THE JOURNAL OF PHILOSOPHY

VOLUME LXXVI, NO. 2, FEBRUARY 1979

THE FOUR-COLOR PROBLEM AND ITS PHILOSOPHICAL SIGNIFICANCE *

THE old four-color problem was a problem of mathematics for over a century. Mathematicians appear to have solved it to their satisfaction, but their solution raises a problem for philosophy which we might call the *new four-color problem.*

The old four-color problem was whether every map on the plane or sphere can be colored with no more than four colors in such a way that neighboring regions are never colored alike. This problem is so simple to state that even a child can understand it. Nevertheless, the four-color problem resisted attempts by mathematicians for more than one hundred years. From very early on it was proved that five colors suffice to color a map, but no map was ever found that required more than four colors. In fact some mathematicians thought that four colors were not sufficient and were working on methods to produce a counterexample when Kenneth Appel and Wolfgang Haken, assisted by John Koch, published a proof that four colors suffice.† Their proof has been accepted by most mathematicians, and the old four-color problem has given way in mathematics to the new four-color theorem (4CT).

The purpose of these remarks is to raise the question of whether the 4CT is really a theorem. This investigation should be purely philosophical, since the mathematical question can be regarded as definitively solved. It is not my aim to interfere with the rights of

* I would like to thank Michael Albertson, Joan Hutchinson, and William Marsh for reading a draft of this paper and for some helpful discussions on a number of points.

† "Every Planar Map Is Four Colorable," *Illinois Journal of Mathematics*, xxi, 84 (September 1977): 429–567. Part I, on Discharging, is by Appel and Haken; part II, on Reducibility, was done in conjunction with Koch. Parenthetical page references to Appel, Haken, and Koch, will be to this article.

0022-362X/79/7602/0057$02.60

mathematicians to determine what is and what is not a theorem. I will suggest, however, that, if we accept the 4CT as a theorem, we are committed to changing the sense of 'theorem', or, more to the point, to changing the sense of the underlying concept of "proof." So, by raising the question of whether the 4CT has really been proved, I will be trying to elucidate the concept of proof and not attempting an evaluation of the mathematical work of Appel and Haken.

What reason is there for saying that the 4CT is not really a theorem or that mathematicians have not really produced a proof of it? Just this: no mathematician has seen a proof of the 4CT, nor has any seen a proof that it has a proof. Moreover, it is very unlikely that any mathematician will ever see a proof of the 4CT.

What reason is there, then, to accept the 4CT as proved? Mathematicians know that it has a proof according to the most rigorous standards of formal proof—a computer told them! Modern high-speed computers were used to verify some crucial steps in an otherwise mathematically acceptable argument for the 4CT, and other computers were used to verify the work of the first.

Thus, the answer to whether the 4CT has been proved turns on an account of the role of computers in mathematics. Even the most natural account leads to serious philosophical problems. According to that account, such use of computers in mathematics, as in the 4CT, introduces empirical experiments into mathematics. Whether or not we choose to regard the 4CT as proved, we must admit that the current proof is no traditional proof, no a priori deduction of a statement from premises. It is a traditional proof with a lacuna, or gap, which is filled by the results of a well-thought-out experiment. This makes the 4CT the first mathematical proposition to be known a posteriori and raises again for philosophy the problem of distinguishing mathematics from the natural sciences.

The plan of the argument is as follows. The paper begins with a preliminary analysis of the concept of 'proof' in order to extract certain features that will be useful to us later. Then the work of Appel, Haken, and Koch is described. The most natural interpretation of this work, I will argue, is that computer-assisted proofs introduce experimental methods into pure mathematics. This fact has serious implications not only for the philosophy of mathematics, but for philosophy in general, and we will examine some of these implications.

I

What is a proof? In this section three major characteristics of proofs will be considered:

(a) Proofs are convincing.
(b) Proofs are surveyable.
(c) Proofs are formalizable.

(a) Proofs are convincing. This fact is key to understanding mathematics as a human activity. It is because proofs are convincing to an arbitrary mathematician that they can play their role as arbiter of judgment in the mathematical community. On a very stark and skeptical position, such as is sometimes suggested in Wittgenstein's *Remarks on the Foundations of Mathematics,* this is all that there is to proofs: they are convincing to mathematicians. This is to be taken as a brute fact, something for which no explanation can be given and none is necessary. Most philosophers are unhappy with this position and instead feel that there must be some deeper characterization of mathematical proofs which explains, at least to some extent, why they are convincing. That proofs are surveyable and that they are formalizable are two such characterizations.

(b) Proofs are surveyable. Proofs are the guarantees of mathematical knowledge and so they must be comprehended by mathematicians. A proof is a construction that can be looked over, reviewed, verified by a rational agent. We often say that a proof must be perspicuous, or capable of being checked by hand. It is an exhibition, a derivation of the conclusion, and it needs nothing outside of itself to be convincing. The mathematician *surveys* the proof in its entirety and thereby comes to *know* the conclusion. Here is an example of a proof, attributed to the young Gauss, which helps to convey the idea of surveyability. It is a proof that the sum of the first one hundred positive numbers is 5050. Write down those numbers in two rows of fifty columns as shown:

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & \cdots & 49 & 50 \\ 100 & 99 & 98 & 97 & \cdots & 52 & 51 \end{array}$$

Observe that the sum of the two numbers in each column is 101 and that there are 50 columns. Conclude that the sum of the first one hundred positive numbers is 5050.

We now know that $1 + 2 + \cdots + 99 + 100 = 5050$. We have surveyed the proof in its entirety and become convinced. If someone actually attempted to add the numbers by hand and arrived at the sum 5048, we would say that he added wrong. The construction that we surveyed leaves no room for doubt. So it is with all mathematical proofs; to say that they can be surveyed is to say that they can be definitively checked by members of the mathematical community. Of course, some surveyable proofs are very long. They might take months for even a trained mathematician to review and work out—an example is Walter Feit and John G. Thompson's famous proof that all groups of odd order are solvable.[1]

Genius in mathematics lies in the discovery of new proofs, not in the verification of old ones. In a sense, the concept of surveyability provides for the democratization of mathematics by making proofs accessible to any competent mathematician. A teacher of mine, a very good mathematician but no genius, once remarked that there were only a few proofs that he couldn't understand, but that there were none that he could not follow.

Surveyability is an important subjective feature of mathematical proofs which relates the proofs to the mathematicians, the subjects of mathematical investigations. It is in the context of surveyability that the idea of 'lemma' fits. Mathematicians organize a proof into lemmas to make it more perspicuous. The proof relates the mathematical known to the mathematical knower, and the surveyability of the proof enables it to be comprehended by the pure power of the intellect—surveyed by the mind's eye, as it were. Because of surveyability, mathematical theorems are credited by some philosophers with a kind of certainty unobtainable in the other sciences. Mathematical theorems are known a priori.

(c) Proofs are formalizable. A proof, as defined in logic, is a finite sequence of formulas of a formal theory satisfying certain conditions. It is a deduction of the conclusion from the axioms of the theory by means of the axioms and rules of logic. Most mathematicians and philosophers believe that any acceptable proof can be formalized. We can always find an appropriate formal language and theory in which the informal proof can be embedded and "filled out" into a rigorous formal proof.

Formal proofs carry with them a certain objectivity. That a proof is formalizable, that the formal proofs have the structural properties that they do, explains in part why proofs are convincing to mathematicians.

[1] "Solvability of Groups of Odd Order," *Pacific Journal of Mathematics*, XIII (1963): 775–1029. It is important to realize that, despite its exceptional length, this proof was surveyed from start to finish by mathematicians including Feit, Thompson, and perhaps several dozen leading group theorists.

We've noted three features of proofs: that they are convincing, surveyable, and formalizable. The first is a feature centered in the anthropology of mathematics, the second in the epistemology of mathematics, and the third in the logic of mathematics. The latter two are the deep features. It is because proofs are surveyable and formalizable that they are convincing to rational agents.

Surveyability and formalizability can be seen as two sides of the same coin. Formalizability idealizes surveyability, analyzes it into finite reiterations of surveyable patterns. Certainly when the two criteria work together, mathematicians do not hesitate to accept or reject a purported proof. Nevertheless the two ideas spring from such different sources that we can wonder whether they will always work together. Can there be surveyable proofs that are not formalizable or formal proofs that cannot be surveyed?

Are all surveyable proofs formalizable? Most mathematicians and philosophers would assent, but not all. Some intuitionists deny that the actual proof constructions of mathematics can be completely captured by formal systems.[2] Intuitionism aside, however, it is well known that no single theory is sufficient to formalize every proof. Given any sufficiently rich theory, we can find a surveyable proof of a statement of that theory which has no formal proof. Such a statement can be a Gödel statement which, when properly interpreted, says that it has no formal proof. Of course the surveyable proof can be formalized in a new and more powerful formal theory; but that theory, in turn, will yield new surveyable proofs that it cannot formalize.

At best, formalizability is a local characteristic of proofs, not a global one. There is not one system in which any proof can be formalized; but rather, given any proof, there is some appropriate formal system in which it can be formalized. The point that formalizability is a local and not a global phenomenon is made by René Thom where he notes the general significance of this distinction for the philosophy of mathematics.[3] However since our concern will not be with surveyable proofs than cannot be formalized, let us turn to the second question.

Are all formalizable proofs surveyable? Consider first the simpler question: Are all formal proofs surveyable? Here the answer is an

[2] See, for example, Arend Heyting, *Intuitionism* (Amsterdam: North-Holland, 1966), ch. 1.

[3] "Modern Mathematics: An Educational and Philosophical Error?" *American Scientist*, LIX, 6 (November/December 1971): 695–699.

easy no. We know that there must exist formal proofs that cannot be surveyed by mathematicians if only because the proofs are too long or involve formulas that are too long. Here "too long" can be taken to mean "can't be read over by a mathematician in a human lifetime." So it is logically possible that mathematicians could come across a statement with no surveyable proof but with a formalized proof.

However, if we stop to think about this situation, it appears unlikely that this logical possibility can ever be realized. How is a mathematician to know that a statement has a formal proof? On the one hand, the mathematician might actually survey or look over the formal proof and check it for correctness. On the other hand, the mathematician can derive the existence of the required formal proof, in effect, by presenting a surveyable proof that the formal proof exists. This sort of thing is standard practice in proof theory, where we find, for example, general surveyable arguments that any proof in, say, elementary arithmetic can be formalized in Zermelo-Fraenkel set theory. Hence it begins to appear that, in practice, at least, mathematicians come to know formal proofs only through the mediation of surveyable proofs. Either the formal proofs are simple enough to be surveyed themselves and verified to be proofs, or their existence is established by means of informal surveyable arguments.

It is not really surprising that we should come to know the existence of specific formal proofs only through some more primitive concept of proof, surveyable proof. After all, in the last analysis, formal proofs are abstract mathematical objects. They can be represented by sets of natural numbers, Gödel numbers, without any loss of information. To state that there is a formal proof of a formula is very much like stating that there is a number with a certain property; and how are we to come to know the latter statement except by a proof?

In summary, although formal proofs outrun surveyable proofs, it is not at all obvious that mathematicians could come across formal proofs and recognize them as such without being able to survey them.

Nevertheless, it is the contention of this paper that the current proof of the 4CT does drive a wedge between the criteria of surveyability and formalizability. In fact, there is no surveyable proof, no proof in the traditional sense, of the 4CT, nor is there likely to be one. Still Appel, Haken, and Koch's work provides mathematically convincing grounds for the 4CT. What can be surveyed, what is presented in their published work, is like a mathematical proof where a key lemma is justified by an appeal to the results of certain computer runs or, as we might say "by computer." This appeal to computer, whether we count it as strictly a part of a proof or as a part of some explicitly non-proof-theoretic component of mathematical knowledge, is ultimately a report on a successful experiment. It helps establish the 4CT (actually, the existence of a formal proof of the 4CT) on grounds that are in part empirical.

The idea that a particular proposition of pure mathematics can be established, indeed must be established, by appealing to empirical evidence is quite surprising. It entails that many commonly held beliefs about mathematics must be abandoned or modified. Consider: [4]

1. All mathematical theorems are known a priori.
2. Mathematics, as opposed to natural science, has no empirical content.
3. Mathematics, as opposed to natural science, relies only on proofs, whereas natural science makes use of experiments.
4. Mathematical theorems are certain to a degree that no theorem of natural science can match.

In order to assess such claims, let us quickly review the proof of the 4CT.

II

Sooner or later any discussion of the 4CT must begin talking of graphs in place of maps, so we might as well begin at once.[5] We can think of a *planar graph* as a finite collection of points in the

[4] To be sure, not all philosophers hold these beliefs, but they are common enough to warrant criticism. Some philosophers have argued against them, notably Imre Lakatos in *Proofs and Refutations* (New York: Cambridge, 1976) and Hilary Putnam in *Mathematics, Matter and Method* (New York: Cambridge, 1975). Putnam, in particular, explicitly rejects the traditional view of mathematics as an absolutely a priori discipline set apart from natural science. He suggests replacing it with the view of mathematics as *quasi-empirical*. The present paper provides additional support for the thesis that mathematics is quasi-empirical.

[5] For a simple account of the proof, see Appel and Haken, "The Solution of the Four Color Map Problem," *Scientific American*, CXXXVII, 8 (October 1977): 108–121. (Parenthetical page references to Appel and Haken are to this article; similarly for the authors cited below.) More detailed summaries can be found in Haken, "An Attempt to Understand the Four Color Problem" and F. Bernhart, "A Digest of the Four Color Theorem," both published in the *Journal of Graph Theory*, I (1977): 193–206 and 207–225, respectively. P. Kainen and T. Saaty provide an account of the theorem along with the required basis in graph theory in *The Four Color Problem: Assaults and Conquest* (New York: McGraw Hill, 1977). The definitive statement of the proof appears in Appel, Haken, and Koch, *op. cit.*

plane, called *vertices*, which are joined to each other by lines, called *edges*, such that no edges meet except at vertices. The number of edges meeting at any vertex is called the *degree* of the vertex, and vertices joined by an edge are said to be *neighboring*, or adjacent. A graph is *4-colorable* if every vertex can be colored by one of four colors in such a way that neighboring vertices never receive the same color.

If every planar graph can be 4-colored, then every planar map can be. This is because every map determines a graph, its *dual graph*, as follows: place one vertex (capital city) in each region (country) of the map and join the capitals of neighboring regions by an edge (road) that crosses their common border. Obviously, the resulting graph is 4-colorable if and only if the original map is.

Next we restrict our attention to graphs in a standard form. We can delete any parallel edges, edges joining two vertices already joined by another edge, without affecting 4-colorability. Graphs without parallel edges or loops are called *simple graphs*. Moreover, we can add edges by a process of triangulation. Given any region or polygon of the *graph* that is bounded by four or more edges, there will be at least two non-adjacent vertices on the boundary. We can join such vertices by a new edge across the region which does not intersect any other edge (except at the vertices). Continuing in this way, we can completely triangulate a graph until all regions have three sides. Since triangulation can only make 4-coloring more difficult because it restricts the possible colorings of a graph, it suffices to prove the 4CT for triangulated graphs.

Now any planar triangulation has only finitely many vertices; so the way to prove that all such graphs can be 4-colored is by induction on the number v of vertices. In case $v \leq 4$, the triangulation can be 4-colored. So we assume as induction hypothesis that any planar triangulation G' with n or fewer vertices is 4-colorable. We wish to show that, if G is a planar triangulation with $n + 1$ vertices, then G can be 4-colored.

There is a well-known formula relating the number of vertices a triangulation can have to the degrees of the individual vertices. If v_i is the number of vertices of degree i and if m is the maximum degree of any vertex in the triangulation, then Euler's formula states that

$$3v_3 + v_4 + v_5 + 0 \cdot v_6 - v_7 - 2v_8 - 3v_9 - \cdots - (m - 6)v_m = 12$$

At least one of v_3, v_4, v_5 must be nonzero; so any triangulated graph has a vertex with five or fewer edges. Incidentally, this fact suffices to prove, by induction, that any graph can be 6-colored.

Look at the triangulation G and delete a vertex of degree 5 along with its edges. The resulting graph has one less vertex and, when triangulated, it can be 6-colored, by the induction hypothesis. However, the missing vertex has at most five neighbors, so one color will be left to color it.

To prove that any graph G can be 4-colored, we consider the following cases.

Case 1. G contains a vertex of degree 3; i.e., $v_3 \neq 0$.

Then, if we delete the vertex along with its adjacent edges, we get a graph with n vertices which can be 4-colored by assumption. Since the missing vertex has only three neighbors, it can be colored by the remaining color.

Case 2. $v_3 = 0$ but $v_4 \neq 0$; the graph G contains a vertex of degree 4.

Again, delete the vertex of minimal degree, call it v_0, and its adjoining edges, to obtain a smaller graph which is 4-colorable.

Subcase 2a. If the four neighbors of the missing vertex are colored by only three colors, then v_0 can be colored the remaining color.

Subcase 2b. The four neighbors of v_0 are each colored differently. This coloring cannot be extended to G directly, but must first be modified. Call the neighbors of v_0 v_1', v_2', v_3', v_4', and suppose that they are respectively colored a, b, c, d. Look at the smaller graph G' $(G - v_0)$, and consider the subgraph of G' determined by all vertices colored a or c along with any edges connecting two such vertices. One of two alternatives must arise. Either there is an a–c chain of points and edges connecting v_1' to v_3', or there is not.

Subcase 2bi. If there is no such path between v_1 and v_3, we say that v_1 and v_3 belong to separate a–c components of G'. In this case reverse the colors in the a–c component containing v_3'. All vertices in this component formerly colored a are now colored c, and vice-versa. The resulting coloring is still a 4-coloring of G' since no neighboring vertices are colored the same, but the vertex v_3' is now colored a. The color c is not used to color any neighbor of v_0; so c can be used to color v_0.

Subcase 2bii. If there is such an a–c path connecting v_1' and v_3', then these vertices belong to the same a–c component of G', and reversing the colors won't help. However, in this case there cannot be a b–d path connecting v_2' and v_4', for any such path is blocked by the a–c path connecting v_1' and v_3'. Thus v_2' and v_4' belong to separate b–d components of G', and by reversing the colors in the

b–d component containing v_4', we obtain a 4-coloring of G' in which v_4' and v_2' are both colored b, leaving d to color v_0.

In either case the 4-coloring of G' can be modified and extended to a 4-coloring of G. The argument used in subcase 2b is called a *Kempe chain argument.* Incidentally, this type of argument can be applied to a vertex of degree 5 to show that any graph can be 5-colored.

If G has a vertex of degree 3 or 4, then G is 4-colorable; so we may assume that $v_3 = 0 = v_4$, and thus we come to case 3.

Case 3. $v_5 \neq 0$, the minimum degree of any vertex in G is 5. In this case the simple proof breaks down; Kempe chain arguments do not suffice if we delete a single vertex of degree 5. Instead of deleting a single vertex, we must try to delete configurations, or systems of interconnected vertices. If we remove a configuration from a triangulation we are left with a graph with a "hole" in it. The vertices of the remaining graph which are adjacent to the hole form a circuit, or *ring* around the configuration. The size of the ring is determined by the number of vertices in it. A *configuration* can be more precisely defined as a subgraph with specifications of the number of vertices, vertex degrees, and the manner in which it is embedded in the original triangulation.

A configuration is *reducible* if the 4-coloring of any planar graph containing it is deducible from the 4-colorability of any graph with fewer vertices. Reducible configurations transmit 4-colorability upwards. Conversely, if G is a graph that *requires* five colors and if G contains the reducible configuration C, then the subgraph $(G–C)$ requires five colors. By 1913, George Birkhoff had investigated the general methods of showing that a configuration was reducible.[6] In outline what must be proved is that every 4-coloring of the ring around a given configuration can either be extended to a 4-coloring of the configuration, or modified first by one or more Kempe interchanges and then extended, or modified by suitable identification of distinct vertices and then extended. A natural plan for attacking the four-color problem suggests itself. We can try to find a set of reducible configurations which is sufficiently large so that every triangulation contains a configuration from that set. Such an *unavoidable* set of configurations would enable us to complete the induction step in case 3. This plan runs into two related problems: the potential size of the unavoidable set and the potential size of

[6] "The Reducibility of Maps," *American Journal of Mathematics*, xxxv (1913): 114–128.

the reducible configurations in it. As Haken observes, the amount of work required to prove that a configuration is reducible increases considerably with the ring size. For a ring of size 14, the number of possible colorations is $3^{14} + 3$ (about 2×10^5). In principle, each one of these colorations must be examined in showing that the configuration is reducible. On the other hand Edward F. Moore found a triangulation that does not contain any known reducible configuration of ring size less than 12. Thus, in order to find enough reducible configurations to fill out an unavoidable set, we will have to include some with large ring size.

In order then to establish case 3, we must find a finite list of reducible configurations such that every graph contains at least one configuration from the list. Building on some work of Heinrich Heesch, Appel and Haken developed a theory of discharging procedures any of which produces an unavoidable set of configurations, i.e., a set that no triangulation ($v_3 = v_4 = 0$) can avoid. Heesch had noticed that certain kinds of configurations were reduction obstacles in that they could not be reduced by known methods. In a preliminary study, Appel and Haken developed a discharging procedure that produced an unavoidable set of configurations which excluded two of the three major reduction obstacles of Heesch. This set the stage for the final assault on the four-color conjecture.

Appel and Haken began with a discharging algorithm and tested for reducibility the configurations in the resulting unavoidable set. Whenever a configuration in the list could not be shown reducible, the discharging algorithm was modified to produce a new unavoidable set that excluded the recalcitrant configuration although generally it included new configurations. The configurations of the new set were checked for reducibility, and so on. Although the discharging procedure and the reducibility checks on individual configurations went hand in hand, and computer work was in practice necessary to develop both, when they had finished, the work of Appel, Haken, and Koch fell nicely into two parts.

The authors could specify a discharging procedure and prove in a mathematically rigorous fashion that this procedure produced an unavoidable set U of 1834 configurations (in fact, only 1482 of these configurations are really necessary). Although computer work was used to develop the procedure and the resulting set U, once the set was produced it could be surveyed and is listed in figures 1 to 63 of Appel, Haken, and Koch. Moreover, one can give a surveyable proof that this set U is unavoidable (see the Discharging Theorem and corollary in Appel, Haken, and Koch, 460).

However, to complete the proof of case 3, we need the lemma: Every configuration in U is reducible (actually, we need something a little stronger, but this version will suffice for our purposes. See Appel, Haken, and Koch on immersion reducibility). The proof of this lemma *cannot* be surveyed in detail. That these configurations are reducible is established by programming a computer to test for reducibility and running the program on the configurations in U. Since most of the configurations have large ring size (13 or 14), the use of computers to check reducibility is "unavoidable." Appel and Haken define a measure of complexity according to which the complexity of a proof of the D-reducibility of a 13-ring configuration will exceed 10^6 although other reductions (C-reducibility) of the same configuration might be of much less complexity (p. 487). In any case, no computer has printed out the complete proof of the reducibility lemma, nor would such a printout be of much use to human mathematicians. Over 1200 hours of computer time were required for the proof. Because of the complexity and time required, any proof of the reducibility lemma along its present lines must include an appeal to computer analysis. Thus it must presuppose the legitimacy of that appeal.

In its over-all outlines, the logic of the four-color proof is easy to see. It is a proof by induction which requires several cases. The first case is trivial, the second has several subcases, and the third has over a thousand subcases most of which cannot be handled except by high-speed computers. I would like to remove any impression that Appel and Haken's work is simply a "brute force" argument. To a certain extent, the appeal to computers might be regarded as "brute force," but it makes sense only when set in the context of a novel and sophisticated theory developed by the authors. However, establishing a theorem by introducing a novel and sophisticated theory is not in itself a novel mathematical procedure. The appeal to computers in order to ground key lemmas is.

To be sure, the use of computers in mathematics, even very sophisticated use, is not unfamiliar. We can cite programs for solving differential equations or the program of Hao Wang to prove theorems of propositional logic.[7] What makes the use of computers

[7] "Toward Mechanical Mathematics" in K. Sayre and F. Cooson, eds., *The Modeling of the Mind* (Notre Dame, Ind.: University Press, 1963), pp. 91–120. J. Weizenbaum, *Computer Power and Human Reason* (San Francisco: W. H. Freeman, 1976), pp. 230/1.

in the 4CT so dramatic is that it leads to a genuine extension of our knowledge of pure mathematics. It is not merely calculation, but yields a proof of a substantial new result.

Let us conclude this section with some general remarks on the complexity of the mathematical argument. Is the above proof of the 4CT, including computer work, the simplest or shortest proof of the 4CT? Might a surveyable proof be found some day?

Obviously some simplification is possible. Between the write-up of the proof and its publication it was found that 429 configurations could be eliminated from the set U. Further reduction could no doubt be achieved by modifying the discharging procedure. Nevertheless, it seems that any significant simplification of one part of the proof is likely to be matched with an increase in the complexity of another part of the proof. The current consensus among mathematicians is that the present proof is reasonably close to the simplest proof.[8] If this is so, then the appeal to computers would be essential to any mathematical justification of the 4CT.

Of course, no one can completely rule out the possibility that some mathematician will one day come up with a ten-page proof of the 4CT along lines currently unimaginable. (Although even here there are some grounds for skepticism; see Kainen and Saaty, 96.) Still, from a philosophical point of view such a discovery would have to be regarded as mere luck. The philosophical point at issue, obviously, is not simply the status of the 4CT, but the status of computer-assisted proofs in general. The work of Appel, Haken, Koch, and IBM 370-168 guarantees that the possibility of computer-assisted proofs is a real possibility.

III

The materials for our problem have been assembled. We have discussed some general features of proofs and some details of the proof of the 4CT. We can now ask whether the 4CT is really a theorem. Let us consider it with regard to the three characteristics of proofs.

(a) Is the proof of the 4CT convincing? Yes, most mathematicians have accepted the 4CT, and none, to my knowledge, has argued against it. Still, it should be noted that Appel and Haken themselves have recognized that there could be some resistance to their work, particularly from those mathematicians "educated before the development of high-speed computers" (Appel and Haken, 121).

[8] Appel, Haken, and Koch, part I, sec. 5; Bernhart, p. 224.

In any case, that an argument is convincing is not sufficient reason to accept it as a proof.

(b) Has the 4CT a surveyable proof? Here the answer is no. No mathematician has surveyed the proof in its entirety; no mathematician has surveyed the proof of the critical reducibility lemma. It has not been checked by mathematicians, step by step, as all other proofs have been checked. Indeed, it cannot be checked that way. Now Appel, Haken, and Koch *did* produce something that was surveyable in the sense that it could be looked over. Their work, as we have said, is very much like a surveyable proof with a lacuna where a key lemma is justified by nontraditional means—by computer. Incidentally, we must be wary of verbal entanglements here. Of course, if we call the appeal to computers a "new method of proof" in the strictest sense, then, trivially, the 4CT will have a surveyable proof. But the notion of proof itself will have shifted to accommodate the new method.

More serious is the objection that the appeal to computers is not a method of proof at all and that the idea that it is arises from a confusion between a proof and a description of a proof. Often mathematicians forgo a complete proof and make do with a description or a sketch of the proof sufficiently detailed for their purposes. In such descriptions, mathematicians may justify a lemma by reference to some already published work, by indicating the general method (e.g., "by diagonalizing") or by simply leaving the proof of the lemma as an exercise for the reader. Of course, these are not necessarily new methods of proof; in point of fact, they are more like shorthand, a brief way of indicating a proof. These devices belong to the description of the proof and not to the proof itself. The objection suggests that we regard Appel, Haken, and Koch's papers as descriptions of a proof (which they are) and try to assimilate the appeal to computers to the pragmatic shortcuts we've just noted.

The objection fails because there is a major difference between the cases. Traditionally any such abbreviation has been backed by a surveyable proof, even more, by a surveyed proof. Some mathematician and usually several mathematicians have surveyed the real thing and verified it. In principle this surveyable backing is available to any member of the mathematical community, either directly, as when the mathematicians can work it out for themselves, or indirectly, when they look it up in the archives, to use Wittgenstein's phrase. But it is just this surveyable backing that is lacking in the 4CT! Mathematicians cannot work out the missing steps for themselves, not even in a lifetime of work; and it is nowhere recorded in the archives. What is recorded is the evidence that a computer once worked out the missing steps. So it would be a grave mistake to classify the appeal to computers as a theoretically dispensable convenience, like the appeal to published journal articles. Of course the appeal "by computer" does mark an abbreviation, and later we will consider it in more expanded form. The point at hand, however, is that surveyability is preserved in traditional descriptions of proofs, but not in the appeal to computers.

Let us consider a hypothetical example which provides a much better analogy to the appeal to computers. It is set in the mythical community of Martian mathematicians and concerns their discovery of the new method of proof "Simon says." Martian mathematics, we suppose, developed pretty much like Earth mathematics until the arrival on Mars of the mathematical genius Simon. Simon proved many new results by more or less traditional methods, but after a while began justifying new results with such phrases as "Proof is too long to include here, but I have verified it myself." At first Simon used this appeal only for lemmas, which, although crucial, were basically combinatorial in character. In his later work, however, the appeal began to spread to more abstract lemmas and even to theorems themselves. Oftentimes other Martian mathematicians could reconstruct Simon's results, in the sense of finding satisfactory proofs; but sometimes they could not. So great was the prestige of Simon, however, that the Martian mathematicians accepted his results; and they were incorporated into the body of Martian mathematics under the rubric "Simon says."

Is Martian mathematics, under Simon, a legitimate development of standard mathematics? I think not; I think it is something else masquerading under the name of mathematics. If this point is not immediately obvious, it can be made so by expanding on the Simon parable in any number of ways. For instance, imagine that Simon is a religious mystic and that among his religious teachings is the doctrine that the morally good Martian, when it frames the mathematical question justly, can always see the correct answer. In this case we cannot possibly treat the appeal "Simon says" in a purely mathematical context. What if Simon were a revered political leader like Chairman Mao? Under these circumstances we might have a hard time deciding where Martian mathematics left off and Martian political theory began. Still other variations on the Simon theme are possible. Suppose that other Martian mathematicians begin to realize that Simonized proofs are possible where the attempts at more traditional proofs fail, and they begin to use "Simon says" even when Simon didn't say! The appeal "Simon says" is an anomaly in mathematics; it is simply an appeal to authority and not a demonstration.

The point of the Simon parable is this: that the logic of the appeals "Simon says" and "by computer" are remarkably similar. There is no great formal difference between these claims: computers are, in the context of mathematical proofs, another kind of authority. If we choose to regard one appeal as bizarre and the other as legitimate, it can only be because we have some strong evidence for the reliability of the latter and none for the former. Computers are not simply authority, but warranted authority. Since we are inclined to accept the appeal to computers in the case of the 4CT and to reject the appeal to Simon in the hypothetical example, we must admit evidence for the reliability of computers into a philosophical account of computer-assisted proofs. The precise nature of this evidence will concern us later. For now it suffices to note that, whatever the evidence is, it cannot take the form of a traditional, surveyable proof. Otherwise Appel and Haken would have given that proof and dispensed with the appeal to computers altogether.

The conclusion is that the appeal to computers does introduce a new method into mathematics. The appeal is surveyable, but what it appeals to is not.

(c) Has the 4CT a formalizable proof? Most mathematicians would concur that there is a formal proof of the 4CT in an appropriate graph theory. We can describe the formal proof in some detail, actually exhibit sections of it, calculate the total length, and so on. Nevertheless, this belief in the formal proof cannot be used to legitimize the appeal to computers. Rather, we believe that the formal proof exists only because we accept the appeal to computers in the first place. It is important to get the order of justification correct. Some people might be tempted to accept the appeal to computers on the ground that it involves a harmless extension of human powers. On their view the computer merely traces out the steps of a complicated formal proof that is really out there. In fact, our only evidence for the existence of that formal proof presupposes the reliability of computers.

This point can be clarified by the Simon parable. Martian mathematicians could say that "Simon says" incorporates no new method of proof and say that any Martian proof was still formalizable. They could claim that all of Simon's work was formalizable, only they themselves couldn't always provide the formalization. This is much the same position we claim to be in with respect to the appeal to computers. The comparison makes clear that formalization comes in only after the fact. It cannot be used as the criterion for accepting computer-assisted proofs.

In summary, the proof of the 4CT, although much like a traditional proof, differs in certain key respects. It is convincing, and there is a formal proof. But no known proof of the 4CT is surveyable, and there is no known proof that a formal proof exists. The crucial difference between the 4-color proof and traditional proofs is that the 4-color proof requires the appeal to computers to fill the gap in an otherwise traditional proof. The work of the computer is itself not surveyable. However, there are very good grounds for believing that this computer work has certain characteristics, e.g., that it instantiated the pattern of a formal proof of the reducibility lemma. Let us consider these grounds.

What does the appeal to computers amount to? Remember, we are now considering the appeal in the context of justifying a mathematical result, not yet in the context of discovery. We have a given mathematical question: Are the configurations in the unavoidable set U reducible? As part of the question, we are given procedures for testing configurations for reducibility. Second, we have a given machine with such and such characteristics. On the basis of our question and the machine's characteristics we construct a program of instruction for the machine. In this case the program is intended to "cause" the machine to "search" through the set U, testing each configuration for reducibility and reporting yes or no as the case may be. Finally we run this program on the computer and note the results. The appeal to computers, in the case of the 4CT, involves two claims: (1) that every configuration in U is reducible if a machine with such and such characteristics when programmed in such and such a way produces an affirmative result for each configuration, and (2) that such a machine so programmed did produce affirmative results for each configuration. The second claim is the report of a particular experiment. It has been experimentally established that a machine of type T when programmed by P will give output O.

But even the conditional conjunct is, at best, an empirical truth and not subject to traditional proof. Its truth depends on two interrelated factors, the reliability of the machine and the reliability of the program. The reliability of the machine is ultimately a matter for engineering and physics to assess. It is a sophisticated natural science that assures us that the computer "does what it's supposed to" in much the same way that it assures us that the electron microscope "does what it's supposed to." Of course, even if we grant that the machine does what it is supposed to—follow the program—there remains the question of whether the program does what *it* is supposed to. This question can be difficult to answer. The task of evaluating programs is a topic of computer science, but at present there are no general methods for accomplishing it at this level. Programs themselves are written in special "languages," and many of them can be quite complex. They can contain "bugs," or flaws that go unnoticed for a long time. The reliability of any appeal to computers must ultimately rest on such diffuse grounds as these.

In the case of the 4CT, most mathematicians feel that the reliability is sufficiently high to warrant a qualified acceptance of the theorem. In the first place, the problem was reducible to computer-manageable complexity. There is a very clear idea of what the computer is supposed to be doing—we have a good understanding of reduction techniques. Moreover, there is a great deal of accumulated evidence for the reliability of computers in such operations, and the work of the original computers was checked by other computers. Finally, there is good reason to believe that the theorem could not be reached by any other means. It is natural for mathematicians, at least for those educated after the development of high-speed computers and pocket calculators, to accept the truth of the 4CT. The reliability of the 4CT, however, is not of the same degree as that guaranteed by traditional proofs, for this reliability rests on the assessment of a complex set of empirical factors.

A digression on the reliability of computer-assisted proofs. No detailed estimate of this reliability, nor a general account of how such estimates should be made is offered here. Instead, let us try to probe our own subjective idea of computer reliability in mathematics by means of the following hypothetical examples.

In the case of the 4CT we understand the general shape of the computer proof. Would we be prepared to rely on computers even when we could not perceive the general shape of their work? Suppose that advances in computer science lead to the following circumstances. We can program a computer to initiate a search through various proof procedures, with subprograms to modify and combine procedures in appropriate circumstances, until it finds a proof of statement A. After a long time, the computer reports a proof of A, although we can't reconstruct the general shape of the proof beyond the bare minimum (e.g., by induction). Perhaps we could describe this hypothetical example by saying that the supercomputer found a human-assisted proof. Mathematicians served to aim the computer in a certain direction, to provide it with certain techniques, and it went on to find a cumbersome patchwork proof consisting of thousands of cases. Again, the question is whether mathematicians would have sufficient faith in the reliability of computers to accept this result.

The idea that a computer program can surprise its originators is not really very farfetched. The Appel-Haken program did surprise them.

> It was working out compound strategies based on all the tricks it had been taught, and the new approaches were often much cleverer than those we would have tried. In a sense the program was demonstrating superiority not only in the mechanical parts of the task but in some intellectual areas as well (Appel and Haken, 117).

Suppose some such supercomputer were set to work on the consistency of Peano arithmetic and it reported a proof of *inconsistency*, a proof which was so long and complex that no mathematician could understand it beyond the most general terms. Could we have sufficient faith in computers to accept this result, or would we say that the empirical evidence for their reliability is not enough? Would such a result justify a mathematician's claim to know that Peano arithmetic was inconsistent, and would such a mathematician have to abandon Peano arithmetic? These are bizarre questions, but they suggest that the reliability of computer-assisted proofs in mathematics, though easy to accept in the case of the 4CT, might some day be harder to swallow.

In conclusion, we have seen why it is reasonable to accept the 4CT, even the crucial reducibility lemma. There is no surveyable proof of the lemma, but we know that there is a formal proof. Our knowledge of this is grounded, in part, in the results of a well-conceived computer experiment. A wedge has been driven between the two explanations of proof in terms of surveyability and formalizability. In addition, a new technique has been developed for establishing mathematical truths. It is largely a matter of notational convention whether we choose to describe the new technique—appeal to computers—as a method of proof or refuse to call it a proof and insist on describing it as an experiment. In the former case, we would count the 4CT as a bona fide theorem. In the latter case we would not count it a theorem in the strict sense but admit it as a new kind of mathematical knowledge. Mere choice of labels cannot mask the underlying reality, which is an unavoidable reliance on computer experiments to establish the 4CT. Let us now turn to the consequences of this fact for philosophy.

IV

The acceptance of the 4CT is significant for philosophy at a number of points. In the first place, it is relevant to philosophy in general, especially to the theory of knowledge. Obviously, it is relevant to the details of any philosophy of mathematics. Finally, it is relevant to some issues in the philosophy of science.

Mathematics has always been important to philosophical theorizing about knowledge and reason, of course, both because mathematics stands as one of the pinnacles of human reason and rational thought and because mathematical knowledge can appear so perplexing if not actually mysterious.

> The science of pure mathematics, in its modern developments, may claim to be the most original creation of the human spirit.[9]

> The apparent contrast between the indefinite flux of sense-impressions and the precise and timeless truths of mathematics has been among the earliest perplexities and problems not of the philosophy of mathematics only, but of philosophy in general.[10]

A widely shared assumption among philosophers is that there is a significant gulf between mathematics and mathematical knowledge on the one hand, and natural science and scientific knowledge on the other. Thoroughgoing empiricists have denied that this gulf exists and have tried to explain mathematical truth, for example as Mill did, as a very general type of empirical truth. Such explanations have not been very persuasive, and, in general, philosophy has assumed that the gulf between mathematics and natural science exists and has tried to characterize the different kinds of knowledge involved by some contrasting pair, e.g., a priori, a posteriori; innate, learned; formal, empirical; certain, dubitable; analytic, synthetic. Once established, these characterizations become philosophical tools that can be applied elsewhere in the theory of knowledge.

[9] A. N. Whitehead, *Science and the Modern World* (New York: New American Library, 1959), p. 25.

[10] S. Körner, *The Philosophy of Mathematics* (New York: Harper, 1960), p. 9.

Mathematical knowledge plays a role in establishing these characterizations by serving as a paradigm of one pole in the dichotomy. The proof of the 4CT, however, undercuts this role. Knowledge of the 4CT does not have any of the characteristics that the paradigm suggests. Let us examine the case of the a priori/a posteriori distinction; the other cases proceed along similar lines.

Traditionally, a priori truths are those truths which can be known independently of any experience and a posteriori truths are those which can be known only on the basis of particular experiences. An a priori truth might be immediately evident, stipulated by convention, or, most common, known by reason independently of any experience beyond pure thought. It is plausible to maintain that such theorems as the mini-theorem that the sum of the first one hundred positive numbers is 5050 are known by reason alone—we all know it and could demonstrate its truth if we desired. However, it is not plausible to maintain that the 4CT is known by reason alone.

By reason alone, we know that the reducibility lemma implies the 4CT; but our knowledge of the reducibility lemma does not take the form of a proof. Our knowledge rests on general empirical assumptions about the nature of computers and particular empirical assumptions about Appel and Haken's computer work. Moreover, it is unlikely that anyone could know the 4CT by reason alone. The only route to the 4CT that we can ever take appears to lead through computer experiments. Thus the 4CT is an a posteriori truth and not an a priori one; mathematicians, I suggest, will never know the 4CT by a priori means.[11]

It is with the claim that the 4CT is not a priori that I differ from the position suggested taken by Saul Kripke when he considers the example of a computer verification that some very large number is a prime.[12] Kripke argues that such a theorem would be known a posteriori for the same reasons that I give that the 4CT is known a posteriori. But he leaves open the question of whether his theorem can be known a priori. I have argued that the 4CT cannot be known a priori by us.

The 4CT is a substantial piece of pure mathematics which can be known by mathematicians only a posteriori. Our knowledge must be qualified by the uncertainty of our instruments, computer

[11] See the qualifications expressed on page 69 of this paper, at the end of sec. II.

[12] "Naming and Necessity," in D. Davidson and G. Harman, eds., *Semantics of Natural Language* (Boston: Reidel, 1972), p. 261.

and program. There surely are truths from electrical engineering about current flow through switching networks which have a higher degree of certainty than the 4CT. The demonstration of the 4CT includes not only symbol manipulation, but the manipulation of sophisticated experimental equipment as well: the four-color problem is not a formal question. In fact, the argument for the 4CT is very like an argument in theoretical physics where a long argument can suggest a key experiment which is carried out and used to complete the argument.

This is a bit of a puzzle. In the first place, it blurs the intuitive distinction between mathematics and natural science which we began with. In the second place, we are left with the question of how to explain the role of experiment in pure mathematics. It is easy to see how experiments play a role in the arguments of physical theory. The physical theory can predict phenomena of space-time which equipment can be designed to register. Are we to say that the computer registered a phenomenon of mathematical space? If not, then how else are we to explain the role of experiment in mathematics? Such puzzles are one aspect of what I have called "the new four-color problem." I will not attempt any solutions to the puzzles here, but simply note these puzzles as among the consequences of the 4CT.

Not every way of characterizing the difference between mathematics and natural science falls to the 4CT. Following Kripke, we can argue that all mathematical truths, even the 4CT, are necessary, or true in all possible worlds. The 4CT, we might say, records an essential property of planar maps. (The truths of natural science, on the other hand, might be counted as contingent, or true in our world but false in some possible world.) In this case the 4CT would be an important example of an a posteriori necessary truth and, a fortiori, a counterexample to the claim that all known necessary truths are known a priori.

The new four-color problem then might serve as a stimulus to general philosophy to rethink the commonly accepted relations among knowledge, reason, and experience. Nevertheless, the most significant impact of the 4CT in philosophy obviously will concern the details of our philosophy of mathematics.

Accepting the 4CT forces us to modify our concept of proof. We can modify it by admitting a new method (computer experiment) of establishing mathematical results in addition to proofs. Or we can modify it by allowing proof to include computer-assisted proofs. I prefer the latter terminology. Either way, the details of this new method can have a substantial impact on the way mathematics is done.

This points to one of the most exciting aspects of Appel, Haken, and Koch's work, but one we have not touched on yet. So far we have been concerned with the 4CT only in the context of its justification: given the purported proof, does it prove the theorem? We have not treated it in the context of discovery. Any conclusions based only on discovery would have invited the Fregean retort that what matters to philosophy is justification and not genesis. It is time to widen our perspective; for there is much of interest about the discovery of the 4CT both to mathematics and to philosophy.

How does one decide to attempt a computer experiment in mathematics? Even where questions of the form $P(n)$ are decidable and we have the techniques to program a computer to check the instances, we cannot simply run the computer as long as it will go, hoping that it finds, say, that $(\exists x)P(x)$ before the computer reaches its limits. There must be some reason to expect that the computer will stop with an answer within a reasonable time. In the case of the 4CT we can ask why anyone thought that an unavoidable set of reducible configurations each of ring size less than or equal to 14 could be found. From the outside, 14 looks no more probable as a bound than 20 or 50 or even 100. Yet, if the minimum ring size were 20 or more, the required proof experiment could not be conducted at present! From the other direction, we know because of Moore's map that we must include configurations whose ring size is at least 12. Perhaps Moore would discover a map requiring the minimum ring size to be 20. Why did Appel and Haken think that a computer experiment could work?

What happened was that they developed a sophisticated probabilistic argument, not a proof, that the ring size could be restricted to 17 or less, and that the restriction to 14 was a good bet. They provided an argument that invested statements of the form "There is an unavoidable set of reducible configurations each of which has a ring size less than or equal to n" with a probability derived from the ratio of the number of vertices in the configuration to the ring size n (Haken, 202). With $n = 14$, the statement was very likely. Together with this probabilistic argument was an argument that the required techniques could be programmed into a computer. Koch did much of the work on the programming, and in their earlier paper Appel and Haken had showed that there was an unavoidable set of geographically good configurations of manageable size. These two arguments made it feasible to conduct the experiment.

The first type of argument is especially interesting. It is a new kind of argument endowing mathematical statements with a probability. This probability cannot be accounted for in ontological terms according to which any statement is true, or false, in all possible worlds. Having modified the concept of proof to include computer-assisted proofs, we might want to modify it again to include the kind of probabilistic argument required to set up a computer experiment. In practice this would amount to permitting mathematicians to make such arguments as part of their mathematical work. That is, it might be counted as a significant mathematical step if someone were to argue that a certain statement is very likely to be true, while leaving it to someone else to design and run the actual computer experiment. We must take this possibility much more seriously after the work of Appel and Haken, who established that such probabilistic arguments can have an important function in mathematics.

On the other hand, such probabilistic arguments inevitably contain the possibility of error; they can go wrong in a way strict proofs cannot.

> To use the computer as an essential tool in their proofs, mathematicians will be forced to give up hope of verifying proofs by hand, just as scientific observations made with a microscope or telescope do not admit direct tactile confirmation. By the same token, however, computer-assisted mathematical proof can reach a much larger range of phenomena. There is a price for this sort of knowledge. It cannot be absolute. But the loss of innocence has always entailed a relativistic world view; there is no progress without risk of error (Kainen and Saaty, 98).

These shifts in the concept of proof initiated by the 4CT force us to reevaluate the role of formal proofs in the philosophy of mathematics. Of course such shifts cast no doubt whatever on the legitimacy of formal proof theory as a branch of mathematical logic. Formal proofs, as idealized abstraction, still figure in our account of the 4CT. Nevertheless, after the 4CT, formal proofs cannot continue to serve the philosophy of mathematics as the sole paradigm of mathematical activity. Philosophers and mathematicians have already noted the limitations of the formal paradigm, but the 4CT aggravates these limitations to the point of a problem.[13] The old idea that a proof is a thought-experiment suggests

[13] See, for example, Lakatos, *op. cit.*

itself here. There is not such an apparent gulf between thought-experiments and computer-experiments as there is between formal proofs and experiments. On the other hand, there is not such a gulf between thought-experiments in mathematics and thought-experiments in physics either.

The primary impact of the new four-color problem in the philosophy of mathematics is on the concept of proof. We have discussed some of the consequences here.[14]

The relevance of the new four-color problem to the philosophy of science is largely a reworking of the earlier consequences. It is especially relevant to that branch of the philosophy of science which looks upon science as diachronic, or developing over time. In particular, it is relevant to the concept of paradigm outlined by Thomas Kuhn.[15] Paradigms, according to Kuhn, are scientific achievements that some scientific community accepts as supplying a foundation for its further practice. To qualify as a paradigm, the achievement must be both "sufficiently unprecedented to attract an enduring group of adherents away from competing modes of scientific activity" and "sufficiently open-ended to leave all sorts of problems for the redefined group of practitioners to resolve" (10). The concept of paradigms plays an important role in Kuhn's explanation of the development of science. It is natural to wonder whether the methodology leading to the 4CT can serve as a paradigm in mathematics; Kainen and Saaty have suggested that it will. "In fact, the Appel-Haken methodology suggests a new paradigm for mathematics. This paradigm includes the traditional elements of intuition and standard logic, as well as heuristic and probabilistic techniques combined with the high order computational abilities of a modern computer" (96).

Looking at the 4CT from the viewpoint of paradigms and thereby placing it in a historical perspective can be very illuminating. I suggest that if a "similar" proof had been developed twenty-five years earlier, it would not have achieved the widespread acceptance that the 4CT has now. The hypothetical early result would probably have been ignored, possibly even attacked (one thinks of the early reaction to the work of Frege and of Cantor). A necessary condition for the acceptance of a computer-assisted proof is wide

[14] For another approach that focuses on the idea of "difficult proof" and its relation to incompleteness results, see Haken, *op. cit.*

[15] *The Structure of Scientific Revolutions* (Chicago: University Press, 1962).

familiarity on the part of mathematicians with sophisticated computers. Now that every mathematician has a pocket calculator and every mathematics department has a computer specialist, that familiarity obtains. The mathematical world was ready to recognize the Appel-Haken methodology as legitimate mathematics.

Before we can satisfactorily describe the 4CT in terms of paradigms, however, there are two obstacles that must be overcome. The concept of paradigm has been developed primarily for the natural sciences with some extensions to the social sciences. We would first have to extend the notion of paradigm to mathematics, both by example and by explanation of the nature of mathematical paradigms.[16] Many philosophers would resist the extension of paradigms to mathematics, of course. In the current philosophy of mathematics, mathematics is viewed solely as a synchronic or timeless structure. Against this position it might be argued that it is simply working out of another paradigm of mathematics, the formal paradigm provided by Cantor, Frege, Russell, and Hilbert. The controversy will be decided, in part, by whether the paradigm model of mathematics can provide a more satisfactory account of achievements like the 4CT than can the formal model.

A second difficulty in extending the notion of paradigm to mathematics is historical. Paradigms are defined in terms of their past performance; they are achievements that had a major effect on the development of their fields. It is one thing to characterize an achievement as a paradigm on the basis of the historical record. It is quite another to predict that a recent achievement will function as a paradigm on the basis of the limited data currently available. It is clear that claims of the second kind must be much more tentative. However, if any such claims succeed, they are likely to provide much more information to the metatheory of paradigms than is provided by the simple classification based on the historical record. Although there are obstacles to treating the 4CT as providing a new paradigm for mathematics, any attempts to solve these problems can be important exercises in the philosophy of science.

Mathematicians have solved their four-color problem, but there is a new four-color problem that has arisen for philosophy. I have tried to explain what this problem is and how it arises. I have argued for its philosophical significance by noting some of the consequences that our acceptance of the 4CT has for the theory

[16] Much material useful for this enterprise can be found in the works of Lakatos and in Raymond Wilder, *Evolution of Mathematical Concepts* (New York: Wiley, 1968).

of knowledge, the philosophy of mathematics, and the philosophy of science.

THOMAS TYMOCZKO

Smith College

COUNTERFACTUALS AND CONSISTENCY *

ALL the things that would have been true if Bizet and Verdi had been compatriots should form a coherent if somewhat sparse picture of a possible state of affairs. From this simple consideration can be extracted several conditions of adequacy on theories of counterfactuals. First and foremost is a consistency condition, to the effect that "would" implies "could"—to ensure that all the things that would be true under any properly entertainable hypothesis are things that at least could be jointly true.

I shall argue that the semantics that David Lewis has presented in his book *Counterfactuals*,† does not secure this adequacy condition. On his assumptions concerning comparative similarity over possible worlds, what will be called *counterfactual inconsistencies* arise in a systematic way. The source of these violations of the adequacy condition will be traced back to Lewis's rejection of what he calls the "Limit Assumption" (sec. 1.4), a restriction on comparative similarity relations which he regards as objectionable and unjustified. The present paper offers a justification for that restriction, relative to the general framework Lewis has provided: under his semantics the Limit Assumption is both necessary and sufficient for guaranteeing counterfactual consistency.

I

For any sentence A, consider the set of all its *counterfactual consequents:* the set ΘA of all sentences B such that $A \,\Box\!\!\rightarrow B$ is true.[1] Following Lewis, let a sentence A be called *counterfactually entertainable* if and only if it is true at some world accessible from the

* I am indebted throughout to ch. I of John Pollock's *Subjunctive Reasoning* (Boston: Reidel, 1976; parenthetical page references to Pollock are to this book). I have also had the benefit of criticism from David Lewis, extended discussions with Isaac Levi, and very helpful comments from a number of other colleagues

† Oxford: Blackwell; Cambridge, Mass.: Harvard, 1973. Parenthetical page references to Lewis are to this book.

[1] Lewis calls ΘA the counterfactual theory for A, and he indexes it for reference worlds, thus: $\Theta(A, i) = \{B\colon A \,\Box\!\!\rightarrow B \text{ holds at } i\}$ (133).

0022-362X/79/7602/0083$00.50

LINKED CITATIONS

The Four-Color Problem and Its Philosophical Significance
Thomas Tymoczko
The Journal of Philosophy, Vol. 76, No. 2. (Feb., 1979), pp. 57-83.
Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28197902%2976%3A2%3C57%3ATFPAIP%3E2.0.CO%3B2-M

[Footnotes]

[6] **A Theorem for the Development of a Function as an Infinite Product**
A. F. Carpenter
American Journal of Mathematics, Vol. 35, No. 1. (Jan., 1913), pp. 105-114.
Stable URL:
http://links.jstor.org/sici?sici=0002-9327%28191301%2935%3A1%3C105%3AATFTDO%3E2.0.CO%3B2-4

NOTE: *The reference numbering from the original has been maintained in this citation list.*