on the part of men and other animals, can be explored and mapped in the laboratory by experiments in conditioning and extinction.[9] Needed as they are for all learning, these distinctive spacings cannot themselves all be learned; some must be innate.

If then I say that there is an innate standard of similarity, I am making a condensed statement that can be interpreted, and truly interpreted, in behavioral terms. Moreover, in this behavioral sense it can be said equally of other animals that they have an innate standard of similarity too. It is part of our animal birthright. And, interestingly enough, it is characteristically animal in its lack of intellectual status. At any rate we noticed earlier how alien the notion is to mathematics and logic.

This innate qualitative spacing of stimulations was seen to have one of its human uses in the ostensive learning of words like "yellow." I should add as a cautionary remark that this is not the only way of learning words, nor the commonest; it is merely the most rudimentary way. It works when the question of the reference of a word is a simple question of spread: how much of our surroundings counts as yellow, how much counts as water, and so on. Learning a word like "apple" or "square" is more complicated, because here we have to learn also where to say that one apple or square leaves off and another begins. The complication is that apples do not add up to an apple, nor squares, generally, to a square. "Yellow" and "water" are mass terms, concerned only with spread; "apple" and "square" are terms of divided reference, concerned with both spread and individuation. Ostension figures in the learning of terms of this latter kind too, but the process is more complex:[10] And then there are all the other sorts of words, all those abstract and neutral connectives and adverbs and all the recondite terms of scientific theory; and there are also the grammatical constructions themselves to be mastered. The learning of these things is less direct and more complex still. There are deep problems in this domain, but they lie aside from the present topic.

Our way of learning "yellow," then, gives less than a full picture of how we learn language. Yet more emphatically, it gives less than a full picture of the human use of an innate standard of similarity, or innate spacing of qualities. For, as remarked, every reasonable expectation depends on similarity. Again on this score, other animals are like man. Their expectations, if we choose so to conceptualize their avoidance movements and salivation and pressing of levers and the like, are clearly dependent on their appreciation of similarity. Or, to put matters in their methodological order, these avoidance movements and salivation and pressing of levers and the like are typical of scientific findings, all scientific conjectures that are at present plausible, are therefore in my view as welcome for use in philosophy as elsewhere. For me then the problem of induction is a problem about the world: a problem of how we, as we now are (by our present scientific lights), in a world we never made, should stand better than random or coin-tossing chances of coming out right when we predict by inductions which are based on our innate, scientifically unjustified similarity standard. Darwin's natural selection is a plausible partial explanation.

It may, in view of a consideration to which I next turn, be almost explanation enough. This consideration is that induction, after all, has its conspicuous failures. Thus take color. Nothing in experience, surely, is more vivid and conspicuous than color and its contrasts. And the remarkable fact, which has impressed scientists and philosophers as far back at least as Galileo and Descartes, is that the distinctions that matter for basic physical theory are mostly independent of color contrasts. Color impresses man; raven black impresses Hempel; emerald green impresses Goodman. But color is cosmically secondary. Even slight differences in sensory mechanisms from species to species, Smart remarks,[12] can make overwhelming differences in the grouping of things by color. Color is king in our innate quality space, but undistinguished in cosmic circles. Cosmically, colors would not qualify as kinds.

Color is helpful at the food-gathering level. Here it behaves well under induction, and here, no doubt, has been the survival value of our color-slanted quality space. It is just that contrasts that are crucial for such activities can be insignificant for broader and more theoretical science. If man were to live by basic science alone, natural selection would shift its support to the color-blind mutation.

Living as he does by bread and basic science both, man is torn. Things about his innate similarity sense that are helpful in the one sphere can be a hindrance in the other. Credit is due man's inveterate ingenuity, or human sapience, for having worked around the blinding dazzle of color vision and found the more significant regularities elsewhere. Evidently natural selection has dealt with the conflict by endowing man doubly: with both a color-slanted quality space and the ingenuity to rise above it.

He has risen above it by developing modified systems of kinds, hence modified similarity standards for scientific purposes. By the trial-and-error process of theorizing he has regrouped things into new kinds which prove to lend themselves to many inductions better than the old.

A crude example is the modification of the notion of fish by excluding whales and porpoises. Another taxonomic example is the grouping of kangaroos, opossums, and marsupial mice in a single kind, marsupials, while excluding ordinary mice. By primitive standards the marsupial mouse is more similar to the ordinary mouse than to the kangaroo; by theoretical standards the reverse is true.

A theoretical kind need not be a modification of an intuitive one. It may issue from theory full-blown, without antecedents; for instance the kind which comprises positively charged particles.

We revise our standards of similarity or of natural kinds on the strength, as Goodman remarks,[13] of second-order inductions. New groupings, hypothetically adopted at the suggestion of a growing theory, prove favorable to inductions and so become "entrenched." We newly establish the projectibility of some predicate, to our satisfaction, by successfully trying to project it. In induction nothing succeeds like success.

Between an innate similarity notion or spacing of qualities and a scientifically sophisticated one, there are all gradations. Sciences, after all, differs from common sense only in degree of methodological sophistication. Our experiences from earliest infancy are bound to have overlaid our innate spacing of qualities by modifying and

Another dim notion, which has intimate connections with dispositions and subjunctive conditionals, is the notion of cause; and we shall see that it too turns on the notion of kinds. Hume explained cause as invariable succession, and this makes sense as long as the cause and effect are referred to by general terms. We can say that fire causes heat, and we can mean thereby, as Hume would have it, that each event classifiable under the head of fire is followed by an event classifiable under the head of heat, or heating up. But this account, whatever its virtues for these general causal statements, leaves singular causal statements unexplained.

What does it mean to say that the kicking over of a lamp in Mrs. Leary's barn caused the Chicago fire? It cannot mean merely that the event at Mrs. Leary's belongs to a set, and the Chicago fire belongs to a set, such that there is invariable succession between the two sets: every member of the one set is followed by a member of the other. This paraphrase is trivially true and too weak. Always, if one event happens to be followed by another, the two belong to *certain* sets between which there is invariable succession. We can rig the sets arbitrarily. Just put any arbitrary events in the first set, including the first of the two events we are interested in; and then in the other set put the second of those two events, together with other events that happen to have occurred just after the other members of the first set.

Because of this way of trivialization, a singular causal statement says no more than that the one event was followed by the other. That is, it says no more if we use the definition just now contemplated; which, therefore, we must not. The trouble with that definition is clear enough: it is the familiar old trouble of the promiscuity of sets. Here, as usual, kinds, being more discriminate, enable us to draw distinctions where sets do not. To say that one event caused another is to say that the two events are of *kinds* between which there is invariable succession. If this correction does not yet take care of Mrs. Leary's cow, the fault is only with invariable succession itself, as affording too simple a definition of general causal statements; we need to hedge it around with provisions for partial or contributing causes and a good deal else. That aspect of the causality problem is not my concern. What I wanted to bring out is just the relevance of the notion of kinds, as the needed link between singular and general causal statements.

We have noticed that the notion of kind, or similarity, is crucially relevant to the notion of disposition, to the subjunctive conditional, and to singular causal statements. From a scientific point of view these are a pretty disreputable lot. The notion of kind, or similarity, is equally disreputable. Yet some such notion, some similarity sense, was seen to be crucial to all learning, and central in particular to the processes of inductive generalization and prediction which are the very life of science. It appears that science is rotten to the core.

Yet there may be claimed for this rot a certain undeniable fecundity. Science reveals hidden mysteries, predicts successfully, and works technological wonders. If this is the way of rot, then rot is rather to be prized and praised than patronized.

Rot, actually, is not the best model here. A better model is human progress. A sense of comparative similarity, I remarked earlier, is one of man's animal endowments. Insofar as it fits in with regularities of nature, so as to afford us reasonable success in our primitive inductions and expectations, it is presumably an evolutionary product of natural selection. Secondly, as remarked, one's sense of similarity or one's system of kinds develops and changes and even turns multiple as one matures, making perhaps for increasingly dependable prediction. And at length standards of similarity set in which are geared to theoretical science. This development is a development away from the immediate, subjective, animal sense of similarity to the remoter objectivity of that mechanism. This embarrassment of riches is, I suspect, a characteristic outcome. That is, once we can legitimize a disposition term by defining the relevant similarity standard, we are apt to know the mechanism of the disposition, and so bypass the similarity. Not but that the similarity standard is worth clarifying too, for its own sake or for other purposes.

Philosophical or broadly scientific motives can impel us to seek still a basic and absolute concept of similarity, along with such fragmentary similarity concepts as suit special branches of science. This drive for a cosmic similarity concept is perhaps identifiable with the age-old drive to reduce things to their elements. It epitomizes the scientific spirit, though dating back to the pre-Socratics: to Empedocles with his theory of four elements, and above all to Democritus with his atoms. The modern physics of elementary particles, or of hills in space-time, is a more notable effort in this direction.

This idea of rationalizing a single notion of relative similarity, throughout its cosmic sweep, has its metaphysical attractions. But there would remain still need also to rationalize the similarity notion more locally and superficially, so as to capture only such similarity as is relevant to some special science. Our chemistry example is already a case of this, since it stops short of full analysis into neutrons, electrons, and the other elementary particles.

A more striking example of superficiality, in this good sense, is afforded by taxonomy, say in zoology. Since learning about the evolution of species, we are in a position to define comparative similarity suitably for this science by consideration of family trees. For a theoretical measure of the degree of similarity of two individual animals we can devise some suitable function that depends on proximity and frequency of their common ancestors. Or a more significant concept of degree of similarity might be devised in terms of genes. When kind is construed in terms of any such similarity concept, fishes in the corrected, whale-free sense of the word qualify as a kind while fishes in the more inclusive sense do not.

Different similarity measures, or relative similarity notions, best suit different branches of science; for there are wasteful complications in providing for finer gradations of relative similarity than matter for the phenomena with which the particular science is concerned. Perhaps the branches of science could be revealingly classified by looking to the relative similarity notion that is appropriate to each. Such a plan is reminiscent of Felix Klein's so-called *Erlangerprogramm* in geometry, which involved characterizing the various branches of geometry by what transformations were irrelevant to each. But a branch of science would only qualify for recognition and classification under such a plan when it had matured to the point of clearing up its similarity notion. Such branches of science would qualify further as unified, or integrated into our inclusive systematization of nature, only insofar as their several similarity concepts were *compatible*; capable of meshing, that is, and differing only in the fineness of their discriminations.

Disposition terms and subjunctive conditionals in these areas, where suitable senses of similarity and kind are forthcoming, suddenly turn respectable; respectable and, in principle, superfluous. In other domains they remain disreputable and practically indispensable. They may be seen perhaps as unredeemed notes; the theory that would clear up the unanalyzed underlying similarity notion in such cases is still to come. An example is the disposition called intelligence—the ability, vaguely speaking, to learn quickly and to solve problems. Sometime, whether in terms of proteins or colloids or nerve nets or overt behavior, the relevant branch of science may reach the stage where a similarity notion can be constructed capable of making even the notion of intelligence respectable. And superfluous.

In general we can take it as a very special mark of the maturity of a branch of science that it no longer needs an irreducible notion of similarity and kind. It is that final stage where the animal vestige is wholly absorbed into the theory. In this career of the similarity notion, starting in its innate phase, developing over the years in the light of accumulated experience, passing then from the intuitive phase into theoretical similarity, and finally disappearing altogether, we have a paradigm of the evolution of unreason into science.

Notes

1. C. G. Hempel, *Aspects of Scientific Explanation and Other Essays* (New York: Free Press, 1965), p. 15.
2. Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, Mass., 1955, or New York: Bobbs-Merrill, 1965), p. 74. I am indebted to Goodman and to Burton Dreben for helpful criticisms of earlier drafts of the present paper.
3. Goodman, *Fact*, pp. 82 f.
4. *Ibid.*, pp 95 ff.
5. I mean this only as a sufficient condition of lawlikeness. See Donald Davidson, "Emeroses by other names," *Journal Of Philosophy* 63 (1966), 778–780.
6. This relevance of kind is noted by Goodman, *Fact*, first edition, pp. 119 f; second edition, pp. 121 f.
7. Nelson Goodman, *The Structure of Appearance*, 2d ed. (New York: Bobbs-Merrill, 1966), pp. 163 f.
8. Rudolf Carnap, *The Logical Structure of the World* (California, 1967), pp. 141–147. (German edition 1928).
9. See my *Word and Object*, pp. 83 f, for further discussion and references.
10. See *Word and Object*, pp. 90–95.
11. This was noted by S. Watanabe on the second page of his paper "Une explication mathématique du classement d'objets," in S. Dockx and P. Bernays, eds., *Information and Prediction in Science* (New York: Academy Press, 1965).
12. J. J. C. Smart, *Philosophy and Scientific Realism* (New York: Humanities, 1963), pp. 68–72.
13. Goodman, *Fact*, pp. 95 ff.
14. Carnap, "Testability and meaning," *Philosophy of Science* 3 (1936), 419–471; 4 (1937), 1–40.
15. Here there followed, in previous printings, 26 lines which I have deleted. They were concerned with explaining certain subjunctive conditionals on the basis of the notion of kind. Paul Berent pointed out to me that the formulation was wrong, for it would have equated those conditionals to their converses.