

WOLFGANG HAKEN AND THE FOUR-COLOR PROBLEM

ROBIN WILSON

ABSTRACT. In 1852, Augustus De Morgan, Professor of Mathematics at University College, London, was asked: *Can every map be colored with just four colors in such a way that neighboring countries are colored differently?* Over a century later, in a controversial proof that made substantial use of a computer, Wolfgang Haken and Kenneth Appel of the University of Illinois answered the question in the affirmative. But how did Haken come to be involved with the problem, and what was his role in its solution?

1. The first eighty years

In October 1852 Francis Guthrie, a former student of Augustus De Morgan, was coloring the counties of a map of England. Noticing that only four colors were needed, he asked whether four colors suffice for all maps (see Figure 1). De Morgan quickly became fascinated by this problem and communicated it to other mathematicians. The first known printed reference to it is in *The Athenaeum* of June 10, 1854.

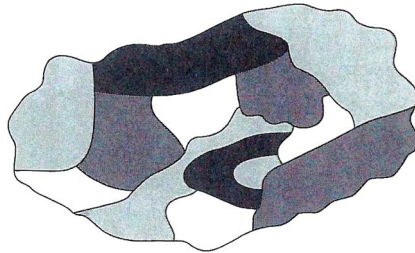


FIGURE 1. A map colored with four colors.

Received September 21, 2015; received in final form January 8, 2016.
2010 *Mathematics Subject Classification*. 05C15.

No progress was made on its solution until 1878, when Arthur Cayley posed the question again at a meeting of the London Mathematical Society. In a short note for the Royal Geographical Society [14], he subsequently gave a simple argument to show that when trying to prove that four colors are always enough it is sufficient to consider cubic maps—those in which exactly three countries (and therefore three boundary lines) meet at each point of intersection. In what follows, we shall usually assume that the maps under consideration are cubic maps, and our approach will be to explain why a *minimal counter-example*—a cubic map with the smallest possible number of countries that *cannot* be colored with four colors—cannot exist.

In the following year, Alfred Kempe (a barrister and former student of Cayley) answered Guthrie’s question in the affirmative. His paper ‘On the geographical problem of the four colours’ [19] appeared in the newly founded *American Journal of Mathematics* and was commissioned from him by Cayley’s friend J. J. Sylvester. Although Kempe’s ‘solution’ was incorrect, it contained a number of important ideas that resurfaced in later attempts on the problem, including the eventual proof by Appel and Haken.

First, using Euler’s polyhedron formula for maps,

$$\begin{aligned} &(\text{number of countries}) + (\text{number of meeting points}) \\ &= (\text{number of boundary edges}) + 2, \end{aligned}$$

Kempe proved that every cubic map must contain at least one country bounded by not more than five edges—a digon, triangle, quadrilateral, or pentagon, as illustrated in Figure 2. We call any such set an *unavoidable set of configurations*—every cubic map must contain at least one of them.



FIGURE 2. Kempe’s unavoidable set.

Kempe also proved by mathematical induction that any map containing a digon or triangle can be colored with four colors. For, if the remainder of the map is colored with four colors, then this coloring can immediately be extended to the digon or triangle, since the countries surrounding it require at most three colors. He then extended this idea to maps that contain a quadrilateral—here the four countries surrounding it may all be different, but Kempe showed that we can always interchange pairs of colors in the map so that a color becomes available for the quadrilateral. Such an interchange of colors is now called a *Kempe-interchange*.

A *reducible configuration* of countries in a map is a collection of countries with the property that any coloring of the rest of the map can be extended to these countries, either directly or after a succession of Kempe-interchanges of color; thus, a digon, a triangle, and a quadrilateral are all reducible configurations. Note that no reducible configurations can appear in a minimal counter-example to the four-color theorem.

Where Kempe ran into difficulties was in trying to show that a pentagon is reducible. To do so, he carried out two interchanges of color for the countries surrounding the pentagon, so that a color then became available for the pentagon. Since this dealt with every case, he believed that he had completed the proof.

In 1890, P. J. Heawood [18] showed that carrying out two such simultaneous interchanges of color is invalid in general, so that Kempe's proof contained a fatal flaw. But Heawood was able to salvage enough from Kempe's argument to prove that *every map can be colored with five colors*—itself a worthwhile result.

For future reference, we note that:

To prove the four-color theorem, it is enough to find an unavoidable set of reducible configurations.

This is because every map must contain at least one of these configurations, but each is reducible and so cannot appear in a minimal counter-example.

The next advance took place in 1904 when Paul Wernicke, an American mathematician studying for his doctorate in Göttingen, produced a new unavoidable set [31]. Since the pentagon had not been proved to be reducible, he replaced it by two other configurations (see Figure 3), hoping that these new ones would turn out to be reducible—specifically, he proved that any cubic map with no digon, triangle, or quadrilateral must contain two adjacent pentagons or a pentagon adjacent to a hexagon. We prove this later in this article.

Later, in 1922, Philip Franklin of Princeton University (and later M.I.T.) extended this result by proving that any cubic map with no digon, triangle, or quadrilateral must contain a pentagon adjacent to two other pentagons, a pentagon and a hexagon, or two hexagons [15]. Further unavoidable sets were discovered by Henri Lebesgue (of Lebesgue integral fame) in 1940 [20].

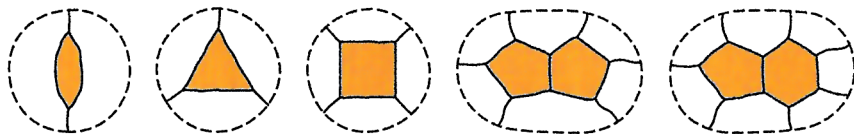


FIGURE 3. Wernicke's unavoidable set.

Meanwhile, progress was being made on reducible configurations. Up to now it was known that digons, triangles, and quadrilaterals were reducible, but little further progress had been made. This situation changed dramatically in 1913 with the publication of a ground-breaking paper by G. D. Birkhoff, one of the outstanding American scholars of the early 20th century who made substantial contributions to many areas of mathematics. Birkhoff regarded solving the four-color problem as one of his greatest aspirations, even though he later regretted the time he had spent on it.

Birkhoff paper on ‘The reducibility of maps’ [13] appeared in the *American Journal of Mathematics*, where Kempe had published his ‘solution’ over thirty years earlier. In this paper Birkhoff gave a systematic treatment of reducible configurations and laid the groundwork for all later developments in this direction. In particular, he showed that a particular configuration, the so-called *Birkhoff diamond* of four pentagons (shown in Figure 4), is reducible; this important configuration was once described as enjoying *as much renown in graph theory as the Kohinoor diamond does in fictional criminal mysteries*.

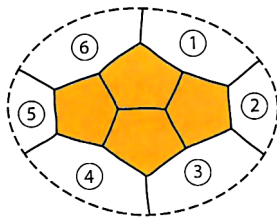


FIGURE 4. The Birkhoff diamond.

To prove that the Birkhoff diamond is reducible, we suppose that we have a minimal counter-example that contains it. Removing the diamond yields a new map with fewer countries which can be colored with four colors. We now try to extend this coloring to the pentagons in the diamond.

To do so, we look at all the possible ways of coloring the ring of six countries surrounding the diamond with the colors red (r), blue (b), green (g), and yellow (y). It turns out that there are essentially thirty-one different colorings of the countries 1 to 6, sixteen of which (such as $rgrbrg$) can be extended directly to the countries of the diamond—these are called *good colorings*—while all the others (such as $rgbrgy$) can be converted into good colorings by suitable Kempe-interchanges of color. Thus, all 31 colorings of the surrounding ring can be extended to the Birkhoff diamond, which is therefore reducible.

After this the flood-gates were open, and several mathematicians on both sides of the Atlantic began to develop Birkhoff’s ideas, producing reducible configurations galore. In particular, Philip Franklin of Princeton proved that each of the following configurations is reducible and so cannot appear in a minimal counter-example:

a pentagon in contact with three pentagons and a hexagon;
a pentagon surrounded by two pentagons and three hexagons;
a hexagon surrounded by four pentagons and two hexagons;
and any n -sided polygon in contact with $n - 1$ pentagons (such as an octagon in contact with seven pentagons).

By applying a counting argument derived from Euler's polyhedron formula, he deduced that *every map with up to 25 countries can be colored with four colors*, and therefore that *a minimal counter-example must have at least 26 countries*.

Later mathematicians continued this work, obtaining further reducible configurations and using counting arguments to prove the four-color theorem for maps with more countries. By the middle of the 20th century it was known that *all maps with up to 35 countries can be colored with four colors*—but there was still a long way to go.

But much of this work on the four-color problem was still piecemeal. Attempts to find unavoidable sets and reducible configurations were largely independent of each other, and a coordinated search for the holy grail—*an unavoidable set of reducible configurations*—had not been forthcoming. The credit for advocating such a search belongs to Heinrich Heesch, whose contributions would pave the way for the eventual solution of the problem by Appel and Haken in the 1970s.

2. Enter Heesch and Haken

Heesch's early career in mathematics was spectacular. After obtaining a doctorate in mathematics from the University of Zürich for a dissertation on symmetry in three-dimensional geometry, he proceeded to Göttingen, the home of such distinguished mathematicians as David Hilbert, Richard Courant, and Hermann Weyl, and became assistant to Weyl, studying the geometry of crystals. While in Göttingen, Heesch came across the 23 celebrated 'Hilbert problems' posed by Hilbert at the International Congress of Mathematicians in Paris in 1900. The four-color problem was not one of these, but the 'regular parquet problem', involving tilings of the plane, was part of Hilbert's Problem 18. Heesch solved this problem in 1932 by constructing tilings that cover the plane according to the rules laid down in the problem.

The mid-1930s were difficult years for Heesch, as for many other German mathematicians. Objecting to the Nazi work camps set up for those desiring to be professors, he soon found himself without university employment. For twenty years he taught mathematics and music in various schools and worked on industrial problems that involved tiling patterns, while continuing with his mathematical researches. He eventually secured a teaching position at the Technical University of Hanover.

Heesch's interest in the four-color problem dates from around 1935. His friend Ernst Witt believed that he had found a proof and they went to show this to Courant. But Courant was about to embark on a train journey to Berlin, so Heesch and Witt bought train tickets too and accompanied him on his journey. Failing to convince Courant that the proof was correct they caught the train back to Göttingen: on the return journey, Heesch found an error in Witt's proof.

Unavoidable sets were still rather scarce, and in order to produce them Heesch invented his *method of discharging*, first published in 1969; the term *discharging* is due to Haken. The general approach is as follows. To prove that a given set of configurations is an unavoidable set, we assume the opposite—that there is a cubic map containing none of them. We next assign to each country a number—like an *electrical charge*—so that *the total charge on the countries of the map is positive*. We then try to move these charges around the map in such a way that no charge is created or destroyed, and so that each individual country ends up with zero or negative charge (this is called *discharging the map*). Then *the total charge on the map cannot be positive*. This contradiction proves that the given set of configurations is an unavoidable set.

To illustrate the method of discharging we prove Wernicke's result that the configurations in Figure 3 form an unavoidable set, and to do so we assume that we have a cubic map that contains none of them and obtain a contradiction. Our assumption tells us that no pentagon can be adjacent to a digon, triangle, or quadrilateral, or to another pentagon or a hexagon, so *each pentagon can adjoin only countries bounded by seven edges or more*.

We now assign the electrical charges. To each country with k boundary lines, we assign a charge of $6 - k$, so that each pentagon ($k = 5$) receives a charge of 1, each hexagon ($k = 6$) receives zero charge, each heptagon ($k = 7$) receives a charge of -1 , and so on. So if the map has C_5 pentagons, C_6 hexagons, C_7 heptagons, etc., then the total charge on the map is

$$\begin{aligned}
 (*) \quad & (1 \times C_5) + (0 \times C_6) + (-1 \times C_7) + (-2 \times C_8) \\
 & \quad + (-3 \times C_9) + (-4 \times C_{10}) + \cdots \\
 & = C_5 - C_7 - 2C_8 - 3C_9 - 4C_{10} - \cdots .
 \end{aligned}$$

But one can easily prove from Euler's polyhedron formula that, for a cubic map with C_2 digons, C_3 triangles, C_4 squares, etc.,

$$4C_2 + 3C_3 + 2C_4 + C_5 - C_7 - 2C_8 - 3C_9 - 4C_{10} - \cdots = 12,$$

and so, since our map has no digons, triangles, or quadrilaterals ($C_2 = C_3 = C_4 = 0$),

$$(*) \quad C_5 - C_7 - 2C_8 - 3C_9 - 4C_{10} - \cdots = 12.$$

Comparing the two starred results, we see that *the total charge on the map is 12*, which is positive.

We now move the charges around the map, transferring *one-fifth* of a unit of charge from each pentagon to each of its five negatively charged neighbors: recall that each of these neighbors has seven or more sides. Then the total charge on the map remains 12, but each pentagon now has zero charge and each hexagon still has zero charge.

What happens to a heptagon? If a heptagon with initial charge -1 receives enough contributions of $\frac{1}{5}$ to acquire positive charge, then it must have at least six neighboring pentagons—but two of these pentagons would then have to be adjacent, which is disallowed; thus, after the discharging, each heptagon retains negative charge. In the same way we can prove that each octagon, nonagon, decagon, etc., retains negative charge.

Thus, each country ends up with zero or negative charge, contradicting our assertion that the total charge on the map is 12. This contradiction proves that every cubic map must contain at least one of Wernicke's configurations of countries, which therefore form an unavoidable set.

By modifying the above method of discharging we can show that many sets of configurations are unavoidable, but the details of the discharging process will vary from situation to situation: in other cases it may be more advantageous to transfer one-fourth or one-third of a unit, or to distribute the charge on each pentagon equally to all of its neighbors (as we see later). As the century progressed, unavoidable sets with thousands of configurations were constructed. To deal with such enormous sets it proved necessary to keep on modifying the discharging process until it could deal with every possible case that arose.

Heesch also contributed to the theory of reducible configurations, obtaining many thousands of them. When discussing the Birkhoff diamond, we saw that all thirty-one possible colorings of the ring of countries surrounding the diamond are all good colorings, or can be converted to good colorings by a succession of Kempe color-interchanges. Heesch introduced the term *D-reducible* for such configurations of countries.

In fact, we do not have to consider all thirty-one colorings. If we modify the map by removing five of the boundary lines, as shown in Figure 5, then we obtain a new map with fewer countries which can therefore be colored with four colors. Heesch used the term *C-reducible* for those configurations that can be proved reducible after they have been modified in some way. The concepts of *D-reducible* and *C-reducible configuration* will feature again later.

In the late 1940s Heesch presented his findings publicly for the first time at the Universities of Hamburg and his home town of Kiel. He had increasingly come to believe that the right approach to the four-color problem was to seek an unavoidable set of reducible configurations. In his lectures Heesch

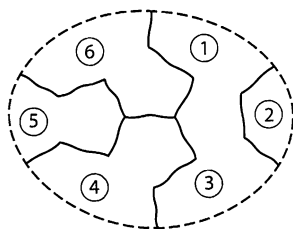


FIGURE 5. Modifying a map.

expounded on his belief that there existed such a set and that its configurations would not be particularly large, but that there would be a great number of them.

One student who attended Heesch's 1948 lectures at the University of Kiel was the young Wolfgang Haken, who had entered the University as its youngest student and was studying mathematics, philosophy, and physics. Haken still recalls hearing Heesch's lecture, much of which he did not understand at the time, but remembers Heesch mentioning that there might be some ten thousand cases to be investigated, that five hundred configurations had already been checked at an average rate of about one per day, and that he seemed optimistic about dealing with the remaining nine-and-a-half thousand cases.

Among the most stimulating series of lectures that Haken attended at Kiel were those on topology by the only mathematics professor, Karl Weise, who had himself worked on the four-color problem. In these lectures Weise described three long-standing problems that remained unsolved at that time: the *knot problem* (determining whether a given tangle of string in three dimensions contains a knot), the now-solved *Poincaré conjecture* (concerning spheres in four-dimensional space), and the *four-color problem*. Haken was to become fully involved with all three problems, in each case using 'very elementary means on which everyone else had given up'. The first problem he solved completely—a great achievement—and announced his results at the 1954 International Congress of Mathematicians in Amsterdam. The second he attacked vigorously, reducing it to 200 particular cases, 198 of which he systematically and successfully managed to deal with, but (frustratingly) not the final two. The third he battled with for several years, as we shall see.

Haken's work on the knot problem so impressed Bill Boone, a logician at the University of Illinois at Urbana-Champaign, that Haken was invited to Illinois as a visiting professor. Shortly after arriving there, he gave several lectures on his researches into the knot problem. His painstaking approach to solving mathematical problems stimulated one of his colleagues to observe:

Mathematicians usually know when they have gotten too deep into the forest to proceed any further. That is the time Haken takes out his penknife and cuts down the trees one at a time.

The 1960s proved to be an exciting time for map coloring. In 1967 the first major book devoted exclusively to map coloring, Oystein Ore's authoritative and influential *The Four-Color Problem* [22], was published, and in 1968 Ore and his research student Joel Stemple extended the earlier methods of Franklin and others to prove that all maps with up to 40 countries can be colored with four colors: their proof involved the consideration of so many special cases that the full details could not be published, but had to be deposited with the Mathematics Department Library at Yale University.

In 1967 Haken contacted Heesch. With his experience of the Poincaré problem, in which just two out of two hundred cases had failed to work, Haken feared that the same thing might have happened to Heesch with his ten thousand configurations, and that Heesch might have given up—but no: Heesch was still working on the problem. By this time, Heesch had invented the method of discharging and had discovered thousands of reducible configurations. But although he lectured widely about his work, it remained unpublished until 1969 when he produced a paperback in German containing the method of discharging and his many other contributions.

Heesch's aim was to 'systematize' the ideas in Birkhoff's 1913 paper by developing an approach for generating reducible configurations, looking at both D -reducible and C -reducible ones. If a configuration was not D -reducible, Heesch could frequently see how to modify it so as to determine whether it was C -reducible. By this means he could restrict his attention to a smaller number of possible colorings, as happened for the Birkhoff diamond.

Over the years Heesch had developed an uncanny knack of being able to look at a configuration and saying whether it was reducible, with at least 80 per cent accuracy. As Haken remarked:

What fascinated me most was that Heesch looked at the configuration, and he either said, 'No, there is no chance. That cannot be reducible', or he said, 'But this one: that is certainly reducible'.

Haken invited Heesch to the University of Illinois to give a lecture, and raised the question of whether computers might be helpful in the examination of large numbers of configurations. Heesch had already had the same idea, and in the mid-1960s had obtained the help of Karl Dürre, a mathematics graduate from Hanover who had become a schoolteacher. Dürre developed a method for testing D -reducibility that was sufficiently routine and algorithmic to be implemented on a computer, even though it might take a long time. By November 1965, using Algol 60 on the University of Hanover's CDC 1604A computer, Dürre confirmed that the Birkhoff diamond is D -reducible and ascertained the D -reducibility of many more configurations of increasing complexity.

The complexity of a configuration can be measured by its *ring-size*, the number of countries surrounding the configuration. As we saw earlier, the Birkhoff diamond has ring-size 6, and there are 31 essentially different colorings of the countries surrounding it. Unfortunately, the number of different colorings increases rapidly as the ring-size increases, as we can see from the following table:

Ring-size	6	7	8	9	10	11	12	13	14
Colorings	31	91	274	820	2461	7381	22144	64430	199291

So, for a configuration of eight countries with ring-size 14, we have almost two hundred thousand essentially different colorings to consider. Moreover, informal calculations first indicated that a solution to the four-color problem might involve configurations with ring-size 18, where the number of colorings exceeds 16 million.

Heesch and Dürre found that the time taken to analyse a configuration increases rapidly as its ring-size increases. On their machine a typical configuration with ring-size 12 might take six hours to analyse, while some configurations with ring-size 13 took anything between 16 and 61 hours, and those of ring-size 14 were beyond reach. They estimated that to verify all ten thousand cases might involve up to fifty thousand hours of computer time; this was not a realistic proposition on any computer then available.

By this time it was clear that any solution of the four-color problem along these lines would have to be complicated. In the 1960s E. F. Moore of the University of Wisconsin developed a remarkable knack for constructing maps that contain no known reducible configurations; he thereby hoped to find a map that required five colors. One of his maps contained no reducible configurations with ring-size 11 or less, so any unavoidable set of reducible configurations necessarily contains at least one reducible configuration of ring-size 12 or greater (see [25]).

Although it looked as though configurations of ring-size up to 18 might be necessary, in the event the Appel–Haken solution involved only configurations with ring-size 14 and less. Had it required any larger configurations they would almost certainly not have been able to complete their solution when they did.

By this time, most workers on the four-color problem were using the ‘dual formulation’, introduced by Kempe, of coloring the points of the corresponding graph or linkage. In particular, Heesch devised a useful notation, which soon became widely used, for representing each point by an appropriate ‘blob’ so that it can be easily distinguished. Four of Heesch’s symbols and an example are shown in Figure 6.

It soon became clear that the Hanover computer was insufficiently powerful to carry out the work required of it. Haken tried to help Heesch and Dürre

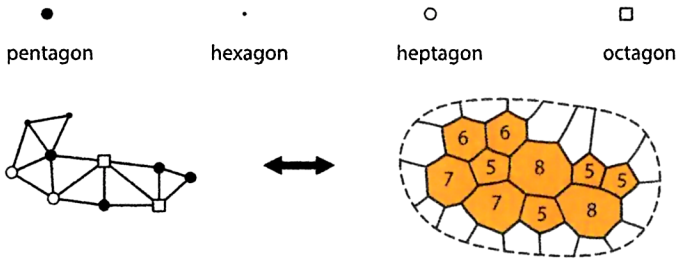


FIGURE 6. Heesch's symbols for points.

by bidding to the University of Illinois for time on a new supercomputer whose construction was nearing completion, but it was not yet ready for use. Eventually, they were referred to Heesch's friend Yoshio Shimamoto, Director of the Computer Centre at the Atomic Energy Commission's Brookhaven laboratory in Upton, Long Island, where there was a Cray Control Data 6600 machine, the most powerful machine of its day.

Shimamoto had always been an enthusiast for the four-color problem and was strongly supportive of Heesch's approach to the problem. He generously invited Heesch and Dürre to visit Brookhaven and continue their reducibility testing on the Cray computer (see Figure 7). Heesch paid two extended visits, including a stay of one year, while Dürre visited for almost two years.

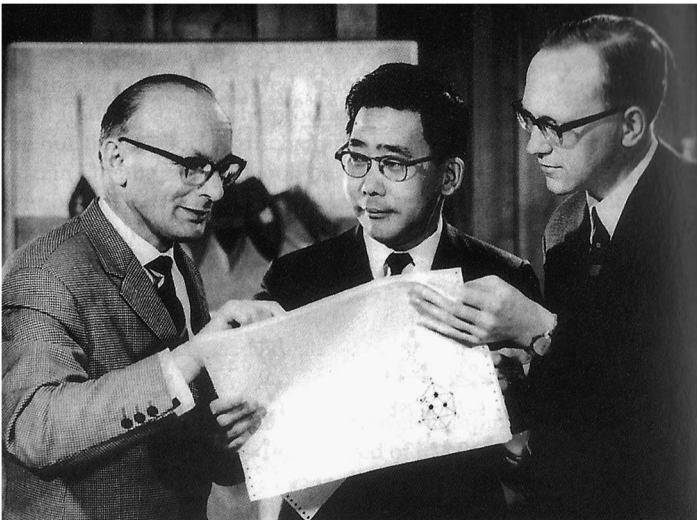


FIGURE 7. Heesch, Shimamoto, and Dürre at Brookhaven National Laboratory.

At Brookhaven progress was quickly made. A configuration with ring-size 13 was not excessively large for the Cray computer, and those with ring-size 14 could be tested for the first time. Eventually, Heesch and Dürre were able to confirm the D -reducibility of more than a thousand configurations.

Meanwhile, Shimamoto was carrying on his own researches into the problem. Unlikely as it may seem, he showed that if he could find a single configuration with certain particular properties, and if this configuration were D -reducible, then the four-color theorem would follow: the whole of the proof would depend on a single configuration! Suddenly, on 30 September 1971, he found what he was seeking, a configuration with ring-size 14 that became known as the *Shimamoto horseshoe* (see Figure 8): if this horseshoe were D -reducible, then the four-color problem would be solved.

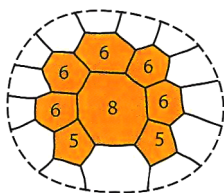


FIGURE 8. The Shimamoto horseshoe.

Coincidentally Heesch and Haken were visiting Brookhaven on that very day. Heesch recognised the horseshoe as one of his list of D -reducible configurations, but Shimamoto was cautious and asked for it to be rechecked for D -reducibility. Excitement ran riot: rumors about Shimamoto's solution were beginning to encircle the world. On the day following the horseshoe's discovery, Haken was due to visit the Princeton Institute for Advanced Study before returning to Illinois, and he asked Shimamoto's permission to tell people about the result. According to Haken, Shimamoto said 'yes', while Shimamoto later denied doing so and became somewhat annoyed about Haken's pronouncement, complaining that:

I am getting a deluge of questions from many people and I am now wondering about the wisdom of having you talk at Princeton and Urbana before a clean manuscript had been prepared.

In the event, things did not work out as hoped. The super-fast Cray computer was allowed to run for a whole weekend, and after grinding on for twenty-six excruciating hours, it confirmed that the horseshoe was *not* D -reducible after all. There was great disappointment all round.

It seemed as though the four-color problem had again come to a dead end. But Heesch's approach was beginning to pay dividends, and in the hands of Haken and Appel would yield the desired goal within the next five years.

In 1970 Heesch sent Haken the results of a new discharging experiment, in which the positive charge on each pentagon was now to be distributed *equally* to all neighboring countries with negative charge. The result of this experiment, if applied to a general map, would yield about 8900 ‘bad’ configurations extending up to ring-size 18, in which some countries would still have positive charge. This approach, the *finitization of the four-color problem*, reduced the problem to the consideration of these 8900 configurations, which Heesch proposed to work through one at a time.

Haken, however, was highly pessimistic about being able to deal with such a large number of configurations, especially since some of them were fairly large. By this time, it had become a simple matter to test configurations of ring-size 11, since there are only 7381 different colorings of the surrounding ring of eleven countries, but each time the ring-size increased by 1 the corresponding computer time seemed to expand by a factor of four, with a corresponding increase in storage. Recalling that the difficult horseshoe configuration with ring-size 14 had taken twenty-six hours to run, Appel and Haken later observed:

Even if the average time required for examining fourteen-ring configurations was only 25 minutes, the factor of four to the fourth power in passing from fourteen- to eighteen-rings would imply that the average eighteen-ring configuration would require over 100 hours of time and much more storage than was available on any existing computer. If there were a thousand configurations of ring-size 18, then the whole process would take over 100,000 hours, or about eleven years, on a fast computer.

For some time, Haken had felt that the complexity of the problem would be substantially simplified if one could find a better discharging method. By restricting his attention to maps without hexagons or heptagons, he succeeded in finding a much simpler procedure. Encouraged by this, he then proceeded to general maps, and communicated his findings to Heesch.

Impressed by these, Heesch invited Haken to collaborate with him, and in 1971 sent him several of his unpublished results concerning reducible configurations. These included three ‘obstacles to reducibility’ whose presence seemed to prevent a configuration from being reducible. Although it has never been *proved* that a configuration containing one of them cannot be reducible, no such reducible configuration has ever been found, and so it seemed sensible to exclude them from consideration. The three obstacles are shown in Figure 9: a *4-legger country* adjoining four consecutive countries of the surrounding ring, a *3-legger articulation country* adjoining three surrounding countries that are not all adjacent, and a *hanging 5-5 pair* of adjacent pentagons that adjoin a single country inside the surrounding ring.

But by this time, Haken was beginning to change his approach to the problem. Unlike everyone else whose objective seemed to be to collect reducible

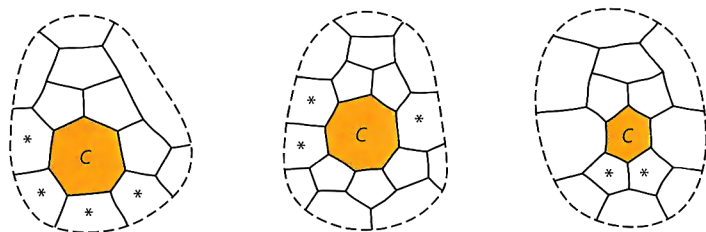


FIGURE 9. Three obstructions to reducibility.

configurations by the hundreds and then package them up into an unavoidable set, Haken's primary motivation, later developed with Appel, was to aim directly for an unavoidable set. In order to avoid wasting time checking configurations that would eventually be of no interest, this set was to contain only configurations that were *likely to be reducible*—in particular, they should contain none of the reduction obstacles. Any configurations that subsequently proved not to be reducible could then be dealt with individually. Haken also considered it inappropriate to spend expensive computer time checking the reducibility of configurations that were unlikely to appear in the eventual unavoidable set. As he later commented:

If you want to improve something, you should not improve that part which is already in good shape. The weakest point is always the one you should improve. This is a very simple answer to why we got it and not the others.

So from this point on, Haken headed off in a different direction from everyone else, concentrating on the unavoidable set and leaving details of the reducibility until much later. Heesch, while initially sympathetic to these ideas, came to reject the idea of a 'likely-to-be-reducible' configuration, and their cooperation was doubtless further damaged by the horseshoe episode. Meanwhile, Heesch was having immense problems obtaining the necessary funding for the use of a powerful computer. As a relatively minor figure in the academic hierarchy he had little 'clout', and his proposals for funding did not receive the consideration they deserved.

3. Enter Appel

With little knowledge of computing and with the Brookhaven machine no longer available, Wolfgang Haken also considered giving up the problem for a few years until more powerful computers had become available to deal with the massive calculations that would clearly be necessary. Computer 'experts' had told him that his ideas could not be programmed, and in a lecture in Illinois on the horseshoe episode he announced:

The computer experts have told me that it is not possible to go on like that. But right now I'm quitting. I consider this to be the point to which and not beyond one can go without a computer.

Attending his lecture was Kenneth Appel, who had received his doctorate from the University of Michigan for a dissertation on the application of mathematical logic to problems in algebra. An experienced computer programmer, he had worked for Douglas Aircraft and for the Institute for Defense Analysis at Princeton before settling at the University of Illinois. His computer experience would prove invaluable for solving the four-color problem.

After the lecture Appel told Haken that he considered the computer experts to be talking nonsense, and offered to work on the problem of implementing the discharging procedures:

I don't know of anything involving computers that can't be done; some things just take longer than others. Why don't we take a shot at it?

Coincidentally, Thomas Osgood, a research student of Haken's, had just submitted his doctoral thesis on solving the four-color problem for maps containing only pentagons, hexagons, and octagons. Since Appel was a member of Osgood's thesis examination panel, the collaboration seemed beneficial for all concerned.

Haken was delighted to accept Appel's offer to take care of the computing side of things. They decided to concentrate their search on unavoidable sets, without taking time to check the configurations for reducibility, and focused on *geographically good* configurations that contain neither of Heesch's first two obstacles to reducibility: such configurations can easily be identified by a computer, or indeed by hand. They would then check their configurations for reducibility when the entire set had been constructed. But when they started work in late 1972 Haken and Appel had no clear idea of where they were heading. As Appel recalled:

We started with certain ideas and kept discovering that we had to become more sophisticated to avoid being swamped by useless or repetitive data.

Their first exploratory computer runs already provided much useful information since geographically good configurations of reasonable size (with ring-size 16 or less) tended to appear close to most countries with ultimately positive charge. Since the computer output was enormous, with many configurations appearing many times, they needed to keep such duplications under control if the eventual list were to be manageable. Fortunately, since the computer program had run in just a few hours, they could experiment as frequently as they needed to.

The necessary changes to the program were easily implemented and the second set of runs, a month later, was a definite improvement. The output was much reduced in thickness, and would eventually come down to a fraction of an inch. The major problems had been overcome and lesser ones would gradually emerge. From then on, every two weeks or so, they modified the discharging

algorithm or the computer program so that the program grew while the output shrank. This two-way dialogue with the computer continued, as problems were sorted out and new ones arose. Within six months of experimenting and improving their procedures, they realised that their method for producing a finite unavoidable set of geographically good configurations in a reasonable amount of time was indeed feasible.

At this stage they changed tack. They decided to prove *theoretically* that their method would provide such an unavoidable set. This proved to be much more complicated than they had expected, taking them more than a year to complete. But the outcome, in late 1974, was a lengthy proof that *an unavoidable set of geographically good configurations exists*, together with an achievable method for constructing such a set.

Their next task was to determine how complicated this process would be. They decided to experiment with the particular case of maps that do not contain two adjacent pentagons. This was much simpler than the general case, and produced a set of 47 geographically good reducible configurations of ring-size 16 or less [3], as shown in Figure 10. Based on their experiment, they estimated that the general problem might be about fifty times as bad as this, and on this basis they decided to proceed: in the event, their estimate turned out to be over-optimistic.

In early 1975 they introduced the third of Heesch's obstacles, the hanging 5-5 pair. This inevitably involved changes in procedure, but was carried out successfully with only a doubling in the size of the unavoidable set. They also programmed the computer to search for sets of configurations with relatively small ring-size.

At this stage, the computer started to think for itself, as Appel and Haken later recalled:

It would work out complex strategies based on all the tricks it had been "taught" and often these approaches were far more clever than those we would have tried. Thus it began to teach us things about how to proceed that we never expected. In a sense it had surpassed its creators in some aspects of the "intellectual" as well as the mechanical parts of the task.

As soon as it seemed probable that they would be able to find an obstacle-free unavoidable set of configurations, which were therefore likely to be reducible, they started the massive detailed check for reducibility. Inevitably, some 'rogue' reducible configurations would appear in the list, but the hope was that they would be relatively few in number. Also, with configurations that might go up to ring-size 16, and others that might cause trouble in other ways, they were expecting to have to devise some clever short cuts.

In the middle of 1974, knowing that they would need help with the reducibility programs, Appel asked whether any graduate student would be interested in writing a thesis in the area of programming, and coincidentally John Koch was looking around for a suitable topic. They set up Koch's thesis

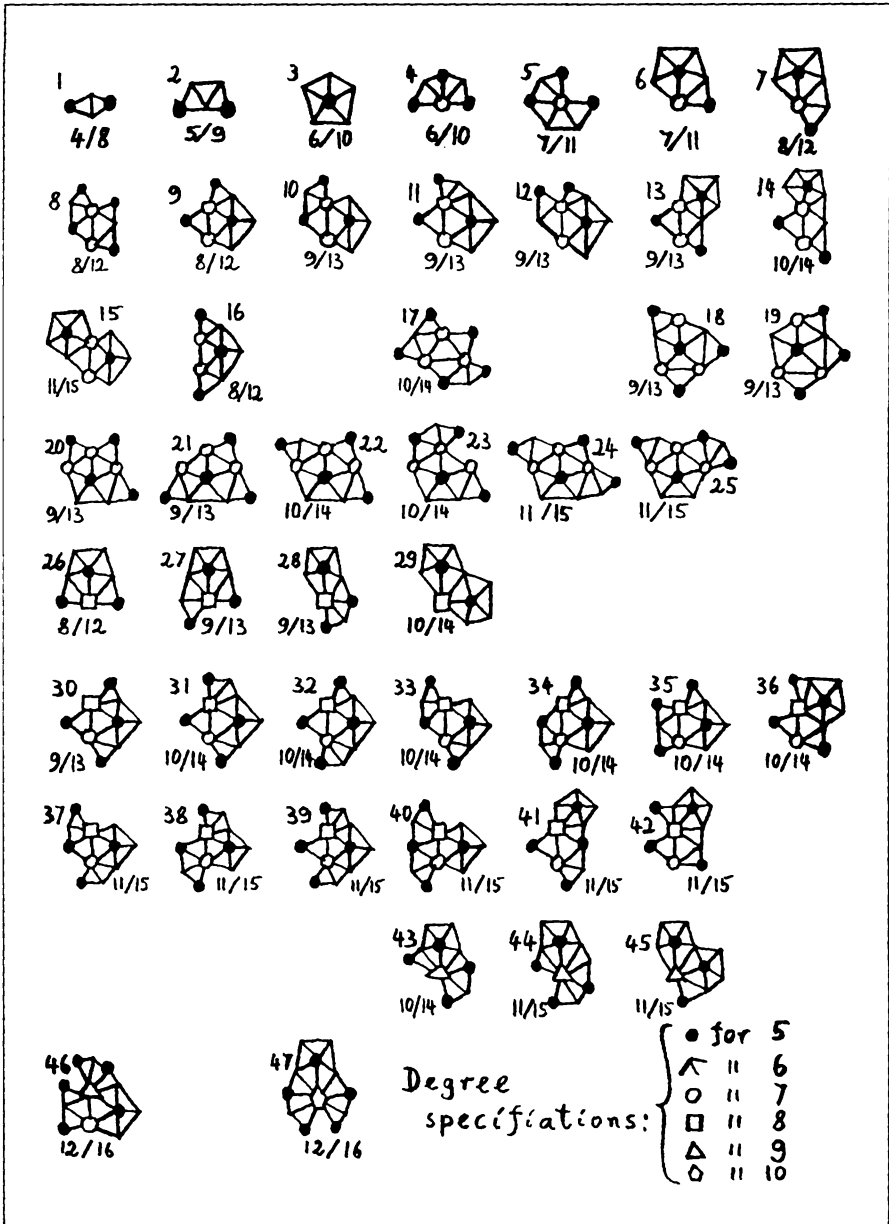


FIGURE 10. Haken and Appel's geographically good configurations.

project so that its completion would not depend on the success of their attack on the four-color problem.

Koch was set to work on the C -reducibility of configurations of ring-size 11. Appel and Haken were particularly interested in two types of modification that were relatively easy to implement, and Koch discovered that 90 per cent of his configurations were of these types. Agreeing that little would be gained by including the other ten per cent, which would have been difficult to program, they restricted their attention to the simple modifications. Koch then devised an elegant and highly efficient method for testing the C -reducibility of these configurations of ring-size 11, which Appel subsequently extended to configurations of ring-sizes 12, 13, and 14.

Towards the end of 1975, their work on the discharging method had run into trouble: structural changes would be necessary, requiring significant modifications to the program. The problem was that, while trying to disperse the positive charge on each pentagon to its immediate neighbors, they regularly came up against 'barriers of hexagons', all with zero charge. Haken asked himself why the positive charge on the pentagons should not be permitted to 'jump over' these hexagons. This would give a more efficient process, but would leave them with a dilemma: *should they rewrite their program from scratch or should they adopt a more ad hoc approach?* Since the former would have been a horrendous task, they opted for the latter, implementing the final version of the discharging process by hand. This inevitably involved much work, but gave them extra flexibility by enabling them to make minor changes whenever necessary. Indeed, they found so many improvements that they could restrict all of their configurations to ring-size 14 or less.

Throughout the first half of 1976, Haken and Appel worked on the final details of the discharging procedure in order to produce the desired unavoidable set of reducible configurations. To do this, they sought out 'problem configurations' that would necessitate further changes to the discharging procedure. Whenever they found such a configuration, they immediately tested it for reducibility—this could usually be done fairly quickly. In this way the reducibility testing by computer could keep pace with the manual construction of the discharging procedure. In the event, the final process involved 487 discharging rules, requiring the investigation by hand of about 10,000 neighborhoods of countries with positive charge and the reducibility testing by computer of some 2000 configurations.

Because the reducibility of an awkward configuration could sometimes take a long time to check, and with memories of the non-reducible Shimamoto horseshoe, they found it convenient to impose on each configuration an artificial limit of ninety minutes checking time on an IBM 370-158 computer, or thirty minutes on an IBM 370-168 computer. If a configuration could not be proved reducible in this time, it was abandoned and replaced by other configurations: finding such alternative configurations was usually straightforward.

By way of comparison, they estimated that checking the computer output of one of the more difficult configurations by hand would take someone working a 40-hour week about five years in total.

The last few months were indeed extremely heavy on computer time, but here Appel, Haken, and Koch were fortunate. Probably no other institution would have given them twelve hundred hours of computer time, but the University of Illinois's computer centre was very supportive throughout. Use was also made of the computer in the Chicago branch of the University: programs were sent to them overnight and the returns were ready by the next morning.

In March 1976 a powerful new computer was bought by the University's administrators. As Haken recalls, local politicians started to ask: 'Why does this administration need a larger computer than the scientists?', and the administrators had to promise: 'OK, half the time, or whenever the thing is not needed by us, the scientists can use it'. Since Appel seemed to be the only scientist who could get the machine running properly, he was initially almost its only user, with a useful fifty hours of computer time over the Easter vacation. Everyone was happy: the administrators could claim a more balanced loading of the system over a 24-hour period, while Appel got all the computer time he needed.

In the event, this new computer proved to be so powerful that everything proceeded far more quickly than they had expected, saving them, by their own estimation, a full two years on the reducibility testing. Meanwhile, with the help of Haken's daughter Dorothea, they spent months of exhausting and stressful effort working through the two thousand or so configurations that would eventually form the unavoidable set:

one of us wrote a section, and then the second one was reading it over, checking it for mistakes, and bringing the errors to the attention of the one who wrote it, and then the third person read it a third time. Reading it over was more strenuous than writing it, because it went faster, and it was more exhausting. It was a terrific amount of work for three people.

Suddenly by late June, almost before they realised what was happening, the entire job was finished: the Hakens had completed the construction of the unavoidable set, and within two days Appel was able to test the final configurations for reducibility. Appel celebrated their achievement by placing a notice on the department's blackboard:

Modulo careful checking it appears that four colors suffice.

This phrase, *four colors suffice*, subsequently became the department's postal meter slogan.

All that remained was the detailed checking, which had to be carried out speedily since—not expecting the reducibility testing to advance so well—Appel had arranged a sabbatical visit to France and was due to leave in late July, leaving them just five weeks to finish the job. Although they had not

fully realised it, time was indeed of the essence, as several other map-colorers were within striking distance of solving the problem.

At the University of Waterloo, Ontario, Frank Allaire had the best reducibility methods around: in Haken's words, these were *even better than Heesch's and much better than ours*. Appel and Haken were aware of Allaire's superior methods, and by 1976 Allaire was several months ahead of them in his investigations into reducibility and was expecting to complete his solution within a few months.

Meanwhile, at the University of Rhodesia was a former chemist, Ted Swart, who was making excellent progress on the four-color problem. Swart submitted a paper to the *Journal of Combinatorial Theory* and received a reply from the Editor-in-Chief, Bill Tutte, to the effect that Allaire was working along very similar lines. Allaire and Swart pooled their results and submitted a paper [2] just before Haken and Appel's proof was announced: this described an algorithm for determining the reducibility of configurations and included a full list of all reducible configurations with ring-size 10 or less.

Yet another person who had been developing some powerful new methods for tackling the four-color problem, and who was expected to complete his solution within a year, was a Harvard doctoral student Walter Stromquist (see [27]).

Although no-one was aware just how close everyone else was to completing their solutions, Haken and Appel suspected that it would be too risky to wait for five months, especially if a rumour leaked out that they had almost reached their goal. They guessed that Allaire, with his superior reducibility methods, was some way ahead of them on that score, but confidently believed that their strategy of spending nine-tenths of their time on unavoidable sets and only one-tenth on reducible configurations (while others operated the other way round) still gave them the edge overall.

But they had no time to lose. Drafting in their children to help they immediately set to work. They had high expectations that, while the checking might turn up many typos and the occasional bad configuration that would need replacing, there would be no major disasters. Laurel Appel checked through the 700 pages of detailed working and found about 800 mistakes in total, most of them typos, and all but fifty of these she immediately corrected herself. As Haken later remarked:

Somebody has worked one month full-time and found eight hundred mistakes. And then we needed only five days to repair all that. This looks so stable—it is this incredible stability. . . .

Haken and Appel knew by this stage that they were safe: even if a few configurations turned out not to be reducible after all, there was more than enough self-correction in the whole system for these configurations to be replaceable easily and quickly. It was not possible for a single faulty configuration, if such existed, to destroy the entire edifice. In fact, there was so

much self-correction in the system that they effectively had many thousands of proofs of the four-color theorem, rather than just one!

Armed with this confidence, they went public. On 22 July, 1976, just a few days before Appel was due to leave for France, they formally informed their colleagues and sent out complete preprints to everyone in the field. One of the recipients was Bill Tutte. Two years earlier he had written an article in the *American Scientist* [29] claiming that people who used their approach were real optimists, since the method seemed extremely unlikely to work. But when Tutte heard the news he waxed eloquent, comparing their achievement with the slaying of a fabled Norwegian sea monster:

Wolfgang Haken

Smote the Kraken

One! Two! Three! Four!

Quoth he: 'The monster is no more'.

Haken and Appel were delighted that a mathematician of Tutte's stature should have given his positive support so quickly. Tutte's endorsement would go a long way to setting people's minds at rest, while a lukewarm response could have cast serious doubts about their solution.

They decided to submit the full solution to the *Illinois Journal of Mathematics*. This was partly because a few years earlier they had published a long and turgid article in its pages, and they felt that they owed it to the *Illinois Journal* to give it the kudos for publishing their solution of this famous problem. But the main reason that they wanted it to appear locally was because they wished to be in a position to suggest appropriate referees for their proof. It was clearly in everyone's interest—theirs and the journal's—that a paper such as this should be thoroughly refereed:

We said to the Illinois Journal, "Look, if this thing is wrong, we want to know it as badly as anyone else does, and you're not going to reject it on the grounds of triviality."

They agreed with the Editor that the best people in the world should be selected as referees—Frank Allaire for the reducibility part and Jean Mayer in France for the discharging arguments. In France, Appel spent much of his sabbatical time in Montpellier working painstakingly through the details of the proof with Jean Mayer, one step at a time, dealing with all of Mayer's queries and suggestions as they arose. On his return in December 1976 Appel and Haken set to work refining the proof, conferring with the referees, and preparing the paper and its associated microfiche for publication.

Heinrich Heesch was naturally very distressed that Appel and Haken had got there first—not surprisingly, since their method was essentially due to him and solving the four-color problem had been his goal for more than forty years. But later he became very cooperative, sending Haken his own list of 2669 reducible configurations, together with supporting data on the number of good configurations. Frank Allaire was also bitterly disappointed when

Appel and Haken announced their result, as he himself was nearing a solution. Moreover, he and Ted Swart had been taking a more systematic approach, as opposed to the more hit-and-miss methods of Appel and Haken. But he quickly suppressed his own feelings for the general good, refereeing the reducibility part conscientiously and constructively.

Appel and Haken's paper was a substantial improvement on the rough-and-ready preprint that they had sent out in July 1976. In particular, they discovered that their preprint contained several 'repeats' of configurations and many instances of one configuration appearing inside another, so that in each case they could omit the larger one. By weeding out these superfluous configurations they managed to reduce the original list of 1936 reducible configurations to the 1482 configurations of the published version: they later reduced this number still further to 1405. However, they saw no point in aiming for the 'smallest possible' number if the total computer time was thereby increased. As Appel pointed out:

If one configuration replaces twelve, but that one configuration takes two hours and the twelve take five minutes, that doesn't make sense.

Appel and Haken's solution appeared in two parts in the December 1977 issue of the *Illinois Journal of Mathematics* [7], [12]. Part I, *Discharging*, written by the two of them, outlined the overall strategy of their proof and described their methods of discharging for constructing the unavoidable set. Part II, *Reducibility*, written with John Koch, described the computer implementation and listed the entire unavoidable set of reducible configurations. These were supplemented by a microfiche containing 450 pages of further diagrams and detailed explanations.

Wolfgang Haken and Kenneth Appel had achieved their goal: the four-color theorem was proved.

4. Aftermath

The Appel–Haken proof of the four-color theorem was greeted with great enthusiasm: after 124 years one of the most famous problems in mathematics had finally been solved. But it was also greeted with scepticism, outright rejection, and (more than anything else) great disappointment.

Every August the American Mathematical Society and the Mathematical Association of America used to organise a joint summer meeting. In 1976 it was held at the University of Toronto, and Wolfgang Haken was one of the main speakers. A report on the event by Donald Albers [1] described the scene:

The elegant and old lecture hall was jammed with mathematicians anxious to hear Professor Haken give the proof. It seemed like the perfect setting to announce a great mathematical result. He proceeded to outline clearly the



FIGURE 11. Wolfgang Haken (seated) and Kenneth Appel.

computer-assisted proof that he and his colleagues had devised. At the conclusion of his remarks, I had expected the audience to erupt with a great ovation. Instead, they responded with polite applause!

Reaction to Haken's lecture was certainly mixed. Ted Swart praised the quality of Haken's presentation, recalling that the applause was a little more than merely polite. But this was by no means the general reaction:

Mathematician after mathematician expressed uneasiness with a proof in which a computer played a major role. They were bothered by the fact that more than 1000 hours of computer time had been expended in checking some 100,000 cases and often suggested (hoped?) that there might be an error buried

in the hundreds of pages of computer print outs. Beyond that concern was a hope that a much shorter proof could be found.

The report concluded:

it seems that the computer-assisted work of Appel, Haken, and Koch on the well-known Four-Color Problem may represent a watershed in the history of mathematics. Their work has been remarkably successful in forcing us to ask What is a Proof Today?

It was clear that Appel and Haken's solution of the four-color problem had created a new situation in mathematics. Is it a proof? And if so, how do we know that it is a proof? Concerns about it, and particularly about its use of the computer, continued to rumble on—and still do. Can a proof be considered valid if it cannot be checked by hand?

As mathematical proofs have become longer and computers are used more widely, the whole question of surveyability has become fraught with difficulties. A computer that performs extensive but completely routine tasks may certainly be considered at least as reliable than a human checking by hand a proof that is long and complicated or breaks up into a large number of special cases. In Ted Swart's words:

Human beings get tired, and their attention wanders, and they are all too prone to slips of various kinds. . . Computers do not get tired.

But surveyability has not been the only problem caused by the Appel–Haken proof. Several mathematicians have complained that their proof is not sufficiently 'transparent'. Among these was Ian Stewart, the well-known writer on mathematics, who complained that it did not explain *why* the theorem is true, partly because it was too long for one to grasp all the details, and partly because it seemed to have no structure:

The answer appears as a kind of monstrous coincidence. Why is there an unavoidable set of reducible configurations? The best answer at the present time is: there just is. The proof: here it is, see for yourself. The mathematician's search for hidden structure, his pattern-binding urge, is frustrated.

What everyone did agree on, however, was that the Appel–Haken proof could not be described as beautiful or elegant mathematics. In his celebrated book *A Mathematician's Apology* [17], the celebrated English mathematician G. H. Hardy had once asserted:

there is no permanent place in the world for ugly mathematics.

—and even Appel did not demur from this in respect to his own proof:

there were people who said, "This is terrible mathematics, because mathematics should be clean and elegant", and I would agree. It would be nicer to have clean and elegant proofs.

However, while everyone would like to see a non-computer solution of the four-color theorem, such a solution along the lines of the Appel–Haken proof is almost certainly unattainable—not least because any unavoidable set of reducible configurations must include configurations with ring-size 12 or more,

as we saw earlier. For a non-computer proof, new ideas are needed, and such ideas have not yet been forthcoming.

Following his lecture at the Toronto meeting, Haken spent a lot of time touring the United States, explaining the details of their solution. Appel was doing a similar thing in Europe, giving lectures in France and England. These lectures were informative and convincing to those who attended. But following the appearance of their unorthodox proof, Appel and Haken were sometimes made to feel most unwelcome. The most dramatic occasion occurred when the head of a mathematics department refused to allow them to meet his graduate students on the grounds that:

Since the problem had been taken care of by a totally inappropriate means, no first-rate mathematician would now work any more on it, because he would not be the first one to do it, and therefore a decent proof might be delayed indefinitely. It would certainly require a first-rate mathematician to find a satisfactory proof, and that was now impossible.

They were made to feel that they had done something very wicked, and that the innocent souls of students needed to be protected from their bad influence.

Other publications appeared in 1977. As well as brief reports on their solution in the *Bulletin of the American Mathematical Society* [4] and *Discrete Mathematics* [5], Appel and Haken wrote a more popular article on ‘The solution of the four-color-map problem’ for the readers of *Scientific American* [6]: this is still one of the clearest descriptions of their approach to solving the problem. Another account appeared in [8], and a further research paper in [9]. Around the same time, a second book on the four-color problem was published: *The Four-Color Problem: Assaults and Conquests*, by Thomas Saaty and Paul Kainen [24].

By the early 1980s, rumors were beginning to spread that there was a major error in Appel and Haken’s proof of the four-color theorem. In view of its complexity, many errors might have been expected to appear, but this hardly happened. These rumors probably arose from a mistake (noticed by Ulrich Schmidt, a German engineering student) that took two weeks for Haken to correct, and a drawing error in one configuration, but other than these and a few misprints no significant error has ever been found in their proof.

In 1986 Appel and Haken received a letter from the Editor of the *Mathematical Intelligencer*, who had heard these rumours that something was wrong with the proof and invited them to set the record straight. Welcoming such controversy, they replied:

With this kind offer we cannot but comply, although it is a pity to spike a good rumor. Mathematical rumors do add interest and excitement to the conversation at mathematical meetings. However, the rumors about the Four Color Theorem seem to be based on a misinterpretation of the results of the independent check of details of the proof by U. Schmidt.

The result was ‘The four color proof suffices’, an upbeat article explaining the error and describing their methods in some detail [10]. They followed this in 1989 by their last word on the subject, a hefty tome published by the American Mathematical Society entitled *Every Planar Map is Four Colorable* [11]: this supplied several more details of the proof, proved some related results, corrected all the errors that had arisen, and included a printed version of all their microfiche pages.

In 1994 Neil Robertson, Daniel Sanders, Paul Seymour, and Robin Thomas added an exciting new chapter to the four-color saga. For the previous ten years, members of this group had been obtaining some spectacular results in graph theory, and they now turned to the four-color problem because it was related to other work in which they were involved.

They were concerned that Appel and Haken’s proof was still not fully accepted, asserting that there remained some doubt as to its validity. The main reason was not that the proof involved a computer, even though a full check of the reducibility part would have involved inputting 1482 configurations by hand into the computer and a lot of programming to test each one for reducibility. It was the unavoidability arguments that caused them most concern: the discharging algorithm that Appel and Haken had carried out by hand was very complicated and no-one had ever made an independent check of it all.

After a week of working through the details of the Appel–Haken proof they gave up. They decided that it would be much more fun, and much more instructive, to discover their own proof, using the same general approach that Appel and Haken had used. It took them a whole year to complete the work, but their eventual proof was shorter and more robust than its predecessor [23].

The unavoidable set that they obtained contains only 633 reducible configurations—they could have reduced it further, but only at the expense of more computer time. Moreover, their discharging procedure for proving unavoidability involved only 32 discharging rules, instead of the 487 used by Appel and Haken. Because they considered computer proofs of long and complicated results to be *more* reliable than hand-checked ones, they used the computer for both the unavoidability and reducibility parts of the proof. What is more, all the steps in their proof can be externally verified by anyone on their home computer in just a few hours.

In August 1998, shortly after their new proof appeared, one of its authors, Robin Thomas, wrote an interesting article giving ‘An update on the four-color theorem’ [28] in the *Notices of the American Mathematical Society*, outlining its main ideas and emphasizing the theorem’s importance by describing some results from other branches of mathematics (the divisibility of integers, the algebra of three-dimensional vectors, and matrices and tensors) that surprisingly turn out to be equivalent to it.

The improved methods employed by Robertson and his team led to a more efficient algorithm for coloring maps. Appel and Haken's approach had given rise to a *quartic algorithm* for map coloring: this means that if a map has n countries, then the running time required for coloring the map is proportional to n^4 . The more efficient methods of Robertson's co-workers led to a *quadratic algorithm*, for which the running time is proportional to n^2 . The resulting improvements are apparent from the following table which compares the running times for various values of n for a computer performing a million calculations per second:

	$n = 10$	$n = 100$	$n = 1000$
n^2	0.0001 seconds	0.01 seconds	1 second
n^4	0.01 seconds	1 minute, 40 seconds	over $11\frac{1}{2}$ days

More recently, good progress has been made in a number of directions. We mention two of these here.

D-reducible configurations. We recall that a configuration is D -reducible if every coloring of the surrounding ring of countries can be extended to the countries of the configuration, directly or after some Kempe-interchanges of color, and that it is C -reducible if it can be proved to be reducible after it has been modified in some way. Since D -reducible configurations are much easier to deal with, we may ask whether there is a proof of the four-color theorem that involves *only* D -reducible configurations.

This question was answered in the affirmative by John P. Steinberger [26] in 2008. Inevitably it involved more configurations and more discharging algorithms than had earlier proofs, and the author needed to consider configurations with larger ring-size. The following table summarizes the differences:

	No. of configurations	No. of discharging rules	Largest ring-size
Appel and Haken	1482	487	14
Robertson <i>et al.</i>	633	32	14
Steinberger	2832	42	16

Checking the proof. As we have seen, the correctness of large computer-assisted proofs is difficult to verify. A major breakthrough occurred in September 2004 when the French computer scientist Georges Gonthier provided a fully machine-checked proof of the four-color theorem [16]—this was a formal language implementation and machine verification of the approach of Robertson and his team. Gonthier's first step was to provide an axiomatic formulation of the terms 'map' and 'four-colorable'. A respected and widely-used machine 'proof-checker' called *Coq* then verified some sixty thousand

lines of formal language proof before pronouncing the proof valid. At last the doubters could consider the four-color problem proved!

So now that the four-colour problem has been proved, what else is there for map-colorers to do? This question was asked back in 1978 in an article by Bill Tutte [30]:

I imagine one of them outgribing in despair, crying “What shall I do now?” To which the proper answer is “Be of good cheer. You can continue in the same general line of research.”

For the four-color theorem is by no means the end of the line—it is indeed more of a beginning. In spite of its enormous difficulty, it is just one special instance of some much harder problems that develop the ideas of the four-color theorem in new and exciting directions (such as finding proofs for Hadwiger’s conjecture and the five-flow conjecture), and on these problems good progress is already being made.

With these thoughts of the future in our minds, we leave our last poetic musings to Bill Tutte:

*The Four Colour Theorem is the tip of the iceberg,
the thin end of the wedge
and the first cuckoo of Spring.*



FIGURE 12. Wolfgang Haken, the author, and Kenneth Appel in 2002.

REFERENCES

Much of this article is derived from the author's book, where further details can be found:

Robin Wilson, *Four Colors Suffice* (Revised Color Edition), Princeton Science Library, Princeton University Press, 2013.

Much material relating to Heesch can be found in the biography: *Hans-Günther Birkhoff, Heinrich Heesch: Kristallgeometrie, Parkettierungen, Vierfarbenforschung*, Birkhäuser, Basel, 1988.

Many of the quotations are taken from unpublished interviews by Donald MacKenzie with Appel, Haken, Swart, and others (see also [21]).

- [1] D. Albers, *Polite applause for a proof of one of the great conjectures of mathematics: What is a proof today?* Two-Year Coll. Math. J. **12** (1981), 82.
- [2] F. Allaire and E. Swart, *A systematic approach to the determination of reducible configurations*, J. Combin. Theory Ser. B **25** (1978), 339–362. MR 0516267
- [3] K. Appel and W. Haken, *The existence of unavoidable sets of geographically good configurations*, Illinois J. Math. **20** (1976), 218–297. MR 0392641
- [4] K. Appel and W. Haken, *Every planar map is four colorable*, Bull. Amer. Math. Soc. **82** (1976), 711–712. MR 0424602
- [5] K. Appel and W. Haken, *A proof of the four color theorem*, Discrete Math. **16** (1976), 179–180.
- [6] K. Appel and W. Haken, *The solution of the four-color-map problem*, Sci. Amer. **237** (1977), no. 4, 108–121. MR 0543796
- [7] K. Appel and W. Haken, *Every planar map is four colorable, Part I: Discharging*, Illinois J. Math. **21** (1977), 429–490. MR 0543792
- [8] K. Appel and W. Haken, *The four-color problem*, Mathematics today (L. A. Steen, ed.), Springer, Berlin, 1978, pp. 153–180. MR 0543791
- [9] K. Appel and W. Haken, *An unavoidable set of configurations in planar triangulations*, J. Combin. Theory Ser. B **26** (1979), 1–21. MR 0525813
- [10] K. Appel and W. Haken, *The four color proof suffices*, Math. Intelligencer **8** (1986), no. 1, 10–20, 58. MR 0823216
- [11] K. Appel and W. Haken, *Every Map is Four Colorable*, American Mathematical Society, Providence, RI, 1989. MR 1025335
- [12] K. Appel, W. Haken and J. Koch, *Every planar map is four colourable, Part II: Reducibility*, Illinois J. Math. **21** (1977), 491–567. MR 0543793
- [13] G. D. Birkhoff, *The reducibility of maps*, Amer. J. Math. **35** (1913), 115–128. MR 1506176
- [14] A. Cayley, *On the colouring of maps*, Proc. Roy. Geogr. Soc. **1** (1879), 259–261.
- [15] P. Franklin, *The four color problem*, Amer. J. Math. **44** (1922), 225–236. MR 1506473
- [16] G. Gonthier, *Formal proof—The four-color theorem*, Notices Amer. Math. Soc. **55** (2008), 1382–1393. MR 2463991
- [17] G. H. Hardy, *A Mathematician's Apology*, Cambridge University Press, Cambridge, 1940. MR 0003386
- [18] P. J. Heawood, *Map-colour theorem*, Quart. J. Pure Appl. Math. **24** (1890), 332–338.
- [19] A. B. Kempe, *On the geographical problem of the four colours*, Amer. J. Math. **2** (1879), 193–200. MR 1505218
- [20] H. Lebesgue, *Quelques conséquences simple de la formule d'Euler*, J. Math. Pures Appl. (9) **19** (1940), 27–43. MR 0001903
- [21] D. MacKenzie, *Slaying the Kraken: The sociohistory of a mathematical proof*, Soc. Stud. Sci. **29** (1999), 7–60. MR 1692830
- [22] O. Ore, *The Four-Color Problem*, Academic Press, New York, 1967. MR 0216979

- [23] N. Robertson, D. Sanders, P. Seymour and R. Thomas, *The four-colour theorem*, J. Combin. Theory Ser. B **70** (1997), 2–44. [MR 1441258](#)
- [24] T. L. Saaty and P. C. Kainen, *The Four-Color Problem: Assaults and Conquest*, McGraw-Hill, New York, 1977. [MR 0480047](#)
- [25] L. Steen, *Solution of the four color problem*, Math. Mag. **49** (1976), 219–222. [MR 0447034](#)
- [26] J. P. Steinberger, *An unavoidable set of D -reducible configurations*, Trans. Amer. Math. Soc. **362** (2010), 6633–6661. [MR 2678989](#)
- [27] W. Stromquist, *The four-color theorem for small maps*, J. Combin. Theory Ser. B **19** (1975), 256–268. [MR 0398880](#)
- [28] R. Thomas, *An update on the four-color theorem*, Notices Amer. Math. Soc. **45** (1998), 848–859. [MR 1633714](#)
- [29] W. T. Tutte, *Map coloring problems and chromatic polynomials*, Amer. Sci. **62** (1974), 702–705.
- [30] W. T. Tutte, *Colouring problems*, Math. Intelligencer **1** (1978), 72–75. [MR 0504663](#)
- [31] P. Wernicke, *Über den kartographischen Vierfarbensatz*, Math. Ann. **58** (1904), 413–428. [MR 1511242](#)

ROBIN WILSON, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OXFORD, ANDREW WILES BUILDING, WOODSTOCK ROAD, OXFORD OX2 6GG, UNITED KINGDOM

E-mail address: r.j.wilson@open.ac.uk