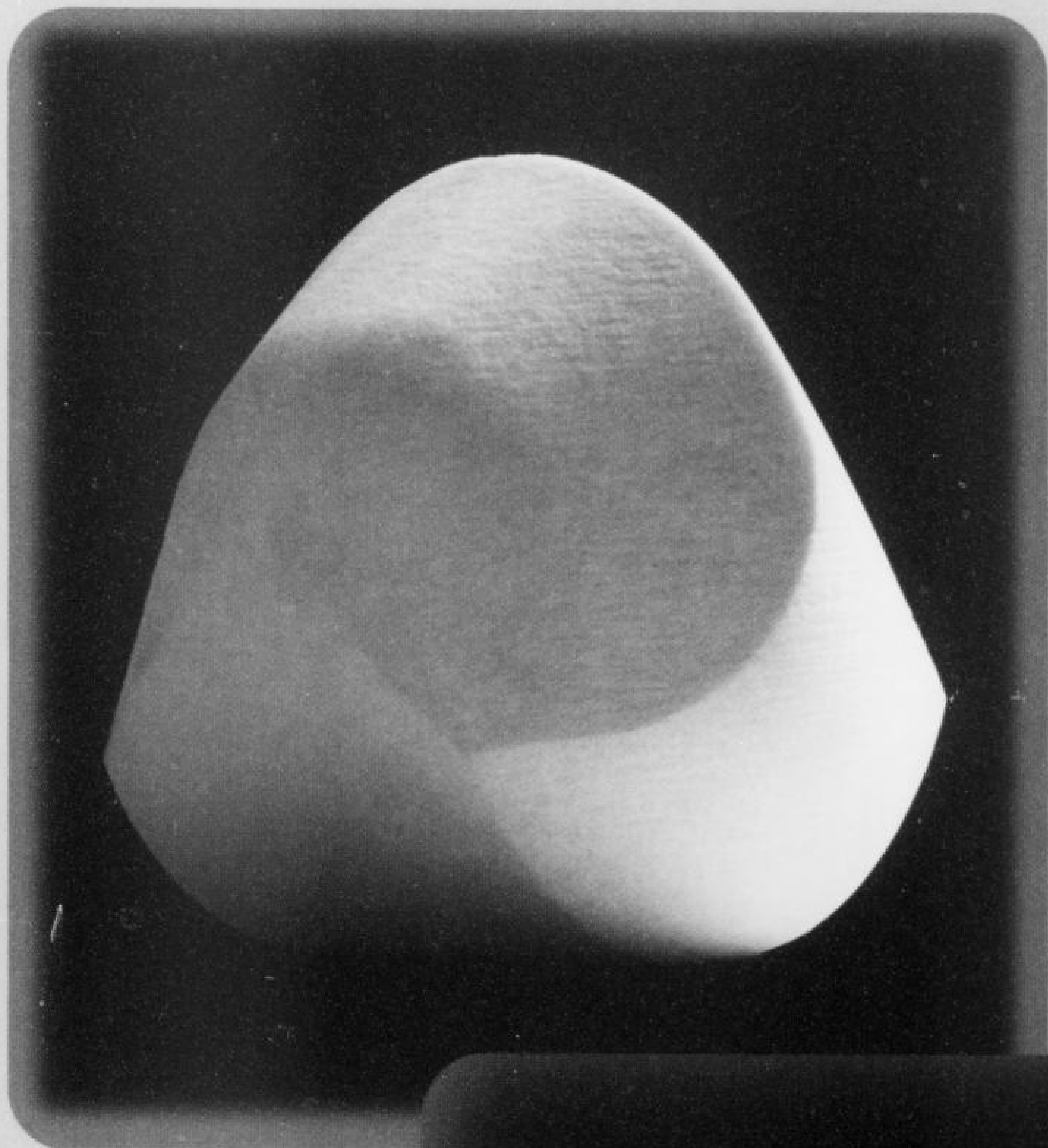
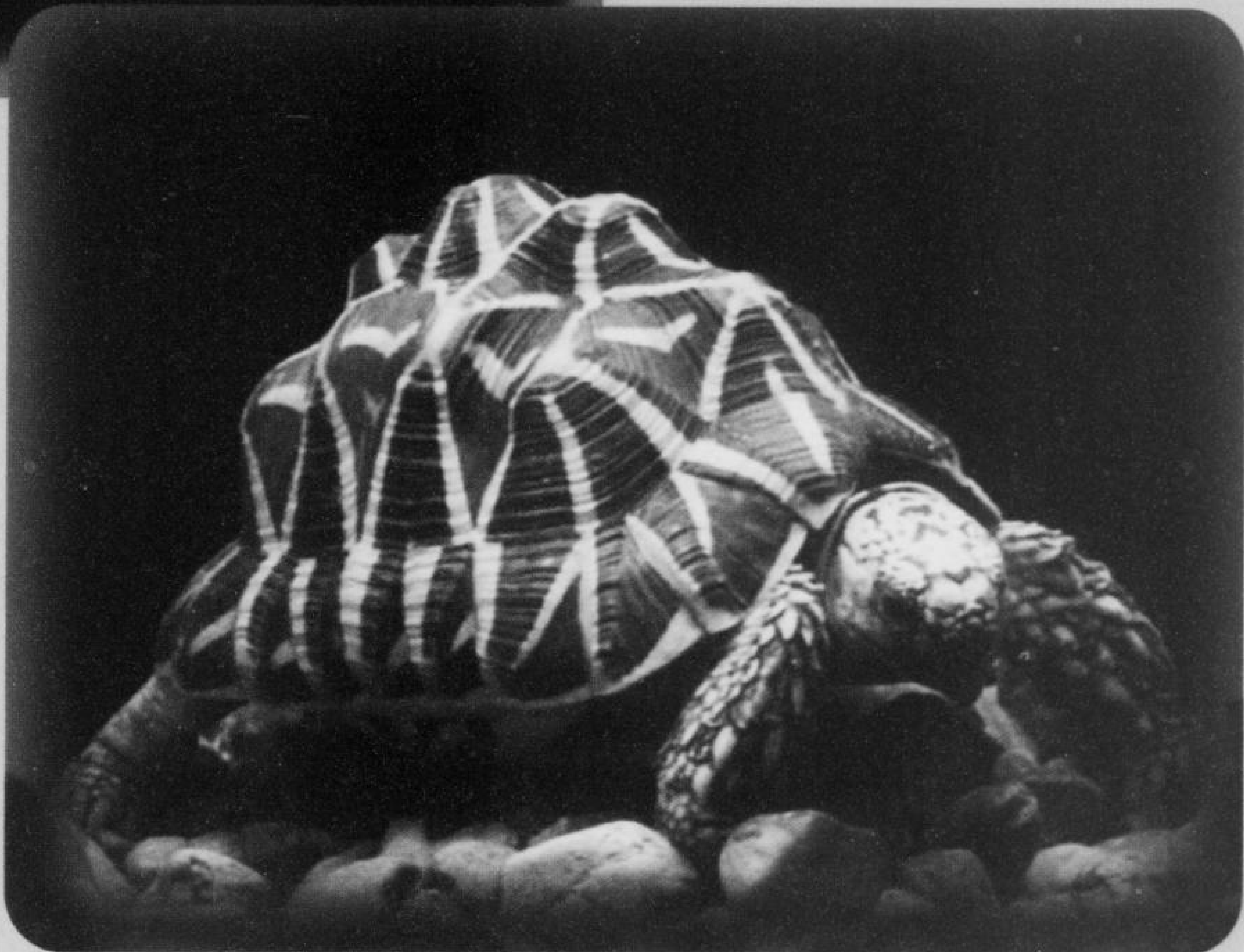
VOLUME 28
NUMBER 4
FALL 2006

The Mathematical Intelligencer





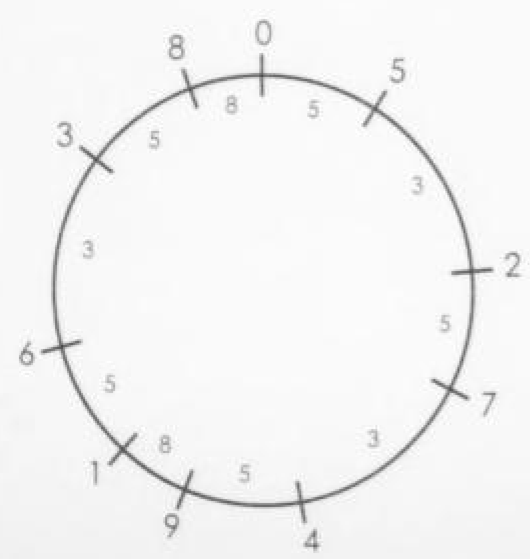
Optimal Turtles

Springer

The Mathematical Intelligencer
FALL 2006

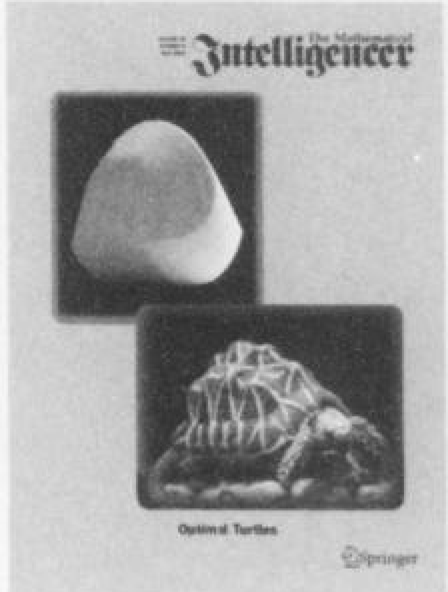
Articles

Departments



Let the sunflower seeds fall where they may.
See pp. 23–30.

ON THE COVER:
Optimal turtles. Does the turtle know, or does the invisible hand of Natural Selection know, how well-chosen the shape of the shell is?
See pp. 34–38



The Mathematical Intelligencer

INSTRUCTIONS TO AUTHORS

The journal encourages authors to submit electronically prepared manuscripts and illustrations. Manuscripts should be submitted to the Editors in LaTeX (plain LaTeX, please! Do not attempt layout—centering, columns, etc.). Illustrations should be high-resolution (.eps, .tif). If you cannot submit in LaTeX, we will consider manuscripts in Word (send disk and two copies, double-spaced) or on paper (send two copies, double-spaced). If you cannot send us digitized illustrations, please send photographs in black-and-white glossy prints, and diagrams on separate sheets of heavy paper, approximately twice the normal size. Corresponding author may order 50 free reprints; additional reprints may be purchased at http://www.sheridan.com/springer/eoc. Print out the order form and fax to the number on the form.

AIMS AND SCOPE: *The Mathematical Intelligencer* publishes articles about mathematics, about mathematicians, and about the history and culture of mathematics.

It informs and entertains a broad audience of mathematicians and the wider intellectual community. *The Mathematical Intelligencer* welcomes expository articles on all kinds of mathematics and interdisciplinary trends, and articles that portray the diversity of mathematical communities and mathematical thought. Humor is welcome, as are puzzles, poetry, fiction, and of course art. Forthcoming issues will feature emergent mathematical communities around the world, new interdisciplinary trends, and relations between mathematics and other areas of culture.

OFFICE OF PUBLICATION: Springer, 233 Spring Street, New York, NY 10013, USA. Telephone: (212) 460-1500; e-mail: journals@springer-sbm.com Production Editor: Susan Hannan; William Payne, Assistant Manager.

The editors welcome suggestions about any aspect of *The Mathematical Intelligencer.*

SUBSCRIPTION INFORMATION: *The Mathematical Intelligencer* is published by Springer Science+Business Media, Inc. winter, spring, summer and fall. Volume 28 (4 issues) will be published in 2006. ISSN: 0343-6993

SUBSCRIPTION RATES:

Please contact our Customer Service department for the latest rates and pricing information:

The Americas (North, South, Central America, and the Caribbean)
MAIL: Journals Customer Service
P.O. Box 2485
Secaucus, NJ 07094-2485
TEL: 800-777-4643; 212-460-1500 (outside North America)
FAX: 201-348-4505
E-MAIL: journals-ny@springer-sbm.com

Outside of the Americas:
MAIL: Journals Customer Service
Springer Distribution Center (SDC)
Haberstr. 7
69126 Heidelberg
GERMANY
TEL: 49-6221-345-4303
FAX: 49-6221-345-4229
E-MAIL: sdc-journals@springer-sbm.com

CHANGES OF ADDRESS: Allow six weeks for all changes to become effective. All communications should include both old and new addresses (with zip codes) and should be accompanied by a mailing label from a recent issue.

BACK VOLUMES: Prices for **back volumes** are available on request. **MICROFORM EDITIONS** are available from: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106, USA.

SPRINGERALERT SERVICE: The SpringerAlert service is an innovative, free-of-charge service that notifies users via e-mail whenever new SpringerLink articles and journals become available, and automatically sends the table of contents and direct links to the abstracts of a new issue of a journal in SpringerLink. Register now for the SpringerAlert service at http://www.springerlink.com/alerting

A Mysterious Subterranean Numeroglyph

GREG HUBER

A recent trip to Portland, Oregon, placed me in the eastbound tube of the Washington Park MAX transit station. It's a dramatic underground location, the Robertson tunnel, being nearly 80 meters under Portland's West Hills. Into one wall of the platform are carved a geological timeline and a number of artworks inspired by science and engineering. There, carved in stone, as a testament to human mathematical accomplishment, I suppose, are about 100 decimal digits of π. Well . . . not quite π (see Figure 1).

After the familiar 3.1415926535 comes the much less familiar (and much less correct) string of digits 821480865144. . . . Now, as everyone and his bubbe knows, the second row of numbers should continue with 897932. . . . But *that* very string of digits turns up in the bottom row of the carving. Very odd!

My colleague Melanie Mitchell of Portland State University found an online source [1] that provided the key to this mystery. The artist who created the display, Bill Will, an ardent admirer of mathematics, apparently took his digits from a reference book. I'll hazard the guess that he consulted the final pages of Petr Beckmann's classic *A History of Pi* [2], wherein are listed ten thousand decimal digits of π, neatly tabulated at one hundred digits per row. Beckmann's table is itself a reprint of the 1961 computation by Shanks and

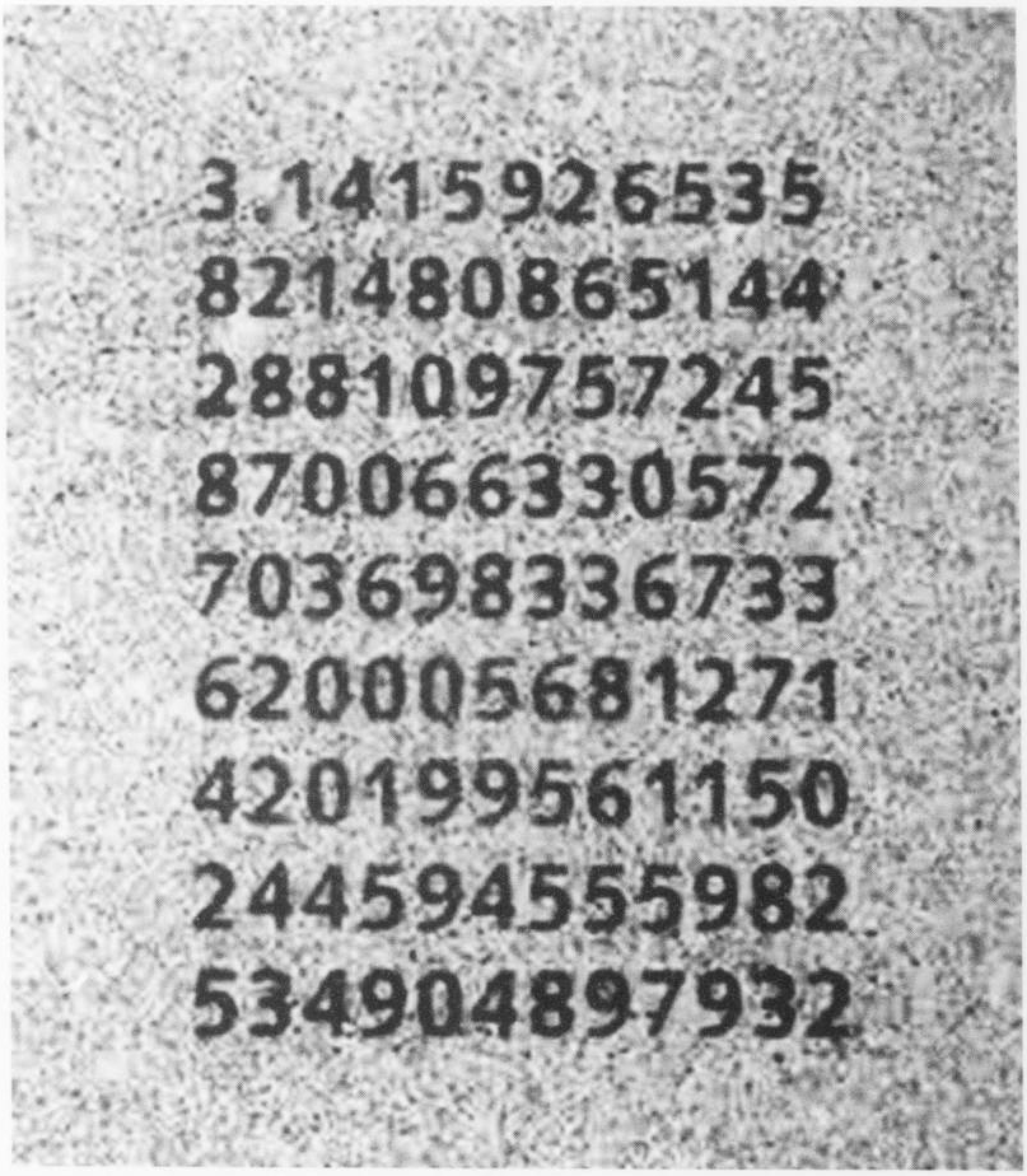


Figure 1.

PI = 3.4

1415926535	8979323846	2643383279	5028841971	6939937510	5820974944	5923078164	0628620899	8628034825	3421170679
8214808651	3282306647	0938446095	5058223172	5359408128	4811174502	8410270193	8521105559	6446229489	5493038196
4428810975	6659334461	2847564823	3786783165	2712019091	4564856692	3460348610	4543266482	1339360726	0249141273
7245870066	0631558817	4881520920	9628292540	9171536436	7892590360	0113305305	4882046652	1384146951	9415116094
3305727036	5759591953	0921861173	8193261179	3105118548	0744623799	6274956735	1885752724	8912279381	8301194912
9833673362	4406566430	8602139494	6395224737	1907021798	6094370277	0539217176	2931767523	8467481846	7669405132
0005681271	4526356082	7785771342	7577896091	7363717872	1468440901	2249534301	4654958537	1050792279	6892589235
4201995611	2129021960	8640344181	5981362977	4771309960	5187072113	4999999837	2978049951	0597317328	1609631859
5024459455	3469083026	4252230825	3344685035	2619311881	7101000313	7838752886	5875332083	8142061717	7669147303
5982534904	2875546873	1159562863	8823537875	9375195778	1857780532	1712268066	1300192787	6611195909	2164201989

Figure 2.

Wrench [3]. What threw Will off were the groups of 10 × 10 blocks that comprised the table (see Figure 2). Instead of copying out 100 digits from the first lengthy row, Will copied the digits (highlighted in red in Figure 2) from the first 10 × 10 block to create his own block design. In fact, he also took some digits from a neighboring block, which explains the reappearance of 897932 at the end. Or so one might surmise. On the other hand, could it be that Will displayed another fundamental number on the wall under the West Hills, one that differs from our familiar π by a wispy one part in 411 billion? I'd like to think that Mathgod is subtle, but not malicious.

REFERENCES

[1] http://www.wweek.com/html/newsbuzz081998.html

[2] P. Beckmann. *A History of Pi*, Golem Press, Boulder CO, 1971.

[3] D. Shanks and J. W. Wrench, Calculation of Pi to 100,000 Decimals, *Mathematics of Computation* **16** (1962), 76–99.

Center for Cell Analysis & Modeling
University of Connecticut
Farmington, CT 06032-1507
USA
e-mail: huber@uchc.edu

United States Postal Service

Statement of Ownership, Management, and Circulation

1. Publication Title: MATHEMATICAL INTELLIGENCER
2. Publication Number: 001 – 656
3. Filing Date: 10/1/06
4. Issue Frequency: QUARTERLY
5. Number of Issues Published Annually: 4
6. Annual Subscription Price: $85.00
7. Complete Mailing Address of Known Office of Publication *(Not printer) (Street,city,county,state,and zip+4)*: 233 SPRING ST - 6TH FL, NEW YORK, NY 10013
   Contact Person: ANTHONY FERRIGNO, Telephone (212)460-1632
8. Complete Mailing Address of Headquarters or General Business Office of Publisher *(Not printer)*: SAME AS ABOVE
9. Full Name and Complete Mailing Addresses of Publisher, Editor, and Managing Editor *(Do not leave blank)*
   Publisher *(Name and complete mailing address)*: SPRINGER SCIENCE & BUSINESS MEDIA, 233 SPRING ST - 6TH FL NEW YORK, NY 10013
   Editor *(Name and complete mailing address)*: Prof Marjorie Lee Senechal - Smith College - Dept of Mathematics, 311 Burton St Northampton, MA 01063
   Managing Editor *(Name and complete mailing address)*: none
10. Owner *(Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names a[nd] addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If owned by a corporation, give the names and addresses of the individual owners. If owned by a partnership, or other unincorporated firm, give its name and address as well as those of each individual owner. If the publicatio[n is] published by a nonprofit organization, give its name and address.)*

Full Name	Complete Mailing Address
SPRINGER SCIENCE & BUSINESS MEDIA	233 SPRING ST 6TH FL NEW YORK, NY 10013

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 percent or more of Total Amount of Bonds, Mortgages, or Other Securities. If none check box → ☒ None

Full Name	Complete Mailing Address

12. Tax Status *(For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)*
    The purpose, function and nonprofit status of this organization and the exempt status for federal income tax purposes:
    ☒ Has Not Changed During Preceding 12 Months
    ☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of ...

13. Publication Title: MATHEMATICAL INTELLIGENCER
14. Issue Date for Circulation Data Below: SPRING 06

15. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies *(Net press run)*	3621	3622
b. Paid and/or Requested Circulation (1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541. *(Include advertiser's proof and Exchange copies)*	1173	1278
(2) Paid In-County Subscriptions Stated on Form 3541 *(Include advertiser's proof and exchange copies)*	0	0
(3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution	1363	1328
(4) Other Classes Mailed Through the USPS	0	0
c. Total Paid and/or Requested Circulation *[Sum of 15b. (1), (2), (3) and (4)]*	2536	2606
d. Free Distribution by Mail (Samples, complimentary and other free) (1) Outside-County as Stated on Form 3541	0	0
(2) In-County as Stated on Form 3541	0	0
(3) Other Classes Mailed Through USPS	0	0
e. Free Distribution Outside the Mail *(Carriers or other means)*	75	69
f. Total Free Distribution *(Sum of 15d. and 15e.)*	75	69
g. Total Distribution *(Sum of 15c. and 15f.)*	2611	2675
h. Copies not Distributed	1010	947
i. Total *(Sum of 15g. and 15h.)*	3621	3622
j. Percent Paid and/or Requested Circulation *(15c. divided by 15g. times 100)*	97%	97%

16. Publication of Statement of Ownership
    Publication required. Will be printed in the FALL 2006 issue of this publication. ☒ Publication not required.
17. Signature and Title of Editor, Publisher, Business Manager, or Owner
    Anthony Ferrigno, Director of Logistics — Date 8/18/2006

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

Instructions to Publishers

1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the stockholder or security holder is a trustee, include in items 10 and 11 the name of the person or corporation for whom the trustee is acting. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing corporation. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 15d, e and f.
4. Item 15h., Copies not Distributed, must include (1) newsstand copies originally stated on Form 3541, and returned to the publisher, (2) estimated return[s] from news agents, and (3), copies for office use, leftovers, spoiled, and all other copies not distributed.
5. If the publication had Periodicals authorized as a general or requester publication, this Statement of Ownership, Management, and Circulation must be published; it must be printed in an issue in October or, if the publication is not published during October, the first issue printed after October.
6. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
7. Item 17 must be signed. *Failure to file or publish a statement of ownership may lead to suspension of Periodicals authorization.*

PS Form 3526, October 1999 (Reverse)

Google's PageRank
The Math Behind the Search Engine

REBECCA S. WILLS

Approximately 94 million American adults use the Internet on a typical day [24]. The number-one Internet activity is reading and writing e-mail. Search engine use is next in line and continues to increase in popularity. In fact, survey findings indicate that nearly 60 million American adults use search engines on a given day. Even though there are many Internet search engines, Google, Yahoo!, and MSN receive over 81% of all search requests [27]. Despite claims that the quality of search provided by Yahoo! and MSN now equals that of Google [11], Google continues to thrive as the search engine of choice, receiving over 46% of all search requests, nearly double the volume of Yahoo! and over four times that of MSN.

I use Google's search engine on a daily basis and rarely request information from other search engines. One day, I decided to visit the homepages of Google, Yahoo!, and MSN to compare the quality of search results. Coffee was on my mind that day, so I entered the simple query "coffee" in the search box at each homepage. Table 1 shows the top ten (unsponsored) results returned by each search engine. Although ordered differently, two webpages, *www.peets.com* and *www.coffeegeek.com*, appear in all three top ten lists. In addition, each pairing of top ten lists has two additional results in common.

Depending on the information I hoped to obtain about coffee by using the search engines, I could argue that any one of the three returned better results; however, I was not looking for a particular webpage, so all three listings of search results seemed of equal quality. Thus, I plan to continue using Google. My decision is indicative of the problem Yahoo!, MSN, and other search engine companies face in the quest to obtain a larger percentage of Internet search volume. Search engine users are loyal to one or a few search engines and are generally happy with search results [14, 28]. Thus, as long as Google continues to provide results deemed high in quality, Google likely will remain the top search engine. But what set Google apart from its competitors in the first place? The answer is PageRank. In this article I explain this simple mathematical algorithm that revolutionized Web search.

Google's Search Engine

Google founders Sergey Brin and Larry Page met in 1995 when Page visited the computer science department of Stanford University during a recruitment weekend [2, 9]. Brin, a second-year graduate student at the time, served as a guide for potential recruits, and Page was part of his group. They discussed many topics during their first meeting and disagreed on nearly every issue. Soon after he began graduate study at Stanford, Page began working on a Web project, initially called BackRub, that exploited the link structure of the Web. Brin found Page's work on BackRub interesting, so the two started working together on a project that would permanently change Web search. Brin and Page realized that they were creating a search engine that adapted to the ever-increasing size of the Web, so they replaced the name BackRub with Google (a common misspelling of *googol*, the number 10^{100}). Unable to convince existing search engine companies to adopt the technology they had developed but certain their technology was superior to any being used, Brin and Page decided to start their own company. With the financial assistance of a small group of initial investors, Brin and Page founded the Web search engine company Google, Inc. in September 1998.

Almost immediately, the general public noticed what Brin, Page, and others in the academic Web search community already knew—the Google search engine produced much higher-quality results than those produced by other Web search engines. Other search engines relied entirely on webpage content to determine ranking of results, and

Table 1. Top Ten Results for Search Query "coffee" at *www.google.com*, *www.yahoo.com*, and *www.msn.com*, April 10, 2006

Order	Google	Yahoo!	MSN
1	www.starbucks.com (◇)	www.gevalia.com (◇)	www.peets.com (*)
2	www.coffeereview.com (†)	en.wikipedia.org/wiki/Coffee (△)	en.wikipedia.org/wiki/Coffee (△)
3	www.peets.com (*)	www.nationalgeographic.com/coffee	www.coffeegeek.com (*)
4	www.coffeegeek.com (*)	www.peets.com (*)	coffeetea.about.com (△)
5	www.coffeeuniverse.com (†)	www.starbucks.com (◇)	coffeebean.com
6	www.coffeescience.org	www.coffeegeek.com (*)	www.coffeereview.com (†)
7	www.gevalia.com (◇)	coffeetea.about.com (△)	www.coffeeuniverse.com (†)
8	www.coffeebreakarcade.com	kaffee.netfirms.com/Coffee	www.tmcm.com
9	https://www.dunkindonuts.com	www.strong-enough.net/coffee	www.coffeeforums.com
10	www.cariboucoffee.com	www.cl.cam.ac.uk/coffee/coffee.html	www.communitycoffee.com
Approximate Number of Results:			
	447,000,000	151,000,000	46,850,246

Shared results for Google, Yahoo!, and MSN (*); Google and Yahoo! (◇); Google and MSN (†); and Yahoo! and MSN (△)

Brin and Page realized that webpage developers could easily manipulate the ordering of search results by placing concealed information on webpages.[1] Brin and Page developed a ranking algorithm, named PageRank after Larry Page, that uses the link structure of the Web to determine the importance of webpages. During the processing of a query, Google's search algorithm combines precomputed PageRank scores with text-matching scores to obtain an overall ranking score for each webpage.

Although many factors determine Google's overall ranking of search engine results, Google maintains that the heart of its search engine software is PageRank [3]. A few quick searches on the Internet reveal that both the business and academic communities hold PageRank in high regard. The business community is mindful that Google remains the search engine of choice and that PageRank plays a substantial role in the order in which webpages are displayed. Maximizing the PageRank score of a webpage, therefore, has become an important component of company marketing strategies. The academic community recognizes that PageRank has connections to numerous areas of mathematics and computer science such as matrix theory, numerical analysis, information retrieval, and graph theory. As a result, much research continues to be devoted to explaining and improving PageRank.

The Mathematics of PageRank

The PageRank algorithm assigns a PageRank score to each of more than 25 billion webpages [7]. The algorithm models the behavior of an idealized *random Web surfer* [12, 23]. This Internet user randomly chooses a webpage to view from the listing of available webpages. Then, the surfer randomly selects a link from that webpage to another webpage. The surfer continues the process of selecting links at random from successive webpages until deciding to move to another webpage by some means other than selecting a link. The choice of which webpage to visit next does not depend on the previously visited webpages, and the idealized Web surfer never grows tired of visiting webpages. Thus, the PageRank score of a webpage represents the probability that a random Web surfer chooses to view that webpage.

Directed Web Graph

To model the activity of the random Web surfer, the PageRank algorithm represents the link structure of the Web as a directed graph. Webpages are nodes of the graph, and links from webpages to other webpages are edges that show direction of movement. Although the directed Web graph is very large, the PageRank algorithm can be applied to a directed graph of any size. To faciliate our discussion of PageRank, we apply the PageRank algorithm to the directed graph with 4 nodes shown in Figure 1.

AUTHOR

REBECCA S. WILLS received her undergraduate degree from High Point University. Since August 2002 she has been a graduate student at North Carolina State University, working on matrix theory and numerical linear algebra. She enjoys a good cup of coffee in the morning and a sitcom or two in the evening.

Rebecca S. Wills
Department of Mathematics
North Carolina State University
Raleigh, NC 27695
USA
e-mail: rmwills@ncsu.edu

[1]That is, a developer could add text in the same color as the background of the page, invisible to the user but detected by automated search engines. If the terms of a search query occurred many times in the hidden text, that webpage could appear higher in rank than webpages that were really more informative.

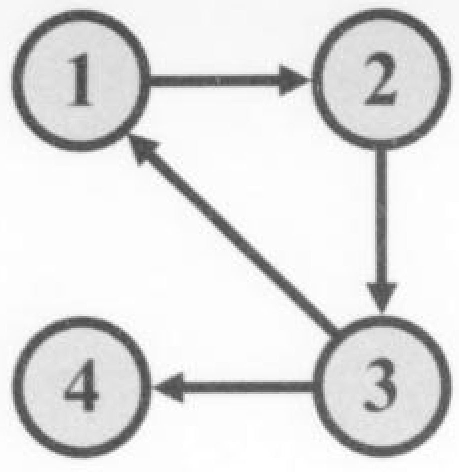


Figure 1. Directed graph with 4 nodes.

Web Hyperlink Matrix

The process for determining PageRank begins by expressing the directed Web graph as the $n \times n$ "hyperlink matrix" H, where n is the number of webpages. If webpage i has $l_i \geq 1$ links to other webpages and webpage i links to webpage j, then the element in row i and column j of H is $H_{ij} = \frac{1}{l_i}$. Otherwise, $H_{ij} = 0$. Thus, H_{ij} represents the likelihood that a random surfer will select a link from webpage i to webpage j. For the directed graph in Figure 1,

$$H = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Node 4 is a *dangling node* because it does not link to other nodes. As a result, all entries in row 4 of the example matrix are zero. This means the probability is zero that a random surfer moves from node 4 to any other node in the directed graph. The majority of webpages are dangling nodes (e.g., postscript files and image files), so there are many rows with all zero entries in the Web hyperlink matrix. When a Web surfer lands on dangling node webpages, the surfer can either stop surfing or move to another webpage, perhaps by entering the Uniform Resource Locator (URL) of a different webpage in the address line of a Web browser. Since H does not model the possibility of moving from dangling node webpages to other webpages, the long-term behavior of Web surfers cannot be determined from H alone.

Dangling Node Fix

Several options exist for modeling the behavior of a random Web surfer after landing on a dangling node, and Google does not reveal which option it employs. One option replaces each dangling node row of H by the same *probability distribution vector*, w, a vector with nonnegative elements that sum to 1. The resulting matrix is $S = H + dw$, where d is a column vector that identifies dangling nodes, meaning $d_i = 1$ if $l_i = 0$ and $d_i = 0$ otherwise; and $w = (w_1 \; w_2 \; . \; . \; . \; w_n)$ is a row vector with $w_j \geq 0$ for all $1 \leq j \leq n$ and $\sum_{j=1}^{n} w_j = 1$. The most popular choice for w is the uniform row vector, $w = (\frac{1}{n} \; \frac{1}{n} \; . \; . \; . \; \frac{1}{n})$. This amounts to adding artificial links from dangling nodes to all webpages. With $w = (\frac{1}{4} \; \frac{1}{4} \; \frac{1}{4} \; \frac{1}{4})$, the directed graph in Figure 1 changes (see Figure 2).

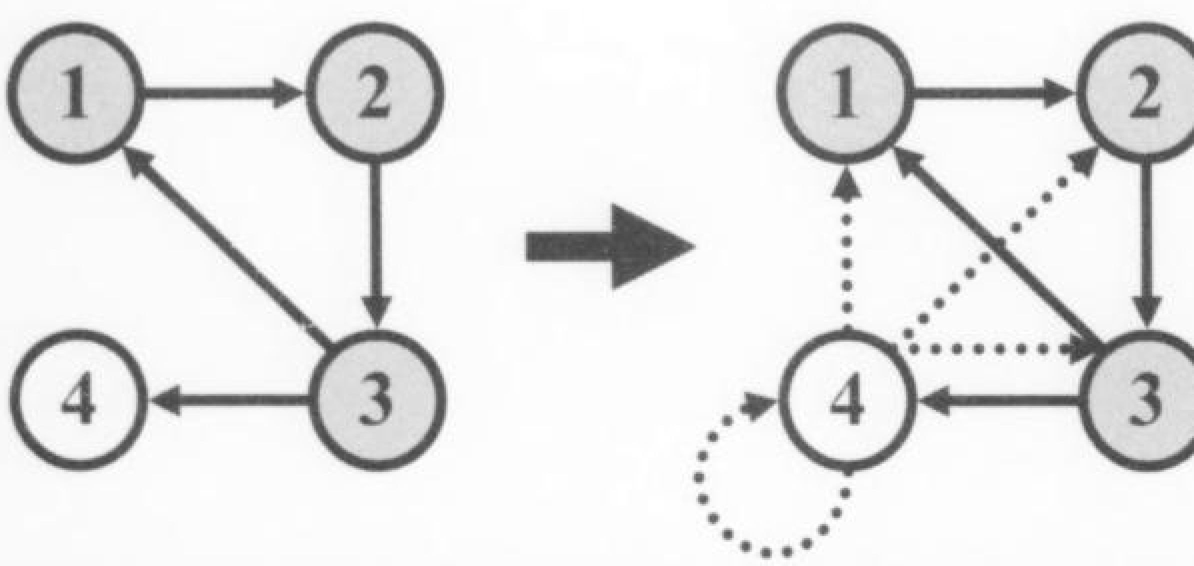


Figure 2. Dangling node fix to Figure 1.

The new matrix $S = H + dw$ is,

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Regardless of the option chosen to deal with dangling nodes, Google creates a new matrix S that models the tendency of random Web surfers to leave a dangling node; however, the model is not yet complete. Even when webpages have links to other webpages, a random Web surfer might grow tired of continually selecting links and decide to move to a different webpage some other way. For the graph in Figure 2, there is no directed edge from node 2 to node 1. On the Web, though, a surfer can move directly from node 2 to node 1 by entering the URL for node 1 in the address line of a Web browser. The matrix S does not consider this possibility.

Google Matrix

To model the overall behavior of a random Web surfer, Google forms the matrix $G = \alpha S + (1 - \alpha)\mathbb{1}v$, where $0 \leq \alpha < 1$ is a scalar, $\mathbb{1}$ is the column vector of ones, and v is a row probability distribution vector called the *personalization vector*. The *damping factor*, α, in the Google matrix indicates that random Web surfers move to a different webpage by some means other than selecting a link with probability $1 - \alpha$. The majority of experiments performed by Brin and Page during the development of the PageRank algorithm used $\alpha = 0.85$ and $v = (\frac{1}{n} \; \frac{1}{n} \; . \; . \; . \; \frac{1}{n})$ [12, 23]. Values of α ranging from 0.85 to 0.99 appear in most research papers on the PageRank algorithm.

Assigning the uniform vector for v suggests Web surfers randomly choose new webpages to view when not selecting links. The uniform vector makes PageRank highly susceptible to *link spamming*, so Google does not use it to determine actual PageRank scores. Link spamming is the practice by some search engine optimization experts of adding more links to their clients' webpages for the sole purpose of increasing the PageRank score of those webpages. This attempt to manipulate PageRank scores is one reason Google does not reveal the current damping factor or personalization vector for the Google matrix. In 2004, however, Gyöngyi, Garcia-Molina, and Pederson developed the TrustRank algorithm to create a personalization vector that decreases the harmful effect of link spamming [17], and Google registered the trademark for TrustRank on March 16, 2005 [6].

Because each element G_{ij} of G lies between 0 and 1 ($0 \leq G_{ij} \leq 1$) and the sum of elements in each row of G is 1, the Google matrix is called a *row-stochastic* matrix. It is known that $\lambda = 1$ is not a repeated eigenvalue of G and is greater in magnitude than any other eigenvalue of G [18, 26]. Hence the eigensystem $\pi G = \pi$ has a unique solution, where π is a row probability distribution vector.[2] We say that $\lambda = 1$ is the *dominant eigenvalue* of G, and π is the corresponding *dominant left eigenvector* of G. The ith entry of π is the PageRank score for webpage i, and π is called the PageRank vector.

Table 2 shows four different Google matrices and their corresponding PageRank vectors (approximated to two decimal places) for the directed graph in Figure 2. The table indicates that the personalization vector has more influence on the PageRank scores for smaller damping factors. For instance, when $\alpha = 0.85$, as is the case for the first and second models, the PageRank scores and the ordering of the scores differ significantly. The first model assigns the uniform vector to v, and node 1 is one of the nodes with the lowest PageRank score. The second model uses $v = (1\ 0\ 0\ 0)$, and node 1 receives the highest PageRank score. This personalization vector suggests that when Web surfers grow tired of following the link structure of the Web, they always move to node 1. For the third and fourth models, $\alpha = 0.95$. The difference in PageRank scores and ordering of scores for these models is less significant. Even though $v = (1\ 0\ 0\ 0)$ in the fourth model, the higher damping factor decreases the influence of v.

Computing PageRank Scores

For small Google matrices like the ones in Table 2, we can quickly find exact solutions to the eigensystem, $\pi G = \pi$. The Google matrix for the entire Web has more than 25 billion rows and columns, so computing the exact solution requires extensive time and computing resources. The oldest and easiest technique for approximating a dominant eigenvector of a matrix is the power method. The power method converges for most starting vectors when the dominant eigenvalue is not a repeated eigenvalue [13, §9.4]. Since $\lambda = 1$ is the dominant eigenvalue of G and π is the dominant left eigenvector, the power method applied to G converges to the PageRank vector. This method was the original choice for computing the PageRank vector.

Given a starting vector $\pi^{(0)}$, e.g. $\pi^{(0)} = v$, the power method calculates successive iterates

$$\pi^{(k)} = \pi^{(k-1)}G, \text{ where } k = 1, 2, \ldots,$$

until some convergence criterion is satisfied. Notice that $\pi^{(k)} = \pi^{(k-1)}G$ can also be stated $\pi^{(k)} = \pi^{(0)}G^k$. As the number of nonzero elements of the personalization vector increases, the number of nonzero elements of G increases. Thus, the multiplication of $\pi^{(k-1)}$ with G is expensive; however, since $S = H + dw$ and $G = \alpha S + (1 - \alpha)\mathbb{1}v$, we can express the multiplication as follows:

$$\begin{aligned}
\pi^{(k)} &= \pi^{(k-1)}G \\
&= \pi^{(k-1)}\,[\alpha(H + dw) + (1 - \alpha)\,\mathbb{1}v] \\
&= \alpha\pi^{(k-1)}\,H + \alpha(\pi^{(k-1)}\,d)\,w + (1 - \alpha)\,(\pi^{(k-1)}\,\mathbb{1})\,v \\
&= \alpha\pi^{(k-1)}\,H + \alpha(\pi^{(k-1)}\,d)\,w + (1 - \alpha)\,v,
\end{aligned}$$

because $\pi^{(k-1)}\mathbb{1} = 1$, $\pi^{(k-1)}$ is a probability vector. This is a sum of three vectors: a multiple of $\pi^{(k-1)}\,H$, a multiple of w, and a multiple of v. (Notice that $\pi^{(k-1)}d$ is a scalar.) The only matrix-vector multiplication required is with the hyperlink matrix H. A 2004 investigation of Web documents estimates that the average number of outlinks for a webpage is 52 [22]. This means that for a typical row of the hyperlink matrix only 52 of the 25 billion elements are nonzero, so the majority of elements in H are 0 (H is very sparse). Since all computations involve the sparse matrix H and vectors w and v, an iteration of the power method is cheap (the operation count is proportional to the matrix dimension n).

Writing a subroutine to approximate the PageRank vector using the power method is quick and easy. For a simple program (in MATLAB), see Langville and Meyer [20, §4.6].

The ratio of the two eigenvalues largest in magnitude for a given matrix determines how quickly the power method converges [16]. Haveliwala and Kamvar were the first to prove that the second-largest eigenvalue in magnitude of G is less than or equal to the damping factor α [18]. This means that the ratio is less than or equal to α for the Google matrix. Thus, the power method converges quickly when α is less than 1. This might explain why Brin and Page originally used $\alpha = 0.85$. No more than 29 iterations are required for the maximal element of the difference in successive iterates, $\pi^{(k+1)} - \pi^{(k)}$, to be less than 10^{-2} for $\alpha = 0.85$. The number of iterations increases to 44 for $\alpha = 0.90$.

An Alternative Way to Compute PageRank

Although Brin and Page originally defined PageRank as a solution to the eigensystem $\pi G = \pi$, the problem can be restated as a linear system. Recall, $G = \alpha S + (1 - \alpha)\,\mathbb{1}v$. Transforming $\pi G = \pi$ to $0 = \pi - \pi G$ gives:

$$\begin{aligned}
0 &= \pi - \pi G \\
&= \pi I - \pi(\alpha S + (1 - \alpha)\,\mathbb{1}v) \\
&= \pi(I - \alpha S) - (1 - \alpha)\,(\pi\mathbb{1})v \\
&= \pi(I - \alpha S) - (1 - \alpha)v
\end{aligned}$$

The last equality follows as above from the fact that π is a probability distribution vector, so $\pi\mathbb{1} = 1$. Thus

$$\pi(I - \alpha S) = (1 - \alpha)v,$$

which means π solves a linear system with coefficient matrix $I - \alpha S$ and right-hand side $(1 - \alpha)v$. Since the matrix $I - \alpha S$ is nonsingular [19], the linear system has a unique solution. For more details on viewing PageRank as the solution of a linear system, see [8, 10, 15, 19].

[2] Though not required, the restriction is often made that the personalization vector v and the dangling node vector w have all positive entries that sum to 1 instead of all non-negative entries that sum to 1. Under this restriction, the PageRank vector also has all positive entries that sum to 1.

Table 2. Modeling Surfer Behavior for the Directed Graph in Figure 2

	Damping Factor (α)	Personalization Vector (**v**)	Google Matrix (**G**)	Page Rank Vector ($\approx \pi$)	Ordering of Nodes (1 = Highest)
Model 1	0.85	$\left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right)$	$\begin{pmatrix} \frac{3}{80} & \frac{71}{80} & \frac{3}{80} & \frac{3}{80} \\ \frac{3}{80} & \frac{3}{80} & \frac{71}{80} & \frac{3}{80} \\ \frac{37}{80} & \frac{3}{80} & \frac{3}{80} & \frac{37}{80} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$	(0.21 0.26 0.31 0.21)	(3 2 1 3)
Model 2	0.85	(1 0 0 0)	$\begin{pmatrix} \frac{3}{20} & \frac{17}{20} & 0 & 0 \\ \frac{3}{20} & 0 & \frac{17}{20} & 0 \\ \frac{23}{40} & 0 & 0 & \frac{17}{40} \\ \frac{29}{80} & \frac{17}{80} & \frac{17}{80} & \frac{17}{80} \end{pmatrix}$	(0.30 0.28 0.27 0.15)	(1 2 3 4)
Model 3	0.95	$\left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right)$	$\begin{pmatrix} \frac{1}{80} & \frac{77}{80} & \frac{1}{80} & \frac{1}{80} \\ \frac{1}{80} & \frac{1}{80} & \frac{77}{80} & \frac{1}{80} \\ \frac{39}{80} & \frac{1}{80} & \frac{1}{80} & \frac{39}{80} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$	(0.21 0.26 0.31 0.21)	(3 2 1 3)
Model 4	0.95	(1 0 0 0)	$\begin{pmatrix} \frac{1}{20} & \frac{19}{20} & 0 & 0 \\ \frac{1}{20} & 0 & \frac{19}{20} & 0 \\ \frac{21}{40} & 0 & 0 & \frac{19}{40} \\ \frac{23}{80} & \frac{19}{80} & \frac{19}{80} & \frac{19}{80} \end{pmatrix}$	(0.24 0.27 0.30 0.19)	(3 2 1 4)

Google's Toolbar PageRank

The PageRank score of a webpage corresponds to an entry of the PageRank vector, π. Since π is a probability distribution vector, all elements of π are non-negative and sum to one. Google's toolbar includes a PageRank display feature that provides "an indication of the PageRank" for a webpage being visited [5]. The PageRank scores on the toolbar are integer values from 0 (lowest) to 10 (highest). Although some search engine optimization experts discount the accuracy of toolbar scores [25], a Google webpage on toolbar features [4] states:

> PageRank Display: Wondering whether a new website is worth your time? Use the Toolbar's PageRank™ display to tell you how Google's algorithms assess the importance of the page you're viewing.

Table 3. Toolbar PageRank Scores for the Top Ten Results Returned by *www.google.com* for April 10, 2006, Search Query "coffee"

Order	Google's Top Ten Results	Toolbar PageRank
1	www.starbucks.com	7
2	www.coffeereview.com	6
3	www.peets.com	7
4	www.coffeegeek.com	6
5	www.coffeeuniverse.com	6
6	www.coffeescience.org	6
7	www.gevalia.com	6
8	www.coffeebreakarcade.com	6
9	https://www.dunkindonuts.com	7
10	www.cariboucoffee.com	6

Results returned by Google for a search on Google's toolbar PageRank reveal that many people pay close attention to the toolbar PageRank scores. One website [1] asserts that website owners have become addicted to toolbar PageRank.

Although Google does not explain how toolbar PageRank scores are determined, they are possibly based on a logarithmic scale. It is easy to verify that few webpages receive a toolbar PageRank score of 10, but many webpages have very low scores.

Two weeks after creating Table 1, I checked the toolbar PageRank scores for the top ten results returned by Google for the query "coffee." Those scores are listed in Table 3. They reveal a point worth emphasizing. Although PageRank is an important component of Google's overall ranking of results, it is not the only component. Notice that *https://www.dunkindonuts.com* is the ninth result in Google's top ten list. There are six results considered more relevant by Google to the query "coffee" that have lower toolbar PageRank scores than *https://www.dunkindonuts.com*. Also, Table 1 shows that both Yahoo! and MSN returned *coffeetea.about.com* and *en.wikipedia.org/wiki/Coffee* in their top ten listings. The toolbar PageRank score for both webpages is 7; however, they appear in Google's listing of results at 18 and 21, respectively.

Since a high PageRank score for a webpage does not guarantee that the webpage appears high in the listing of search results, search engine optimization experts emphasize that "on the page" factors, such as placement and frequency of important words, must be considered when developing good webpages. Even the news media have started making adjustments to titles and content of articles to improve rankings in search engine results [21]. The fact is most search engine users expect to find relevant information quickly, for any topic. To keep users satisfied, Google must make sure that the most relevant webpages appear at the top of listings. To remain competitive, companies and news media must figure out a way to make it there.

Want to Know More?

For more information on PageRank, see the survey papers by Berkhin [10] and Langville and Meyer [19]. In addition, the textbook [20] by Langville and Meyer provides a detailed overview of PageRank and other ranking algorithms.

ACKNOWLEDGMENTS

Many people reviewed this article, and I thank each of them. In particular, I thank Ilse Ipsen and Steve Kirkland for encouraging me to write this article. I thank Ilse Ipsen and my fellow "Communicating Applied Mathematics" classmates, Brandy Benedict, Prakash Chanchana, Kristen DeVault, Kelly Dickson, Karen Dillard, Anjela Govan, Rizwana Rehman, and Teresa Selee, for reading and re-reading preliminary drafts. Finally, I thank Jay Wills for helping me find the right words to say.

REFERENCES

[1] *www.abcseo.com/seo-book/toolbar-google.htm*, Google Toolbar PageRank.

[2] *http://www.google.com/corporate/history.html*, Google Corporate Information: Google Milestones.

[3] *http://www.google.com/technology/index.html*, Our Search: Google Technology.

[4] *http://www.google.com/support/toolbar/bin/static.py?page=features.html&hl=en*, Google Toolbar: Toolbar Features.

[5] *http://toolbar.google.com/button_help.html*, Google Toolbar: About Google Toolbar Features.

[6] *http://www.uspto.gov/main/patents.htm*, United States Patent and Trademark Office official website.

[7] *http://www.webrankinfo.com/english/seo-news/topic-16388.htm*, January 2006, Increased Google Index Size?

[8] Arvind Arasu, Jasmine Novak, Andrew Tomkins, and John Tomlin, *PageRank Computation and the Structure of the Web: Experiments and Algorithms*, Technical Report, IBM Almaden Research Center (November 2001).

[9] John Battelle, *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, Penguin Group, New York, 2005.

[10] Pavel Berkhin, A survey on PageRank computing, *Internet Mathematics* **2** (2005), no. 1, 73–120.

[11] Celeste Biever, Rival engines finally catch up with Google, *New Scientist* **184** (2004), no. 2474, 23.

[12] Sergey Brin and Lawrence Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* **33** (1998), 107–117.

[13] Germund Dahlquist and Åke Björck, *Numerical Methods in Scientific Computing*, vol. II, SIAM, Philadelphia, to be published, http://www.math.liu.se/~akbjo/dqbjch9.pdf.

[14] Deborah Fallows, Search engine users, Pew Internet & American Life Project Report, January 2005.

[15] David Gleich, Leonid Zhukov, and Pavel Berkhin, Fast Parallel PageRank: A Linear System Approach, Technical report, WWW2005.

[16] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.

[17] Zoltán Gyöngyi, Hector Garcia-Molina, and Jan Pedersen, *Combating Web spam with TrustRank*, Proceedings of the 30th International Conference on Very Large Databases, Morgan Kaufmann, 2004, pp. 576–587.

[18] Taher H. Haveliwala and Sepandar D. Kamvar, *The Second Eigenvalue of the Google Matrix*, Technical report, Stanford University, 2003.

[19] Amy N. Langville and Carl D. Meyer, Deeper inside PageRank, *Internet Mathematics* **1** (2004), no. 3, 335–380.

[20] ———, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, 2006.

[21] Steve Lohr, This boring headline is written for Google, *The New York Times*, April 2006.

[22] Anuj Nanavati, Arindam Chakraborty, David Deangelis, Hasrat Godil, and Thomas D'Silva, *An investigation of documents on the World Wide Web*, http://www.iit.edu/~dsiltho/Investigation.pdf, December 2004.

[23] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, The PageRank citation ranking: Bringing order to the Web, Technical report, Stanford University, 1998.

[24] Lee Rainie and Jeremy Shermak, Big jump in search engine use, Pew Internet & American Life Project Memo, November 2005.

[25] Chris Ridings and Mike Shishigin, PageRank uncovered, Technical Paper for the Search Engine Optimization Online Community.

[26] Stefano Serra-Capizzano, Jordan canonical form of the Google matrix: A potential contribution to the PageRank computation, *SIAM J. Matrix Anal. Appl.* **27** (2005), no. 2, 305–312.

[27] Danny Sullivan, Nielsen NetRatings search engine ratings, *Search Engine Watch*, January 2006.

[28] Danny Sullivan and Chris Sherman, *Search Engine User Attitudes*, iProspect.com, Inc., May 2005.

The proof is in the pudding.

Opening a copy of The Mathematical Intelligencer *you may ask yourself uneasily, "What is this anyway—a mathematical journal, or what?" Or you may ask, "Where am I?" Or even "Who am I?" This sense of disorientation is at its most acute when you open to Colin Adams's column.*

Relax. Breathe regularly. It's mathematical, it's a humor column, and it may even be harmless.

Column editor's address: **Colin Adams,** Department of Mathematics, Bronfman Science Center, Williams College, Williamstown, MA 01267 USA
e-mail: Colin.C.Adams@williams.edu

Journey to the Center of Mathematics

Colin Adams

It was only 2:00 in the afternoon on a Thursday when I heard the front door to our little house on the Königstrasse slam.

"Axle, come at once," called Professor Lederhosen, as he rushed through the entry into his study. The professor was not a patient man, so I thrust my Jules Verne novel under the covers, leaped up from the bed, and descended the stairs. Timidly, I poked my head inside the door to his sanctuary.

"Axle, see this amazing book I have purchased," he exclaimed with his characteristic fervor. "It is a twelfth-century tome, written entirely in runic characters." He proffered a well-thumbed text bound in worn leather.

"Quite a find, Professor," I answered, feigning interest. "And now, I shall return to my chores."

"No, Nephew, you do not understand. This is the long lost book of Icelandic doggerel. This completes my collection of doggerel verses of the world." As he spoke these words, a small piece of parchment slipped from the book and floated to the floor.

"What could this be?" he exclaimed, as he reached down to pick it up.

"A meaningless scrap, perhaps," I replied hopefully.

"Oh, Axle, I doubt very much that a twelfth-century book of Icelandic doggerel would contain a meaningless scrap of paper."

He lifted the piece to within an inch of his nose and squinted.

"Why, this appears to have also been written in runic characters. Either that or it was written by a chicken."

I laughed aloud, only to stifle myself the instant I realized he was not joking.

"Who could have written this?" demanded the professor. "And what could it mean?"

"Perhaps it was written by the owner of the book," I said. "It says here on the inside cover 'This book belongs to Arnold Sackmuffin.' "

"Sackmuffin! It cannot be," said my uncle, his face turning ashen.

"Why? Who is Arnold Sackmuffin?"

"Axle, are you a complete illiterate? Only perhaps the greatest savant of Iceland during the years 1655–1659. He must have written this message."

"How can we figure out what it means?," I asked.

"In fact, Axle, I studied runic as a schoolboy. It was that or wood shop. I would wager that I can still translate with the best of them."

The professor proceeded to sit down at the table and slowly transcribe one letter after the other. Upon finishing, he stood slowly and read the result. "Plump dishes pinch the waffle man."

"What could this possibly mean?" he wailed. "Dishes cannot pinch anyone."

"Wait, Uncle," I exclaimed. "Flip it upside down and read it again."

As he did so, I stared over his shoulder, dumbfounded by the decrypted message that could now be read on the parchment.

> Descend if you dare into the crater of Sard's Theorem and you will attain the center of mathematics; which I have already done.—Arnold Sackmuffin.

"The center of mathematics!" exclaimed the professor as he fell into his chair, which immediately tipped over backward, sending him crashing to the floor. As I helped my uncle to his feet, I asked him what was this Sard's Theorem.

"Oh, Axle, for a student of mathematics, you know perilously little," he replied. "Sard's Theorem is the central mountain of differential topology. An inactive volcano, it can only be reached by a long slog through the desert of Differential Topology."

This was beginning to sound ominous.

"Well, this has all been fascinating, Professor. And now I shall return to my work."

"My dear Nephew, don't you see? We have discovered the secret of how to reach the center of mathematics, how to travel to its very core. This is an amazing revelation. We must prepare to leave at once."

"But I do not want to go. Honestly, Uncle, I do not understand mathematics. And I do not think I have what it takes to be a mathematician."

"All the more reason to go, my boy. We will learn from whence it comes. We will travel to its very source. And then perhaps, you will understand. Perhaps we will all understand."

"Very well, Uncle. I know there is no use in arguing with you. When do we start?"

"We leave tomorrow. But first we must gather up the necessary supplies and find a guide. Grab those textbooks off the shelf there. Dump them in this bag."

The next day, we booked travel on a steamer across the Analytic Ocean. All too soon, we found ourselves trudging through the desolate wasteland of Differential Topology following our newly hired guide Hansel, a mathematician from Stockholm who spoke neither English nor Swedish. But he was quite good with hand signals.

Several weeks did it take us to cross the desert. Much did I learn on that journey. I learned that the tangent space at a point to an m-dimensional smooth manifold is a vector space. I learned that a non-singular derivative at a point x of a smooth map f from R^n to R^n implies that f sends a neighborhood of x diffeomorphically onto a neighborhood of $f(x)$. I learned not to share a pillow with Professor Lederhosen, as he drools in his sleep. Finally, after many tortuous days of slow progress, we found ourselves at the foot of Sard's Theorem.

I looked up at the immense cone of the volcano that lay high above us.

"We must climb that?" I asked.

"Fear not, Axle," replied the Professor. "First we sleep. Then in the morning we attack Sard."

When I woke the next morning, Hansel had already packed the gear. Unable to communicate verbally, he was busily kicking me awake.

"Come, sleepyhead," called the Professor, obviously eager to begin the climb. "It is time for us to learn some mathematics."

Seven hours later, we stood on the rim of the crater, looking down into the dark maw of the mountain.

A rock on the rim caught the professor's attention. "See here," he called excitedly. "This symbol hacked into the stone. '∞.' That is the mark of Sackmuffin. This must be the way down."

We threaded our way through a field of boulders the size of cottages, as we descended into the crater. Eventually we arrived at a ledge from which darkness was all we could see below.

"There is no way further down," I said to the professor. "Too bad. This has been exciting, but now we must return home."

"Nonsense, Axle, have you never before become entangled in a proof, unable to continue forward? Do you just give up? Do you throw in the towel?"

"Yes, Professor, that is what I do."

"That is why you are not yet a mathematician, my boy. We will make one of you yet."

"Do not concern yourself with me, Professor. I am happy the way I am."

My uncle ignored me.

"We continue down. All we need is our ropes and ingenuity."

Ten minutes later, I found myself dangling from the ledge by a rope, as I was lowered slowly into the very depths of the core of the silent volcano. As I turned on my electric light, I could see the details of the proof of Sard's Theorem. Here on the right was Fubini's Theorem, which is critical in the proof. And there on the left was the descending sequence of closed sets that form the core of Sard's Theorem.

As I looked upward where the professor and Hansel stood above me, I could see the light of the outside world dwindling quickly. By the time I reached the bottom, signs of the surface could no longer reach me. Within the hour, all three of us stood in the small pool of light cast by my electric lantern.

"Now what, Professor?" I asked.

"See, here, Axle. The mark of Sackmuffin on the wall of this passage. Follow me." He took the light from my hand and proceeded down the passage. Hansel followed. I quickly took up the rear, not wanting to be left behind in the darkness with the fearsome matrices of partial derivatives that surrounded me.

For several days, we traveled ever deeper into the heart of mathematics. From differential topology, we entered the world of point set topology. Many a beautiful basis we passed, sparkling in the reflected light from our electric lanterns. Pathological topologies of complexity too twisted to describe appeared before our eyes. If I had only had the time and courage to write it all down, I would have had many a research paper to my name.

After several days of downhill travel, I noticed the passageway had taken a distinct upward incline. The mathematics around us was no longer becoming simpler as we traveled forward. Rather, it was taking on an ominous complexity. This continued for a few hours, until I could contain my concern no longer.

"Professor, we are not getting closer to the center of mathematics. We are getting farther away with each step."

"You are more discerning than I would have given you credit for, my boy. Our path has taken us into algebraic topology. It appears we will have to continue upward for the time being. I have every hope that the path will crest soon."

I looked with dread upon the tunnel wall, where I saw Čech cohomology groups with coefficients in presheaves. As we continued, we soon found ourselves in the midst of the Relative Hurewicz Isomorphism Theorem. Turning a bend, I saw what appeared to be a Leray-Serre spectral sequence. This was too much for me.

"Uncle, I beseech you. We must turn back. We cannot go on this way. If I am not mistaken, we are now deeply enmeshed in the field of homotopy theory. If we continue, we will just become more entangled in this morass and we shall never find our way out."

The professor turned to me, flashing his light directly into my eyes.

"Axle, what if Norman Steenrod had reacted this way when he first confronted homotopy theory? What if J. H. C. Whitehead had turned tail and run when he came face-to-face with absolute neighborhood retracts? What if Heinz Hopf had buried his head in the sand when he discovered hopficity?"

I motioned to the fiber bundles hanging from the ceiling and jutting up out of the floor.

"But professor, look at our surroundings. We cannot possibly hope to understand what is going on here, let alone anything more complicated."

"Axle, at this point, we have no choice. We have come too far to go back. Our water supply will not last long enough for us to return the way we have come. We must continue in the hopes of finding water. Do not be frightened by a few fiber bundles. I am certain the path will level off soon, and then begin its descent."

We continued along the passage for several more hours, my throat becoming more parched with each step. I ached for just a sip of water. But Hansel guarded the water supply zealously, and warned me off with guttural threats and vigorous hand gestures whenever I grabbed for the canteen. When I had finally given up all hope, and was just looking for a comfortable spot to lie down, the path did level off. And then it began to descend. I saw an Eilenberg-Maclane space embedded in the wall. And there was some homology with Z coefficients. My flagging spirits rose concomitant with the descent in our altitude. Then suddenly, the passageway opened up into a huge cavern, from the far end of which there appeared a blue green luminescence. As we crossed the broad cavern floor, approaching this strange light, the stone at our feet was replaced with sand. I could feel a breeze blowing moist air across my face. I began to run, my colleagues just behind me, until the vista opened upon a miraculous sight. It was a giant underworld ocean extending to a broad horizon.

"What could this be?" I asked, as the three of us stood awestruck on the sandy beach, waves lapping at our toes.

"It is Morse Theory," replied the professor.

"Thank God for it, whatever it is," I said, as I fell down on the sand and submerged my head in the water, drinking greedily.

After replenishing ourselves and resting by the shore, we wandered along the beach until we came across a grove of logarithms growing out of the sand. Cutting them down, we lashed them together to construct a raft. Shortly, we found ourselves on the Morse Sea, making good time as a strong breeze blew our makeshift sail. Hansel steered with a rudder we had attached to the stern.

Quickly, the shoreline receded in the distance, and all that filled our horizon was water in every direction. The cavern ceiling was so far above us that we could no longer make it out through the hazy clouds floating overhead. The water itself gave off the light that illuminated this underworld sea.

After three days of sailing, there appeared a large land mass rising up out of the water. As we approached, we recognized it to be an island, with a peak so high it disappeared into the clouds overhead.

"Do you know it, Professor?"

"Yes, Axle, it can only be Mazur's Theorem, which uses Morse theory to prove that a smoothly embedded n-dimensional sphere sitting inside an $n + 1$-dimensional sphere always separates it into two $n + 1$-dimensional balls."

I stared up as its inspiring cliffs above sailed by, and then watched it sink into the horizon behind us as we progressed, alone again on the flat expanse of water. Suddenly Hansel started to gesticulate excitedly, pointing ahead of us. Darkening clouds approached from the west. The wind began to pick up.

"It appears we are in for some weather," said the professor calmly.

"But Uncle, from whence would come weather on an ocean deep in the interior of mathematics?"

"Perhaps some logical inconsistencies are passing through the bowels of the earth, inconsistencies that are causing indigestion and that may require some adjustments to accommodate."

I was not sure to what extent his digestive analogy was meant to be taken literally. I preferred to think of mathematics as an inanimate object rather than a living breathing creature that was having stomach trouble.

The howl of the wind picked up, and my uncle's further words were lost in the tumult that descended upon us. Our tiny craft was tossed on towering waves, as we clung to the mast. The temperature dropped by 30 degrees. Rain, flying horizontally, soaked us to the core. Hansel grabbed a rope with his free hand, and tied us to the mast. I watched with horror as the rucksack containing our provisions was washed overboard. The storm continued unabated for two days. At some point I lost consciousness, slumped against the mast.

I was lying on a beach, face in the sand, when I regained my senses. It was the sound of good old Hansel yelling gibberish that convinced me that I lived still. Looking up, I saw him pointing to an opening in the rock wall far down the beach. Even from a distance I could make out the mark of Sackmuffin chiseled into the stone.

We gathered the belongings that had survived our voyage and set off down the passage. Almost at once we came to the end of the tunnel and found ourselves face-to-face with a complicated counterexample that had somehow become dislodged from the roof of the tunnel, and fallen to block our way.

"Professor, this is a dead end."

"Yes, nephew, this must have come down since Sackmuffin passed this way, perhaps even during this last storm."

"But what can we do? Is this the end of our journey?"

"Do not give up yet, my boy. We will have to blast our way through."

"How do we do that?"

He reached into his rucksack.

"I brought along a few axioms that should do the trick. That counterexample will cease to exist if we change the axioms. It will be reduced to rubble."

My uncle dropped to his knees and with his hands, dug a small hole underneath the counterexample. Then he gingerly squeezed the axioms in. Pushing in the end of a fuse, he then unrolled it until we were again standing on the beach, one hundred yards away.

"Axle, would you like to do the honors?" he asked, offering me a match. I struck it on the sole of my shoe and lit the fuse.

"Quickly," he said, "Get on the raft. We can float out another hundred feet for safety."

Just as we had pushed off the beach, there was a deafening roar. Rocks shot out of the tunnel and cascaded around us. A large wave lifted us up in the air, as the rock wall in front of us crumbled into the sea.

"Those axioms seem to be doing a

bit more damage than I had expected," yelled the professor.

A huge crevasse opened in the beach and the ocean began to pour in, sweeping us with it. We found ourselves shooting down a large opening, riding the wave at the front of the massive moving body of water.

"Hang on," yelled my uncle as the water swirled about us, and we rocketed down the passageway.

I glanced at the mathematics whizzing by and my heart leapt.

"Look, Professor, sines and cosines. It is trigonometry! We must be getting close."

We passed through high school algebra, word problems, and exponents. And then up ahead was an outcropping of rock alongside an opening in the wall. It was the side-angle-side theorem from geometry.

"Axle, steer the raft toward that theorem and get ready to jump," yelled the professor. "This may be our only chance."

At the last possible second, we leapt for the rocks. The raft careened off the theorem and splintered into pieces. The three of us pulled ourselves to safety, sitting on the lip of a small proof about parallelograms. The water continued to rush by. I turned to see where we were, and there on the wall above us was the mark of Sackmuffin.

I leapt to my feet.

"We must be almost there," I exclaimed, as I started down the passage. Hansel and my uncle got up to follow. I found myself in the midst of arithmetic. Good old arithmetic. Even I could do arithmetic—adding fractions, long division, multiplying three-digit numbers.

Grinning from ear to ear, I continued down the passage as the surrounding equations became even simpler. Here was the multiplication table for integers less than 10 and here was addition of small numbers. This must be it, I thought. I am almost there. I turned the corner, pointing my light ahead of me and stopped dead.

I stood dumbfounded as I stared at the center of mathematics. Was it algebra? Was it topology? Was it number theory? None of the above. In fact none of anything. There was nothing there.

Hansel and the professor collided with me as they rushed around the corner.

"How can this be, Professor?" I asked plaintively. Hansel made desperate hand signals.

The professor stood silently, his brow creased in thought.

"Professor, there is nothing here. The center of mathematics is empty."

Suddenly the professor looked up, a smile on his face.

"Ah, Axle, that is the answer. There is nothing here, because that is how you create the numbers, from nothing."

"What do you mean?"

"First there is nothing. That is the set $\varnothing$, the empty set."

"Yes, I see we have the empty set here. Since there is nothing here."

"But Axle, now we have something. A set that is empty."

"What do you mean?"

"We now have something. It is the set called the empty set."

"Yes, but it isn't there."

"Oh, it assuredly is there. It is something, is it not? It is a set."

"Yes, it is a set."

"Isn't a set something?"

"Yes. . . . "

"So now we have created something! Something from nothing. Does it not make you feel like a god?"

"Not particularly. And what about the rest of mathematics? Now, we have just one set, a set which contains no elements."

"Yes, but you see, my boy, now we can take the set that has this set as element. This is a nonempty set. It contains one element, which is this set. We write it like this. With the toe of his boot, he drew a picture in the sand at my feet: $\{\varnothing\}$"

"Sounds a bit self-referential to me."

"So what? Then we have the set that consists of the empty set, and the set that has element the empty set. This forms a two-element set $\{\{\varnothing\},\varnothing\}$."

"You are making my head spin, Professor. I need to sit down."

But as I went to do so, the ground began to move. A low grumble echoed through the chamber.

"I fear, Axle," said the professor, "that in stating the solution to this puzzle, we have set in motion powers beyond our ability to comprehend. Run!"

The ground began to shake and split. Steam shot up out of the cracks that were forming.

"Quick, into this lemma," yelled the professor. He leaped over the edge, pulling me with him. Hansel jumped in after us, and there was a deafening roar. Logical arguments burst around us. The lemma was thrown upward with incredible force, the three of us holding on for dear life. As the pent up forces exploded, we found ourselves shooting up a large crack in the rock that had been created by the cataclysmic eruptions. We rocketed upward, covering in several minutes a distance that had taken us months to traverse. Suddenly we cleared the underworld and shot out into the real world, right out the top of the Riemann-Roch Theorem.

As we clung to the lemma, it flew through the air, higher and higher before its trajectory peaked. Then we started earthward again, and I prayed for deliverance. As fate would have it, the lemma was wide enough to catch some air resistance, and we were lucky enough to land with a splash, unharmed, in the Algebraic Ocean, more than a thousand miles from where we had embarked on our adventure.

There we were picked up by some commutative algebraists who had been out trawling for lemmas. Within a short time, we found ourselves back home, hailed as the heroes who had reached the center of mathematics.

Of course, to this story there is an aftermath. Following such an adventure, it was difficult for each of us to adjust to the everyday world. The professor eventually gave up mathematics altogether. He spends most of his time writing doggerel in runic and reading it to the chickens he keeps beyond the house. Hansel dabbled with puppetry for a while, but eventually he decided to go to mime school. And me, I have decided to become a logician. There was something special I felt at that instant when I understood how all of mathematics could come from nothing. And I just hope to capture that feeling again, even if only for a moment.

Airborne Weapons Accuracy: Topologists and the Applied Mathematics Panel

JOHN MCCLEARY

Mathematicians love stories, especially stories about the major figures in their speciality. For example, the image of a young Gauss putting down a slate with the sum of 1 to 100 before his grade school classmates is a favorite of most mathematicians. Stories of a recent vintage are passed along by word of mouth and so undergo a whisper-down-the-lane effect, leaving their reliability in question. I heard one such story, from multiple sources, about Hassler Whitney's work in World War II on aerial weapons accuracy in dogfights. Apparently, this formidable geometer had proved that a gunner should aim behind the plane he pursued in order to shoot it down. The story went on that no one believed Whitney, but an unnamed assistant accompanied a gunner on a flight that engaged two enemy planes. The gunner fired in front of the second plane, as intuition would dictate, but he hit the first plane, proving Whitney correct.

The story is entertaining, but it should be verified before being added to the historical record. I decided to consult the records of the research done during World War II by mathematicians, particularly by the group at Columbia University associated with the Applied Mathematics Panel, which included Whitney and Saunders Mac Lane. Although I didn't find the story I was looking for, the story I did find in documents in the National Archive gave a deeper look into wartime activities of some of my favorite mathematicians and a detailed picture of the process of wartime consultancy.[1] Of particular interest, in response to a question posed by June Barrow-Green, I found that the work of research mathematicians did not go unread and ignored: there was contact between these active, sometimes passionate, academicians and the military command, resulting in changes of policy.

The Applied Mathematics Panel

In the years leading up to World War II, Applied Mathematics was not well represented at U.S. universities. Centers at Brown and New York University (NYU) and the teaching of statistics at Princeton and Columbia were exceptions. As the war approached, certain scientists anticipated the need for expert technical support and sought ways to put in place the organizational structures needed to make such support possible. One structure, already in place since 1916, was the National Research Council (NRC) established by President Wilson, which "brought science to bear on World War I" [5]. By 1940, the NRC had outgrown its initial charges and had become committed to projects that would make it, as an organization, too slow to react if a new war effort of a different sort was launched. The Council of National Defense, consisting of the Secretaries of War, Navy, Interior, Agriculture, Commerce, and Labor, was authorized to create subordinate organizations for "support in special circumstances," and so the National Defense Research Committee (NDRC) was established in 1940 with the approval of President Roosevelt [4], who appointed Vannevar Bush (1890–1974), then President of the Carnegie Institution of Washington, to be its chairman. A year later, Executive Order No. 8807 established the Office of Scientific Research and Development (OSRD), a reorganization of the NDRC that emphasized carrying a research project from planning through the engineering stage. The OSRD also provided a place for the research programs for the Army, Navy, NDRC, and National Advisory Committee on Aeronautics (NACA) to discuss common goals, and it maintained parallel committees on weapons and medicine. In 1942, about a year after the declaration of war, the OSRD was reorganized with a more transparent and

Send submissions to **David E. Rowe,**
Fachbereich 17—Mathematik,
Johannes Gutenberg University,
D55099 Mainz, Germany.

Talk given at the joint international AMS-DMV-ÖMG meeting in Mainz, Germany, June 17–19, 2005.

[1]My thanks to Loyd Lee, and Kathleen Williams for suggesting the National Archive, and Marjorie Ciarlante at the National Archive for making my visit a success.

correlated administrative structure. The five divisions of the NDRC with their manifold sections were split into 19 divisions with more focussed charges. For example, "the field covered by Division 7 included directors for airborne, land-based and shipborne uses; servomechanisms for transmitting data to and from the guns and for elevating and traversing the guns; special range finders and testing equipment for assessing aerial gunnery and for evaluating the performance of directors, and the development of gun sights for special purposes." More general panels of experts were also organized under OSRD, including the Applied Psychology Panel and the Applied Mathematics Panel (AMP), which was set up with the goal "... to aid NDRC divisions and Services by providing assistance in the handling of problems needing the special training of mathematicians." Warren Weaver (1894–1978), Director of the Natural Sciences Division of the Rockefeller Foundation, was tapped by Bush to head Section 7.5, Fire Control Analysis. Analytical studies in fire control and in other areas required more mathematical experts and that need was met by the AMP, also headed by Weaver [4].

My goal in this article is to give an account of the engagement of a particular mathematician, Hassler Whitney (1907–1989), in a particular problem, fire control for rockets (air-to-air), and the subsequent results of his efforts. Whitney, a consultant and not an administrator, is a typical case. This account is based on research in the US National Archives where reports commissioned by the OSRD and associated documents are accessible. We will follow a problem from official informal request to the AMP, its assignment to a consultant, through the research, to delivery of results. From this we can get a sense of how mathematicians contributed to the war effort. I include a short account of some of the mathematics involved to illustrate the kind of analysis, and later exposition, that pure mathematicians produced for the AMP.

What Is Fire Control?

The main problem of fire control is mechanical: an operator (called the bomber B) controls both a sight and a gun. Since a projectile requires time to cover distance and the target (the fighter T) is moving, it would be efficient if the mechanism for sighting included a means to calculate, from input data (such as the angular velocity of the target, its range, and the time of flight of the projectile), the direction that the gun must be pointed so that the projectile will meet its target. In simple realistic cases, this problem is susceptible to mathematical analysis which would suggest features in the design and use of equipment. This analysis was the focus of Section 7.5 at Columbia's Applied Mathematics Group (AMG-C).

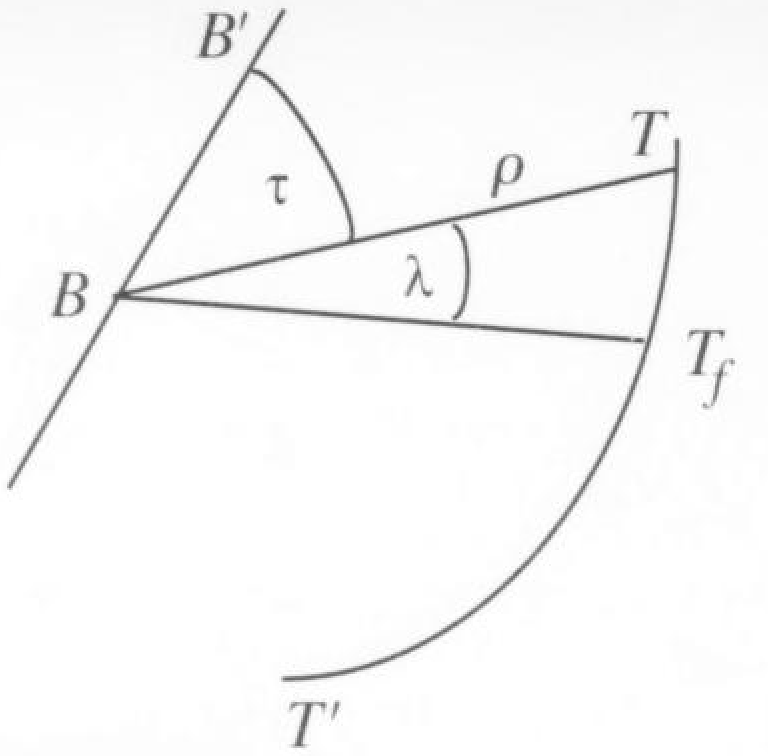


FIGURE 1

The prime case for improvements in fire control systems was data on antiaircraft fire from the bombing of England. Weaver recalled, "... in October 1940 it was estimated that at least 10,000 rounds of fire from 3-inch antiaircraft guns were expended for each plane shot down in the London area. Even that figure, moreover, is almost certainly too low."

To understand the kind of analysis involved, consider the path of a fighter TT' relative to a bomber at B (Figure 1). The bomber can be moving if defending from a plane or a ship, or not moving if on the ground at an antiaircraft gun; it is the relative motion that is relevant. At time t the fighter is at position T and the bomber wants to fire a shot to arrive at the fighter some time later, when the fighter is at the future position T_f. The angle λ, called the kinematic lead angle, is the direction in which the shot should be fired to meet the fighter at T_f. It depends on features of the motion of the fighter, such as present range from the bomber (length of BT), and its angular velocity, as well as the ballistic properties of the bomber's projectile. Choose a reference line BB' from which to measure angles and view the path in polar coordinates:

$\tau = \angle B'BT$ angular position of the target

$\rho = BT$ present range of the target

$u =$ time of flight of projectile to the present position T

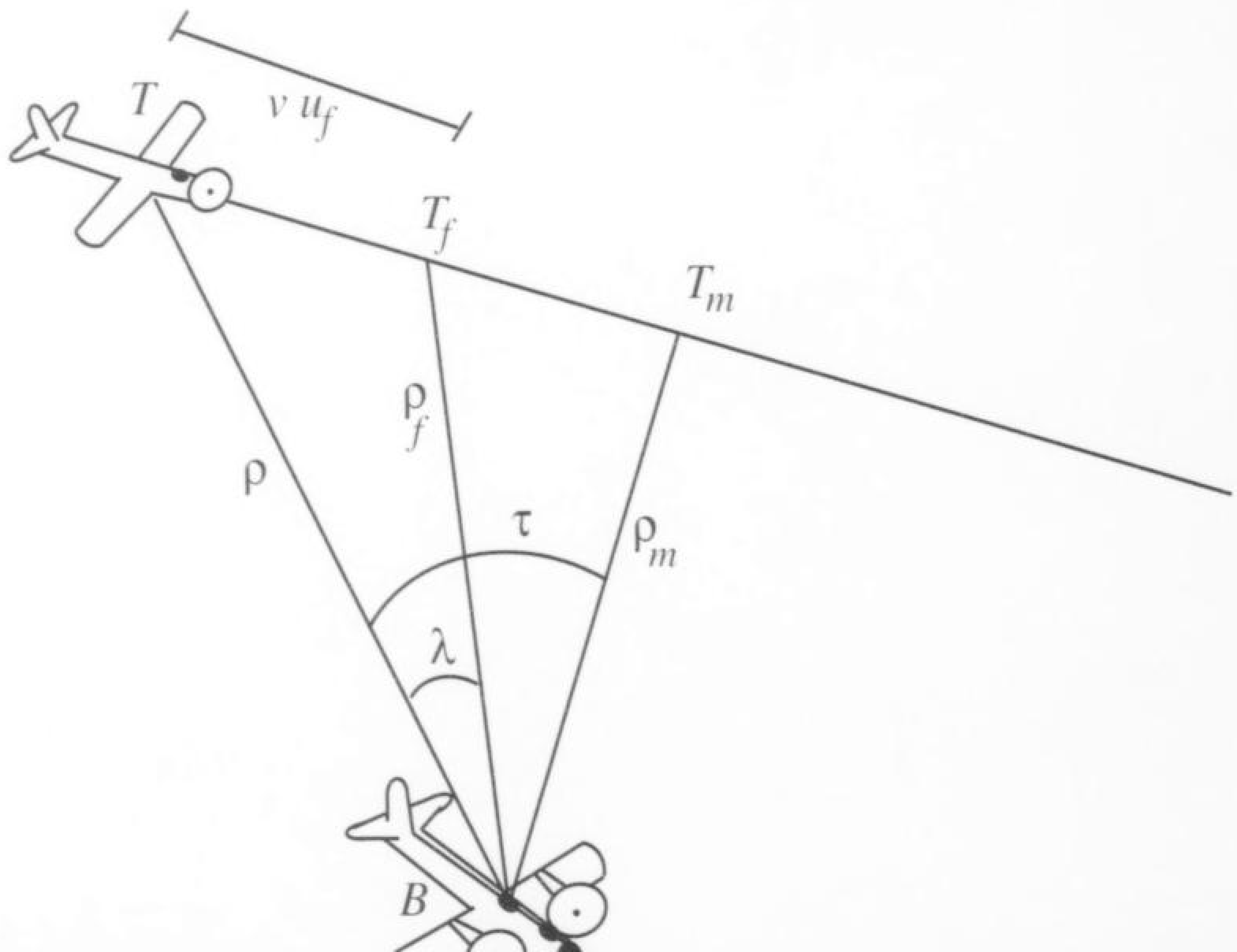


FIGURE 2

u_f = time of flight of projectile to the future position T_f;
= time required for the target to go from T to T_f.

A fire control system takes as input flight times of projectiles u and target rates, such as $d\tau/dt$ and/or $d\rho/dt$, obtained by the operator through the apparatus, and gives as output an approximation of the kinematic lead angle λ.

As an example of the kind of analysis that is possible in this setting, let us assume the fighter is following a straight line path at constant speed relative to the bomber (Fig. 2). Let T_m denote the point along the flight path closest to the bomber, with $BT_m = \rho_m$ and BT_m perpendicular to TT_m; T_m is called the crossover position. Let ν denote the constant velocity of the fighter. It follows that the distance TT_f equals νu_f since u_f is the time of flight of the projectile from B to the future position. To study the kinematic lead angle, we can apply the law of sines:

$$\frac{\sin\lambda}{TT_f} = \frac{\sin(\angle BTT_m)}{\rho_f} = \frac{\cos\tau}{\rho_f}.$$

The last equality follows from the right triangle ΔBT_mT. Thus

$$\sin\lambda = TT_f\,\frac{\cos\tau}{\rho_f} = \frac{\nu u_f\cos\tau}{\rho_f}$$
$$= \nu\cos\tau\,\frac{u_f}{\rho_f}.$$

From $\tan\tau = TT_m/\rho_m$, we can take the derivative, which gives us

$$\frac{d\tau}{dt}\sec^2\tau = \frac{\nu}{\rho_m},$$

since ρ_m is constant and TT_m changes at rate ν. Then $d\tau/dt = (\nu/\rho_m)\cos^2\tau$. From the right triangle ΔBT_mT we know that $\cos\tau = \rho_m/\rho$, and so

$$\frac{d\tau}{dt} = \cos\tau\frac{\rho_m}{\rho}\cdot\frac{\nu}{\rho_m} = \frac{\nu\cos\tau}{\rho}.$$

Substituting this equation into the expression for $\sin\lambda$, we get

$$\sin\lambda = \nu\cos\tau\cdot\frac{u_f}{\rho_f}$$
$$= \frac{\nu\cos\tau}{\rho}\cdot\frac{\rho u_f}{\rho_f} = \frac{u_f}{\rho_f}\cdot\rho\frac{d\tau}{dt}.$$

If we let W and W_f denote the average velocities ρ/u and ρ_f/u_f, then

$$\sin\lambda = (W/W_f)\;u\,d\tau/dt.$$

For small angles λ, the values of $\sin\lambda$ and λ are approximately the same, and for projectiles like rockets, average velocity is essentially independent of range, and so $W\sim 1$. This motivates the choice of $\lambda\sim u\cdot(d\tau/dt)$, which is both a good approximation and calculable.

An application of this analysis appeared in the Memorandum [1] written by Saunders Mac Lane (1909–2005) for AMG-C, *An Introduction to the Analytical Principles of Lead Computing Sights*. Compare the flight path of a fighter going on a straight line with one following a pursuit curve, that is, a curve of constant speed on which the fighter is always pointed at the bomber, which is also following a flight path at a constant speed. In this case, the kinematic lead angle computed from the straight line path is ahead of the lead angle for the pursuit curve (Figure 3). However, the approximation $\lambda = u\tau'$ gives the same value for both paths. To compensate for this situation, a bomber should take aim at a point behind the center of the fighter, a counterintuitive conclusion. Mac Lane described this consequence of the computation in his autobiography [2]: "How should a machine gunner aim at an attacking fighter plane? The fighter is approaching the bomber, but the bomber is moving forward at the same time. The resulting rule is that the gunner should aim toward the tail, which is opposite from the rule for hunting ducks, where the rule for a stationary hunter is to aim ahead of the ducks. A major part of our problem was properly training machine gunners to aim toward the tail." My Whitney story sounds like a misattributed and modified version of this result.

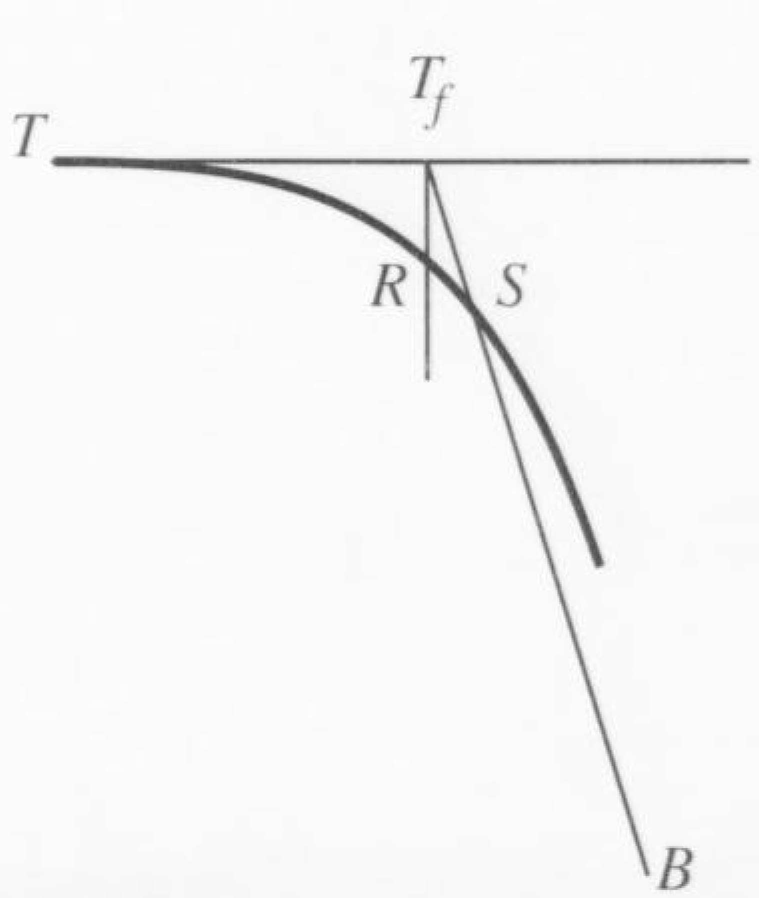


FIGURE 3

A more complicated problem tackled by AMG-C is known as the tracking problem: what kind of design is best for the mechanism that controls a fire control system, that is, the device that translates the action of the bomber in tracking a target into the motion of the gun?

Problems like this arrived at the desk of the AMP through Informal Requests, which were transmitted to the group from the armed forces through Section Chiefs, and then assigned to consultants who were university professors recruited and employed by the OSRD. One such request is the following:

```
Informal Request #18:
Analytical Study of the
Tracking Problem
```

Transmission: Dr. H. C. Wolfe. Section 7.2 (NDRC) at AMG-C

> The extensive use of laboratory machines to study tracking problems, at the Franklin Institute and elsewhere, indicates the desirability of an analytical study of the nature of the tracking problem and of machines for tracking tests. The importance of such study has been recognized by Dr. Hassler Whitney (largely on his own initiative).

This request was assigned to Whitney and led to the AMP Note No. 21 titled *Tracking and the fire control problem*.

Hassler Whitney

Hassler Whitney studied music and physics at Yale University, graduating in 1929. Going on to Harvard University, he studied mathematics, obtaining a PhD in 1932 under G. D. Birkhoff, with a thesis on graph theory and the four-color problem. Appointed instructor at Harvard in 1930, he was promoted to assistant professor in 1932, associate in 1940, and full professor in 1946. In 1931–32 he visited Princeton as a National Research Council Fellow. Whitney left Harvard in 1952 to become a permanent member of the Institute for Advanced Study in Princeton.

In the late 1930s, Saunders Mac Lane was also an assistant professor at Harvard; this connection brought Whitney into the Applied Mathematics Group at Columbia. Whitney was employed as consultant to AMG-C officially 1 November 1943 to 30 September 1945 (also working 6 June 1945 to 31 October 1945 without pay), according to Sard's final report on the AMG-C. In a letter to Weaver a year into his consulting, Whitney expressed the kind of commitment he had for the project and his sense of what was needed to bring it to completion: "If the rocket and bombing sighting problems are really to be pushed (they must be!), it is essential that there be sufficient coordinated work on the problem. . . . With a good group to study all angles of the problem thoroughly, plenty of laboratory facilities for building, and planes and pilots available to carry out a (fairly large) variety of tests—tests to help both in the theoretical stage and in later stages—miracles could be accomplished, and quickly."

What Whitney had in mind was evident in remarks of Mina Rees (1902–1997), technical aide and executive assistant with the AMP, who recalled in 1980 [3], "That part of the program of the Applied Mathematics Panel that was concerned with the use of rockets in air warfare was primarily the responsibility of Hassler Whitney, who served as a member of the Applied Mathematics Group at Columbia. He not only integrated the work carried on at Columbia and Northwestern in the general field of fire control for airborne rockets but maintained effective liaison with the work of the Fire Control Division of NDRC in this field and with the activities of many Army and Navy establishments, particularly the Naval Ordnance Test Station at Inyorkern, the Dover Army Air Base, the Wright Field Armament Laboratory, the Naval Board of Ordnance, and the British Air Commission." Whitney apparently had a feel for this kind of work: as Rees remarked elsewhere [8], Hassler Whitney "turned out to have an absolute genius for airplane problems from guidance studies."

Rees once accompanied Whitney to perform some tests on a training machine for pilots. Whitney was developing mathematical principles to discover the best techniques for aerial gunnery, and he was going to test his theories on a flight simulator. Rees and Whitney each got into a simulator, and "he shot down all his planes and I only shot down a couple. It was quite clear that he really knew all about this mechanism which I didn't think he knew anything about. So [there were] a lot of people you didn't really appreciate until you saw what they'd actually done."

Whitney's Analysis

In a fire control system, much depends on the flow of information from the operator (the bomber) and the gun. If it is particularly difficult to move the sighting device to follow a target, then a computing device determining the kinematic lead angle will deliver an out-of-date prediction while the mechanical turning takes place. A lead computing sight (lcs) incorporates the computing device and the sighting mechanism—when the operator sights the target in the cross-hairs, then the lcs has compensated for the kinematic lead angle and the gun is pointed to meet the target at the future position. The theory of such sights lies in the kind of tracking in use. Whitney assumed that the operator controlled a handlebar, a stick "pivoted several feet in front of the operator's chest, and pointing towards him; it may be moved up and down and to the right and left." The handlebar controlled the movement of the sight, part of a fire control system which aimed the gun appropriately. Three sorts of tracking were studied by the AMG-C:

- Direct tracking. The position of the sight line depends only on the position of the handlebar.
- Velocity tracking. The velocity of the sight line depends on the position of the handlebar.
- Aided tracking. The position and velocity of the sight line depend on the position of the handlebar.

In direct tracking, the handlebar must be moved constantly to follow a moving target, whereas in velocity tracking, if the line between the operator and the target is turning at a fairly constant rate, then only small handlebar motions are required. However, Whitney wrote, this kind of tracking requires more skill and training "since the sight line motion is the integral of the handlebar motion, the tracking can hardly help but feel more complicated to the inexperienced operator."

Such practical concerns were part of the fabric of Whitney's report. His analysis was written with the armed forces as audience, in wartime when effort was meant to give concrete results, and by an eager researcher. He recommended certain systems for ease of use as a "psychological advantage." He could imagine the bomber vividly: "We remark, finally, that an important consideration in assessing a tracking mechanism is the mental effort it requires on the part of the operator. He may have other matters to attend to at the same time, such as ranging and pressing a trigger. In the heat of combat, one cannot expect a gunner to go through mental gymnastics." The analysis Whitney made was presented in general terms, incorporating as many features of a general fire control system as were useful. The basic variables he denoted by

$\sigma(t)$ = sight direction
$\lambda(t)$ = gun direction
$\mu(t)$ = handlebar position
u = time of flight setting

It follows that the lead kinematic angle λ satisfies $\lambda = \gamma - \sigma$, and so our previous discussion applies. Whitney assumed that γ and σ were determined by "linear differential equations between the above quantities and μ, with coefficients depending on u only." For example, aided tracking has an equation of the form

$$\frac{d\sigma}{dt} = A\frac{d\mu}{dt} + B\mu.$$

The basic differential equation governing the lead kinematic angle is derived in Mac Lane's memorandum [1]

$$cu\frac{d\mu}{dt} + \mu = u\frac{d\sigma}{dt}.$$

Equations of this sort resemble the equations for smoothing circuits, where the unknown function is a kind of smoothed version of the forcing function $u\sigma'$. They arise elsewhere, in electrical applications such as a resistance-capacitance circuit. The effect of smoothing is to take jerks in direction and exponentially dampen their effect on the system. Whitney massaged the various classes of equations in a rather technical chapter leading to an analysis of tracking gun error, the difference between the ideal direction and the computed direction of the gun at firing. He ended with, "A general conclusion that may be drawn is that *any change in a turret mechanism designed* *to allow faster turret motions, thus improving tracking, could easily give worse gun positions.*" (Whitney's emphasis.)

Inyokern, California: Meeting of the Advisory Panel, July 18–20, 1945

Whitney brought his analysis, written up in various reports that eventually led to his monograph, AMP Note No. 21, to the Naval Ordnance Test Station (NOTS) about a month before the end of the war. In a diary of the visit that he kept for the leadership of AMP [6], he noted his surprise, on arrival, to find that the working meeting he expected with top members of the Advisory Panel to research and development at NOTS had turned into a school to study the present situation in fire control systems and their assessment. He expressed his impatience, "Being somewhat perturbed at the program, and yet not being at all sure of what could be accomplished by such a large group of people with various degrees of specialization, HW did some lobbying to make the conference accomplish something. . . . However, HW did not feel himself in a position to take matters into his hands to any great extent, did not have much experience (if any) in this kind of situation, and was generally lazy and shy about talking too much."

Whitney hoped the meeting would produce a recommendation for an interim rocket sight. Although a schedule of talks had been laid out in advance, Whitney was added to the schedule, giving a talk, "designed to help the members and spectators (thirty or more in all) fit the various sights into a pattern based on a simple analytical theory." The four sights being tested at NOTS were presented and demonstrated as part of the meeting. Although assessment was the goal, Whitney found the talks on assessment "dull"; the talk in which Comdr. Hayward ranked the sights began the "fireworks." By the afternoon of the 20th, recommendations were drawn up based on the experience of pilots, not the kind of analysis Whitney provided. He took the matter into his own hands and got a smaller group together that evening. For this group, he took two sights that had been rated earlier as 2 and 3 in the list and, after analyzing them "from all angles," suggested that their features be combined. To Whitney's surprise, the ranking commanders decided "at once" that this would be the interim sighting system. His efforts to ground the discussion, and his willingness to do some social maneuvering, like calling special meetings, got an unexpected result.

Summary

Whitney cannot have swayed a panel with purely technical arguments. His attention to detail in his report, his diary, and the descriptions by others of his involvement give a picture of a committed contributor to the war effort, someone willing to do what was necessary to make, in this case, improvements in weapon systems and bring the war to a close. Other work on sights was also successful. Warren Weaver reported in his autobiography that French fliers were particularly happy with a sight for submarine bombing that was a product of analysis and development by OSRD. In this study, we get a sense of the kind of activities that American mathematicians engaged in during World War II, both mathematical and social. Whitney presents an interesting case study of personal depth in pursuit of goals. This was probably typical. The work of the AMG-C was recognized for its contributions, receiving the Naval Ordnance Development Award "for distinguished service to the research and development of Naval Ordnance" [3].

REFERENCES

[1] Mac Lane, Saunders, An Introduction to the analytical principles of lead computing sights, March 27, 1944, AMP Memo. 55.1, AMG-C 137, OSRD Report 4037.

[2] Mac Lane, Saunders, *A Mathematical Autobiography*, A K Peters, Wellesley, MA, 2005.

[3] Rees, Mina, The mathematical sciences and World War II, *Amer. Math. Monthly,* 87 (1980), 607–621.

[4] Stewart, Irvin, *Organizing Scientific Research for War: The Administrative History of the Office of Scientific Research and Development*, Little, Brown and Co., Boston, 1948.

[5] Weaver, Warren, *Scene of Change: A Lifetime in American Science*, Scribner's, New York, 1970.

[6] Whitney, Hassler, July Meeting of the Advisory Panel, Inyorkern: Diary of Hassler Whitney, July 25, 1945, AMG-C working paper No. 460.

[7] Whitney, Hassler, Tracking and the fire control problem, September 14, 1945, AMP Note no. 21, AMG-C 441, OSRD Report 5680.

[8] Williams, Kathleen Bloome, *Improbable Warriors: Women Scientists and the U.S. Navy in World War II*, Naval Institute Press, Annapolis, MD, 2001.

Department of Mathematics
Vassar College
Poughkeepsie, NY 12604-0069
e-mail: mccleary@vassar.edu

Dancing Elves and a Flower's View of Euclid's Algorithm

SUSAN GOLDSTINE

This column is a place for those bits of contagious mathematics that travel from person to person in the community, because they are so elegant, suprising, or appealing that one has an urge to pass them on.

Contributions are most welcome.

Please send all submissions to the
Mathematical Entertainments Editor,
Ravi Vakil, Stanford University,
Department of Mathematics, Bldg. 380,
Stanford, CA 94305-2125, USA
e-mail: vakil@math.stanford.edu

The Dancing Elf Puzzle

"You know those Fibonacci patterns in sunflower spirals? Do you have any idea where they come from? So far, I haven't been able to find anyone who can explain them to me, and it seems like the sort of thing you might know."

I had been asking some version of this question whenever it occurred to me for a year or so. At the time, the dark ages of the 1990s, before everything was two clicks away on the Internet, I was intent on learning why these curious patterns exist but was having no luck tracking down sources on the topic. As I suspected, biologists and mathematicians had long since explained the Fibonacci phenomenon. I just couldn't *find* the explanations. Because I had more pressing research, I was hoping that one of my friends had done the work for me.

On one occasion, I was questioning Ravi Vakil, who replied, "It's funny you should ask that. I just heard this puzzle that suggests a partial answer. Let me show you."

The problem, known as "The Dancing Elf Puzzle," was first posed by Greg Kuperberg in [6] and goes as follows. There is a circle of circumference 1. An elf is dancing around the circle by taking clockwise steps of arclength τ, where $\tau = 0.61803\ldots$ is the golden section, which satisfies the equation $\tau + \tau^2 = 1$ (see Figure 1). If you number the elf's steps in chronological order, stopping whenever you or the elf gets tired, you will observe that the differences between consecutive step numbers are all Fibonacci numbers. Several possible stopping points are illustrated in Figure 2. The puzzle poses the questions: is this always the case no matter where the elf stops, and if so, why?

For a collection of related Fibonacci-themed puzzles, the interested reader may peruse [9, p. 147].

In the Figure 2 diagrams, you will notice that the physical space between steps corresponds to the difference between the step numbers. Each Fibonacci number has a matching arclength, and the larger the number, the smaller the arclength. This can be explained in part by the fact that rotating the entire picture by an angle of $360\tau°$ causes all the step numbers to advance by one, a key simplifying step in solving the problem. For the moment, we defer this solution.

The puzzle has well-established botanical connections (see, for example, [8]). In any sunflower's center, clockwise and counterclockwise spirals are discernible in the arrangement of seeds, and the numbers of spirals in each direction are virtually always consecutive Fibonacci numbers. In fact, it was the act of sitting down with a sunflower photograph (Figure 3) and counting 55 spirals going one way and 89 going the other that sparked my ongoing fascination with the origin of the patterns. Might the elf's dance suggest a mechanism for sunflower growth that would produce Fibonacci families of spirals?

Seed Patterns

The arrangement of the seeds in the center of a flower is one of many botanical structures, including leaves around a stem and scales in a pinecone, described by *phyllotaxis*, a term derived from the linguistic roots for "leaf arrangement." The unifying growth principle is that in all three settings, the leaves or seeds or scales emerge radially from the center of the plant, and the angle between consecutive botanical elements is constant. In leaves around a stalk, the particular phyllotaxis arrangement is commonly described in terms of *leaves per turn*. For instance, a plant that exhibits 5 leaves in 2 turns has a phyllotaxis angle of 144° between successive leaves, which is 2/5 of the circle. In this article, I will describe this in terms of the *phyllotaxis ratio* 2/5. In

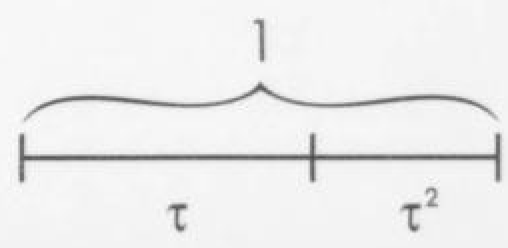


Figure 1. The golden section.

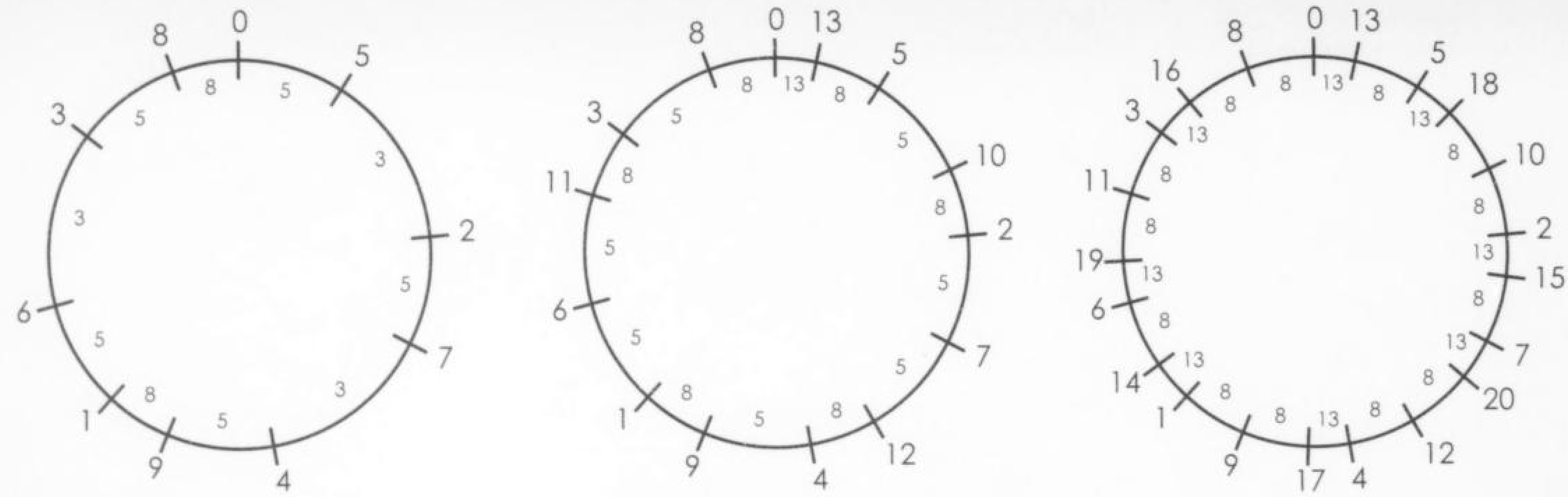


Figure 2. Ten-step, fourteen-step, and twenty-one-step elf dances.

a typical leaved plant, there are relatively few leaves per stalk, and so the phyllotaxis ratio cannot be observed with as much precision as in a flower head with hundreds or thousands of seeds.

Here, then, is the connection between dancing elves and sunflowers. Suppose that the phyllotaxis ratio in a sunflower is τ. The angular positions of the sunflower seeds are now the same as those of the elf's steps. Let us number the seeds in order of growth. If we look at a particular seed, say seed k, and then take the closest seed with a larger number, seed m, the dancing elf puzzle tells us that $m - k$ is a Fibonacci number.

To make the discussion simpler, suppose we have a (fictitious) very small sunflower with a small number of spirals. If we start numbering at seed 0, then the next closest seed to 0 will have a Fibonacci number, say 8. In fact, seeds 0 and 8 will be so close that they touch. Since the angle between seeds 0 and 1 is the same as the angle between seeds 8 and 9, seeds 1 and 9 will be correspondingly close and so will touch, as will seeds 2 and 10, seeds 3 and 11, and so forth. By the same token, we will see a spiral strand of touching seeds starting with 0 (consisting of seeds 0, 8, 16, 24, . . .), and another spiral starting with 1, a third spiral starting with 2, and so forth up to an eighth spiral starting with 7. Thus, we find a Fibonacci family of spirals. The closest seed to 0 on the opposite side of 8 will likewise be a Fibonacci seed, and by the same reasoning we will find another Fibonacci family of spirals going in the opposite direction. A closer look at the dancing elf puzzle, which we'll come to later, explains why the two Fibonacci numbers are consecutive.

It is worth noting that the Fibonacci differences associated to neighboring seeds, in contrast to the elf's steps, will be much smaller than the total number of seeds. In a flower head, the seeds push each other out from the center, and they are separated by radial distance as well as angular distance. This effectively makes each seed part of a much shorter elf dance. In Figure 4, the seeds of a daisy, which contains the sunflower pattern on a smaller scale, are numbered in order of growth. The daisy has approximately 112 seeds, whereas the sunflower in Figure 3 has approximately 1500. In the top image, the family of 21 spirals is highlighted, and in the middle image, the family of 13 spirals is highlighted. However, as the seeds get closer to the center, their radial separation increases. This accounts for the fact that inside the blue circles, seeds that are 21 or 13 apart no longer touch. The seeds in the bottom image have been colored according to parity: the even seeds are dark green and the odd seeds are light green. In the fringe outside the blue circle, seeds that are 34 apart touch. Were the daisy a bit larger, this proximity would produce a 34-spiral family that would overshadow the 13-spiral family. Inside the blue circle is a clearly visible family of 8 spirals.

Why on earth would a sunflower, daisy, or pinecone prefer a phyllotaxis ratio of τ?

This still leaves a gap in this interpretation of the Fibonacci phenomenon. Why on earth would a sunflower, daisy, or pinecone prefer a phyllotaxis ratio of τ? While the rest of this article will give a hint of an explanation, this gap has been eloquently bridged else-

AUTHOR

SUSAN GOLDSTINE received her A.B. in Mathematics and French from Amherst College and her Ph.D. in Mathematics from Harvard University. She is currently Assistant Professor of Mathematics at St. Mary's College of Maryland, where her research interests include number theory, Diophantine geometry, and integral lattices, and her professional hobbies include modular origami, topological fiber arts, and pinecone numbering. At well over forty chimes, she has the largest windchime collection of anyone she knows personally.

Department of Mathematics and Computer Science
St. Mary's College of Maryland
18952 East Fisher Rd
St. Mary's City, MD 20686-3001
e-mail: sgoldstine@smcm.edu
http://www.smcm.edu/users/sgoldstine

where, and I direct the interested reader to [7] for a thorough exposition. In fact, although the ratio τ is predominant, it is not universal, and there are plants that exhibit different phyllotaxis ratios and hence non-Fibonacci spirals. For some lovely photographs of plants with various phyllotaxis ratios, see [1].

Three years or so after I stumbled onto the dancing elf puzzle, another mathematical connection presented itself out of the blue. I was making diagrams for a talk about phyllotaxis and sunflowers when the pictures I was drawing suddenly struck me as eerily familiar. I had seen this process before, and when I had, it had nothing to do with flowers or Fibonacci. It had to do with Euclid.

Euclid's Algorithm

Nowadays, Euclid's Algorithm is usually shown in numeric form. As is the case with much of Euclid's *Elements*, the original algorithm is geometrically presented.

The geometric Euclid's Algorithm (Figure 5), also known as *antenaresis*, takes as inputs two segments, whose lengths we call a and b. The first step of the algorithm, shown on the second segment in Figure 5, is to use a compass to lay off segments of length b on our segment of length a until we can no longer do so. If a cannot be divided evenly into segments of length b, then we are left with a remainder segment of length r_1. At each succeeding step, we lay off copies of the remainder segment from the previous step on the dividing segment from the previous step.

> *The extremes of scaling that occur in Euclid's Algorithm go beyond the pale.*

A drawback of the geometric algorithm is that it is usually physically unfeasible. With most constructions, staying in the range between the maximum reach of the compass and the error tolerance of the compass is next to impossible, but the extremes of scaling that occur in Euclid's Algorithm go beyond the pale. For our particular example, three iterations is close to the limit of what can be legibly drawn, even with a computer's precision. Naturally, we can determine the theoretical outcome of the algorithm without compass and straightedge. In this example, the segments were chosen in a ratio of 360 to 157, and so the numerical analogue of the calculation begun in Figure 5 is the following list of long divisions.

$$\begin{aligned} 360 &= 2(157) + 46 \\ 157 &= 3(46) + 19 \\ 46 &= 2(19) + 8 \\ 19 &= 2(8) + 3 \\ 8 &= 2(3) + 2 \\ 3 &= 1(2) + 1 \\ 2 &= 2(1) + 0 \end{aligned} \tag{1}$$

Figure 3. Sunflower. Photograph by Donna D'fini. © Copyright 2006 Donna D'fini. Printed with permission.

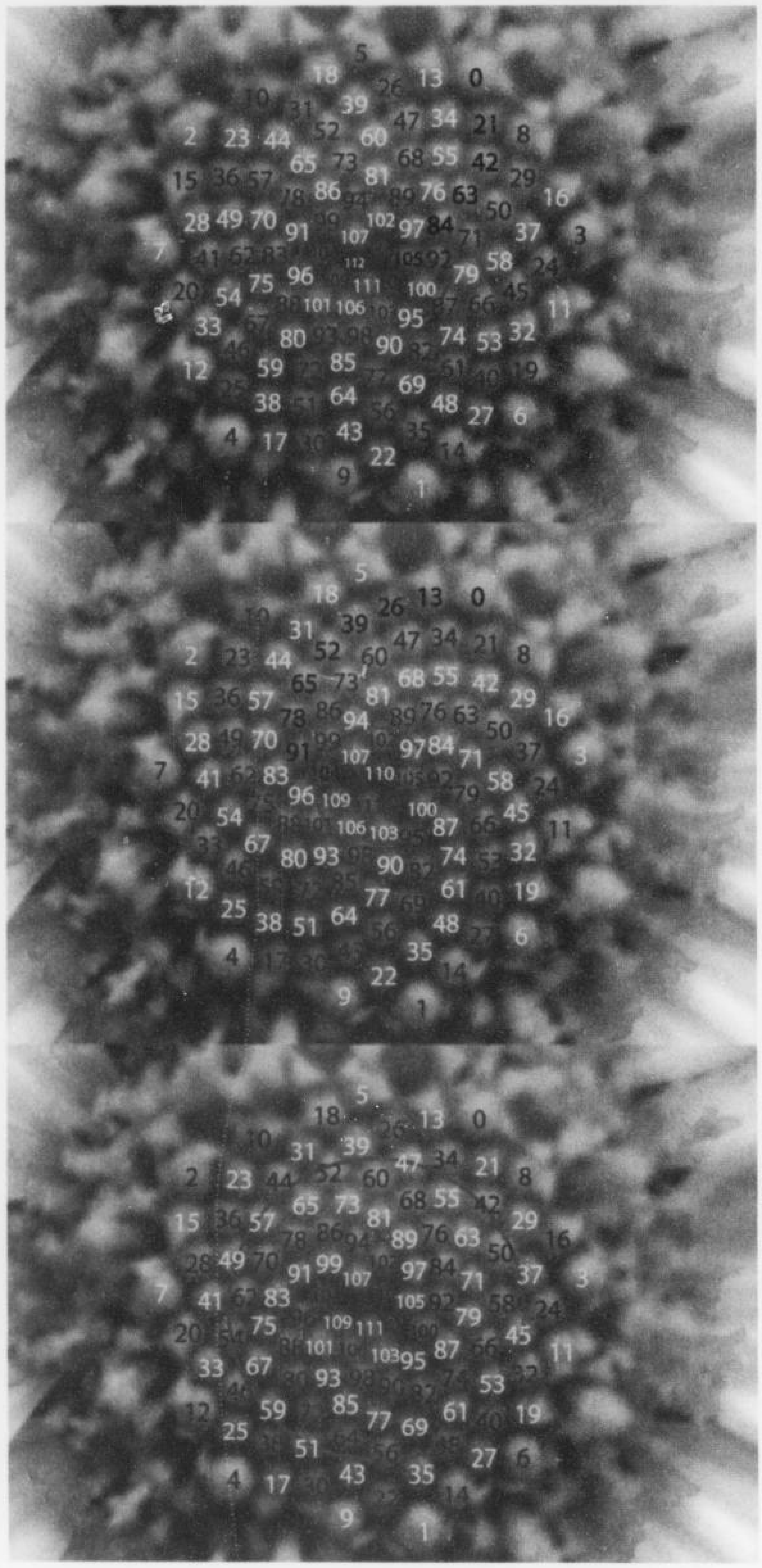


Figure 4. The seeds of a daisy. Photography by Marian Goldstine. © Copyright 2006 Marian Goldstine. Modified and printed with permission.

Therefore, in our example, the algorithm will terminate with a sixth and final remainder segment of length m, where $a = 360m$ and $b = 157m$. In light of the Pythagoreans' shocking discovery that not all numbers are rational, it is also possible for antenaresis not to terminate.

In the general scenario, if the algorithm terminates, then the smallest segment we construct will be the largest segment that *measures* both of the input segments, i.e., a segment of the largest length into which both a and b can be divided evenly. This is Proposition 3 of Book X of Euclid's *Elements*. On the other hand, if the algorithm does not terminate, then there is no segment that commonly measures segments of length a and b, and thus the original segments are *incommensurable*. This statement is Proposition 2 of Book X. The first reference in the *Elements* to Euclid's Algorithm occurs in the first of the three number theory books, Book VII, where Propositions 1 and 2 state that the output of the algorithm is the greatest common divisor of the inputs. However, even here the algorithm is described in terms of repeated subtraction of segments, all of which are integer multiples of a fixed unit.

An alternative form of Euclid's Algorithm for numbers allows for the non-terminating form of the algorithm. For the integer calculation in Equations 1, we obtain the second numeric algorithm by dividing each line by the divisor for that line. The initial input is now the improper fraction 360/157, and the first step is to write this as its integer part plus a remainder between 0 and 1. Each new equation begins with the reciprocal of the remainder from the previous step.

$$\begin{aligned}
\frac{360}{157} &= 2 + \frac{46}{157}\\
\frac{157}{46} &= 3 + \frac{19}{46}\\
\frac{46}{19} &= 2 + \frac{8}{19}\\
\frac{19}{8} &= 2 + \frac{3}{8}\\
\frac{8}{3} &= 2 + \frac{2}{3}\\
\frac{3}{2} &= 1 + \frac{1}{2}\\
\frac{2}{1} &= 2 + \frac{0}{1}
\end{aligned} \tag{2}$$

This form of Euclid's Algorithm has two major benefits. First, we no longer need to begin with a rational number, and it is not hard to see that the algorithm will terminate if and only if the initial input is rational. Second, this form of the algorithm allows us to compute a continued fraction expansion of the initial value. For the example above, we obtain

$$\frac{360}{157} = 2 + \frac{1}{\frac{157}{46}} = 2 + \cfrac{1}{3 + \cfrac{1}{\frac{46}{19}}} = \cdots$$

$$= 2 + \cfrac{1}{3 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2}}}}}} \tag{3}$$

In the interests of space, we adopt the common convention (as in [4]) of de-

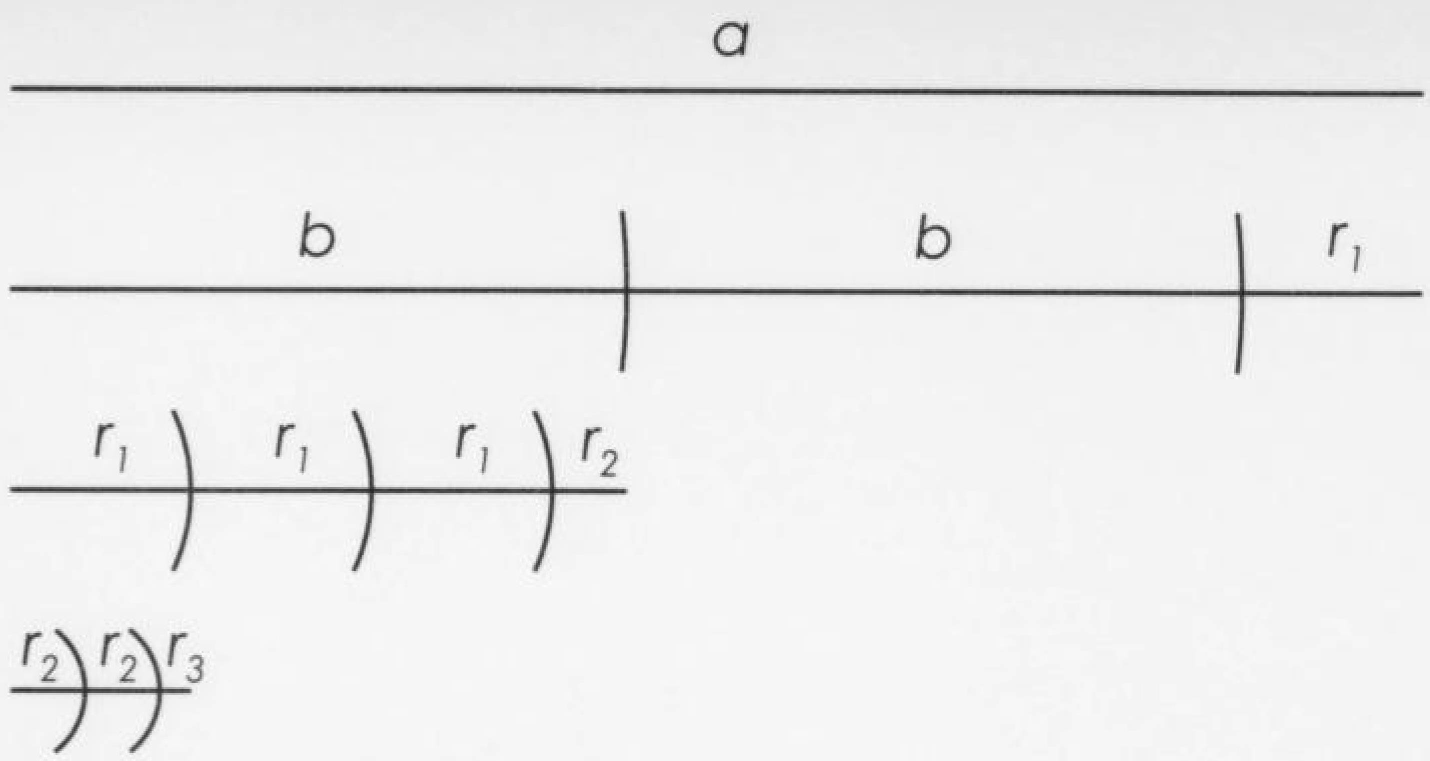


Figure 5. The geometric Euclid's Algorithm.

noting this continued fraction by [2;3,2,2,2,1,2]. If we begin with an irrational number, we obtain an infinite continued fraction expansion.

A neat thing happens when we apply Euclid's Algorithm to the golden ration ϕ, the number that satisfies $\phi = 1 + 1/\phi$. This equation is the first line of Euclid's algorithm, and since ϕ is the reciprocal of $1/\phi$, it is also the second line, the third line, and so on. Consequently, the continued fraction expansion of ϕ is

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{\ddots}}} = [1;1,1,1,\ldots]$$

Naturally, this continued fraction satisfies the equation $x = 1 + 1/x$, since taking the reciprocal and adding 1 yields the self-same continued fraction. Because the two solutions of $x = 1 + 1/x$ are ϕ and $-\tau$, we can see directly that this continued fraction *must* be ϕ.

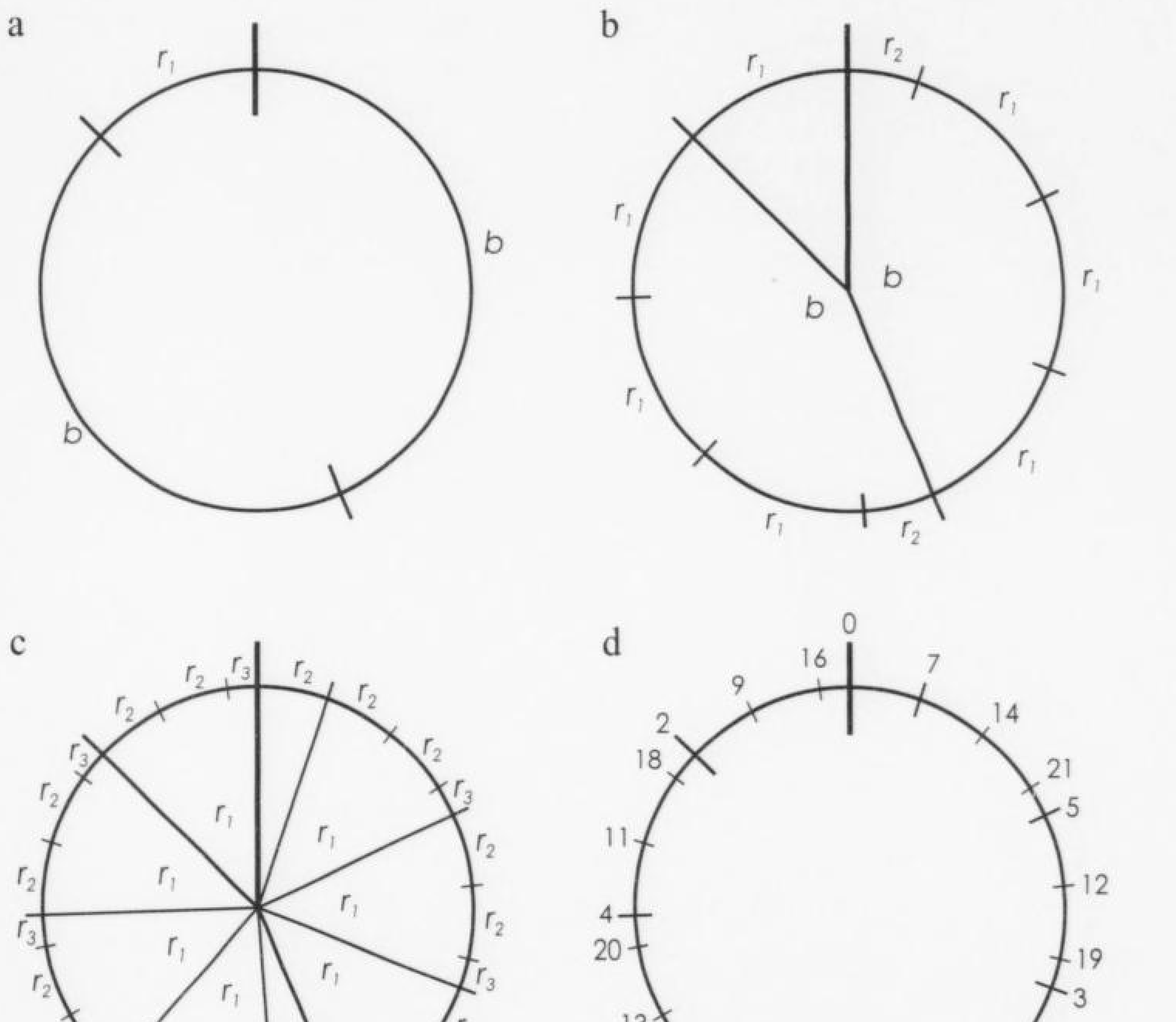


Figure 6. The circular Euclid's Algorithm.

Furthermore, because dividing the defining equation for ϕ by ϕ gives the formula $1 = 1/\phi + (1/\phi)^2$,

$$\tau = \frac{1}{\phi} = \phi - 1 = [0;1,1,1,\ldots]$$

is the continued fraction expansion for the golden section.

Elves and Euclid's Algorithm

Let us now return to a broader form of the dancing elf puzzle. Instead of a circle of circumference 1 and clockwise steps of arclength τ, consider the general scenario of a circle of circumference a and clockwise steps of arclength b. This dance corresponds to the sequence of angles produced by phyllotaxis with a phyllotaxis ratio of b/a. Figure 6 shows several stages of the dance in which $a/b = 360/157$.

> *Eventually, the elf will wind up closer to step 0 than* r_1.

At the outset, the elf lays off arcs of length b on the full circle of length a until the elf is closer to his starting point than b. Let us call this step number S_1 and the length of the remaining arc r_1. In Figure 6a, we reach this stage at $S_1 = 2$.

Look closely at this last step. The elf arrived here by starting at 0 and taking S_1 clockwise steps of arclength b. For conceptual convenience, we will call a consecutive string of S_1 steps a 1-jig. Performing a 1-jig starting at step 0 places the elf r_1 *counterclockwise* from step 0. By the same token, step $S_1 + 1$ is a 1-jig from step 1, landing the elf r_1 counterclockwise from step 1, and the following step lands him r_1 counterclockwise from step 2, and so forth. In fact, from this point on, the elf is laying off arcs of length r_1 counterclockwise on all the arcs of length b. More precisely, the elf cycles around the arcs of length b in the order he originally stepped them, laying off a further counterclockwise r_1 on each pass. Furthermore, the endpoints of each arc of length r_1 are a 1-jig apart, and so the difference between the step numbers is always S_1.

Eventually, the elf will wind up closer to step 0 than r_1, at which point he will be standing at an arclength of r_2 *clockwise* from step 0. Let us call this new step S_2, and a string of S_2 consecutive steps a 2-jig. This is the beginning of the final pass through the arcs of length b. When this final pass is complete, the circle will be completely divided into arcs of lengths r_1 and r_2, as shown in Figure 6b. At this stage, we call the circle r_1-*saturated*, just as in Figure 6a, the circle is b-saturated.

By now, we can see what the elf is really doing. The dance is an iterated circular version of Euclid's Algorithm! The elf will now lay off arcs of length r_2 clockwise on all the arcs of length r_1, until at last he is standing r_3 *counterclockwise* from step 0 at step S_3. This will complete his first 3-jig. Shortly thereafter (Figure 6c), the elf has chopped the entire circle into arcs of length r_2 and r_3, the circle is r_2-saturated, and the division of r_2 by r_3 begins.

Figure 6d shows the steps of the 157/360 elf dance numbered in chronological order. In this case, while the differences between consecutive step numbers are not Fibonacci numbers, they are all either $S_2 = 7$ or $S_3 = 16$, the number of steps in 2-jigs and 3-jigs. To see why these particular numbers arise, and for that matter, why the Fibonacci numbers arise in the τ dance, we take a closer look at continued fractions.

Convergents

Euclid's Algorithm allows us to write any real number γ as a continued fraction $[c_0;c_1,c_2,c_3,\ldots]$, where the terms c_k are the integer parts extracted in Euclid's Algorithm, as in Equations 2. When γ is a rational number, the continued fraction is a finite-length expression ending at some denominator c_n. From this expansion, we can extract a sequence of close rational approximations of γ, the *convergents* of γ, which are the truncated continued fractions

$$c_0,\ [c_0;c_1],\ [c_0;c_1,c_2],\ [c_0;c_1,c_2,c_3],\ \ldots$$

We will denote these convergents as P_1/Q_1, P_2/Q_2, and so forth, where P_k/Q_k is the convergent that ends with c_{k-1}.

The launching point for the study of continued fractions is the existence of recursive formulas for P_k and Q_k. It is clear from the definitions that $P_1 = c_0$ and $Q_1 = 1$. To get the recurrence started, we set $P_0 = 1$ and $Q_0 = 0$. With these initial values, all of the numerators and denominators of the convergents of γ are defined by the recurrence

$$P_{k+1} = c_kP_k + P_{k-1} \text{ and } Q_{k+1} = c_kQ_k + Q_{k-1} \quad (4)$$

(see for instance [4, p. 4]).

From the continued fractions we computed above, we get the following data: for 360/157,

c_k	2	3	2	2	2	1	2	
P_k	1	2	7	16	39	94	133	360
Q_k	0	1	3	7	17	41	58	157

;

and for ϕ,

c_k	1	1	1	1	1	1	1	...
P_k	1	1	2	3	5	8	13	21 ...
Q_k	0	1	1	2	3	5	8	13 ...

.

It's hard not to notice that all of the entries in the final table are Fibonacci numbers. Since $c_k = 1$ for all k, the numerators and denominators of the convergents of ϕ satisfy the Fibonacci recurrence

$$F_{k+1} = F_k + F_{k-1}.$$

Indeed, the standard Fibonacci fact that the ratios F_{k+1}/F_k of consecutive Fibonacci numbers converge to ϕ is a special case of the general result that for any irrational γ, the convergents P_k/Q_k of γ converge to γ as k goes to ∞. If γ is rational, then the final convergent to γ will naturally be γ itself.

To see the connection between the P's and Q's and the elf's circular dance, we simply need to ask the right questions.

Question 1 How many steps are there in a k-jig?
Answer P_k.

By definition, a 1-jig consists of $c_0 = P_1$ steps. The first 2-jig is completed when the elf is r_2 clockwise from step 0, a feat accomplished by performing one step (a 0-jig, if you like) to land b clockwise from step 0 and then performing c_1 1-jigs to move r_1 counterclockwise c_1 times. Therefore, the total number of steps in a 2-jig is $1 + c_1P_1 = P_0 + c_1P_1 = P_2$.

The rest follows by induction. For $k \geq 2$, the elf reaches the first r_{k+1} arc

Figure 7. Combining jigs.

by performing a $(k-1)$-jig to land r_{k-1} from step 0 and then performing c_k k-jigs to divide r_{k-1} by r_k, as in Figure 7. The total number of steps in this $(k+1)$-jig is $P_{k-1} + c_kP_k = P_{k+1}$ by Equations 4.

Question 2 How many times does the elf wind around the circle in a k-jig?
Answer Rounding to the nearest full circuit, Q_k.

To be more precise, the total arclength travelled by the elf in a k-jig is $Q_ka + (-1)^kr_k$. Once again, this follows from induction, using the fact that

$$r_{k+1} = r_{k-1} - c_kr_k.$$

Since a 1-jig is r_1 short of a full circuit, the total arclength for a 1-jig is $c_1a - r_1 = Q_1a - r_1$. Subsequently, if we consider one step a 0-jig, a $(k+1)$-jig consists of a $(k-1)$-jig plus c_k k-jigs, as described above. So if we set $r_0 = b$ for convenience, the total arclength for a $(k+1)$-jig is

$$\begin{aligned} &Q_{k-1}a + (-1)^{k-1}r_{k-1} \\ &\qquad + c_k(Q_ka + (-1)^kr_k) \\ &= (Q_{k-1} + c_kQ_k)a - (-1)^k\,(r_{k-1} - c_kr_k) \\ &= Q_{k+1}a + (-1)^{k+1}r_{k+1}. \end{aligned}$$

We can make two further observations based on the answers above. The first concerns r_k-saturation. The circle is r_k-saturated when it is divided into arcs all of length r_k or r_{k+1}, which occurs at the moment that the elf completes the division of all the r_{k-1} arcs. There will be a single r_{k+1} remainder arc for each of the original r_{k-1} arcs for a total of P_k arcs of length r_{k+1}. Meanwhile, there will be P_{k-1} remainder arcs of length r_k left over from the r_{k-1}-saturation stage, plus a further c_kP_k arcs of length r_k from the divisions of r_{k-1}, for a total of $P_{k-1} + c_kP_k = P_{k+1}$ arcs of length r_k. This gives us a nice description of r_k-saturation and derives the formula

$$a = P_{k+1}r_k + P_kr_{k+1}.$$

These saturation patterns can be observed in the example of Figure 6. In Figure 6a, there are 2 arcs of length b and 1 arc of length r_1; in Figure 6b, there are 7 arcs of length r_1 and 2 arcs of length r_2; and in Figure 6c, there are 16 arcs of length r_2 and 7 arcs of length r_3.

The second observation is that we can derive one of the fundamental properties of convergents from the properties of the elf dance. We know that at step P_k, the elf is r_k from step 0, and furthermore that this is closer than the elf has been to his starting point on any earlier step. Conversely, the Euclid's-Algorithm structure of the dance tells us that any step that brings the elf closer to step 0 than ever before is some P_k. Now, the distance between step 0 and step n is

$$\min_{m \geq 0} |nb - ma|$$

since each step has length b and the circumference of the circle is a. We conclude that $|Pb - Qa|$ is the minimal value of $|nb - ma|$ for m and n nonnegative integers with $0 < n \leq P$ if and only if $P = P_k$ and $Q = Q_k$ for some $k \geq 0$.

To recast this in terms of convergents, let θ be the phyllotaxis ratio b/a. Since the convergents to a/b are P_k/Q_k for $k \geq 1$, the convergents to $\theta = b/a$ will be Q_k/P_k for $k \geq 0$. (Since $\theta < 1$, its first convergent is $0 = Q_0/P_0$.) Therefore, dividing $|Pb - Qa|$ and $|nb - ma|$ by $|a|$, we find that $|P\theta - Q|$ is the minimum value of $|n\theta - m|$ for $0 < n \leq P$ if and only if Q/P is a convergent of θ.

The Puzzle Resolved

Now we see the general pattern for step differences in elf dances. When the circle is r_k-saturated, there will be exactly two step differences, P_k and P_{k+1}, since any adjacent steps on the circle are either a k-jig or a $(k + 1)$-jig apart. As a consequence, a simulated flower head grown with a phyllotaxis ratio of θ will exhibit families of spirals corresponding to the denominators of the convergents of θ. In the most common case of ratio τ, this yields the famed Fibonacci spirals. The recent article [7], which derives these spiral patterns directly from the properties of convergents, features spectacular diagrams of flowers with phyllotaxis ratios of $\sqrt{2}$ and π.

We also have an explanation for why the families of spirals in either direction in a sunflower come in *consecutive* Fibonacci numbers. In terms of the nearest seed, since the new remainder segments alternate between clockwise and counterclockwise from step 0, seeds F_k and F_{k+1} are always on opposite sides of seed 0.

There is a crucial final piece to the original dancing elf puzzle. In the puzzle, the elf gets to stop whenever he likes, whether or not the circle is saturated. In the τ-dance snapshots in Figure 2, only the third image shows the circle in a saturated position. The reason that the Fibonacci pattern holds in all intermediate positions is that since $c_k = 1$ for all k, every single step the elf takes completes one line of the geometric Euclid's Algorithm, and so the elf never produces any arcs that are not in the sequence of remainders. In fact, from the relation $1 = \tau + \tau^2$ and the resulting relations $\tau^m = \tau^{m+1} + \tau^{m+2}$, we see that all of the arcs in the τ dance have length a power of τ, and we can derive that each arc of length τ^m corresponds to a Fibonacci difference of F_m.

There is a crucial final piece to the original dancing elf puzzle.

On the other hand, with a phyllotaxis ratio of 157/360, where most of the c_k's are greater than 1, there will be intermediate stages containing arcs that have not been fully divided. Such an intermediate stage is shown in Figure 8, an unsaturated stage which contains three arcs of length $r_1 + r_2$ with a step difference of 9.

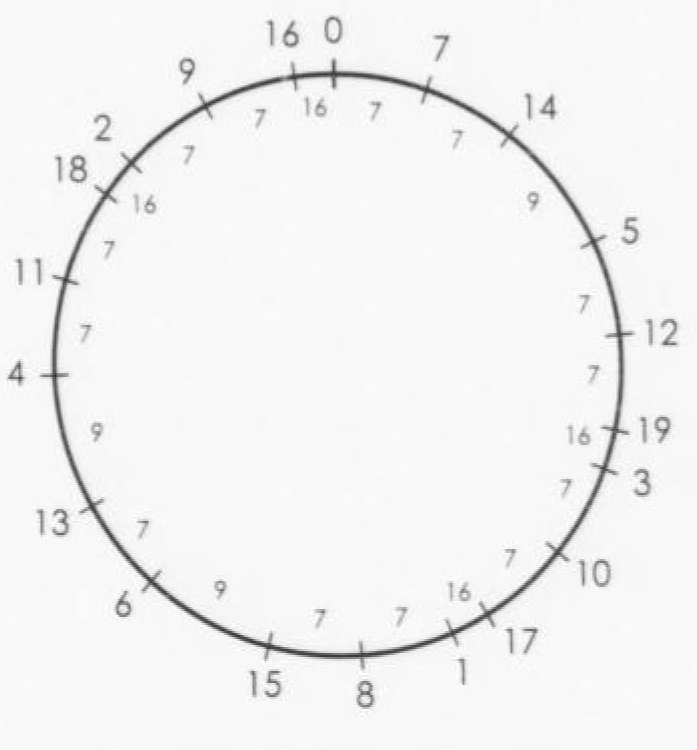


Figure 8. Why $\frac{157}{360}$ is different from τ.

That $c_k = 1$ for all k is also the key to why τ makes an outstanding phyllotaxis ratio. The fundamental goal of phyllotaxis is uniform scattering, which keeps leaves from stealing each other's sunlight and packs seeds into stable seed heads. If the goal is to take exactly n seeds and uniformly distribute them around the circle, then the ideal arrangement is to place them exactly $1/n$ of the circle apart, which can be achieved by a rational phyllotaxis ratio of the form m/n. However, as soon as more than n seeds are deposited, new seed angles coincide with old seed angles, which is as poor a scatter pattern as possible. Thus equal spacing between seeds is impractical, and rational phyllotaxis ratios are less than ideal.

With an irrational phyllotaxis ratio, we will never have the circle divided into equal arcs. The best we can hope for is that the largest arc and the smallest arc are as close in size as possible. In an unsaturated configuration, these arcs will have the form r_{k-1} and r_{k+1}, respectively. Since

$$r_{k-1} = c_k r_k + r_{k+1} > (c_k + 1) r_{k+1},$$

to maximize the ratio r_{k+1}/r_{k-1}, thereby keeping the remainders' arcs as close in length to each other as possible, we set $c_k = 1$ for all k.

Flowers and Euclid's Algorithm

So does all of this mean that flowers know Euclid's Algorithm?

In a word, no.

The fact is that flowers don't grow the way elves dance. The mechanism in the center of a sunflower does not perform the algorithm: *eject seed, turn* $360\tau^\circ$, *repeat.* In biological terms, it's not that the flower selects a phyllotaxis ratio of τ in order to grow properly. It's that the flower grows properly and this causes the phyllotaxis ratio to rapidly lock into τ. There is a lot of wonderful science and mathematics in this process that is still under active investigation. To read more about it, see [2] and [3]. The latter describes a fascinating experiment in which phyllotactic patterns are induced on droplets of ferrofluids using magnetic fields. You can see a video of this experiment through the *Gallery* link at [1]. The conclusion is that while elves may view Euclid's Algorithm as a deliberate process, flowers are gradually trained into the algorithm by primal forces.

For that matter, some τ-flowers are less reliable than others. My primary experience is with sunflowers, which are quite faithful in producing Fibonacci spirals. However, in scouting daisy photographs for Plate 2, I found that a substantial percentage of the photographs I studied were of daisies in which one or both families of spirals were not Fibonacci, or in which a few spirals merged in midstream, muddying the count. Moreover, I am dismayed to report that on more than one occasion, I have spotted a 7-12 pineapple in a supermarket. One of my sharper colleagues, on hearing of this culinary catastrophe, remarked, "Clearly, you must get rid of the evidence."

Now, what to do with the defective daisies . . .

ACKNOWLEDGMENTS

I am deeply indebted to Ravi Vakil for showing me the dancing elf puzzle and to John Conway for his botanical insights and flower-numbering tips. Greg Kuperberg was kind enough to explain the original context of the puzzle to me and thereby increase my understanding of the topological connections in [6]. I am also grateful to Marian Goldstine and Donna D'Fini for allowing me to use their flower photographs, to Michael Kleber, Marjorie Senechal, and Jonathan Goldstine for meticulous proofreading, to Ron Knott for his online compendium of all things Fibonacci [5] and the many valuable resources I have found there, and to all my colleagues, friends, and relatives who have suffered and supported my obsession with sunflowers and pinecones over the years.

And to Dan Shapiro, thanks for the pineapple quip.

REFERENCES

[1] Pau Atela, Christophe Golé, *et al.*, Phyllotaxis: an interactive site for the mathematical study of plant pattern formation, http://maven.smith.edu/~phyllo/index.html.

[2] John Horton Conway and Richard K. Guy, *The Book of Numbers*, Copernicus Books, 1996.

[3] Stephane Douady and Yves Couder, Phyllotaxis as a self-organised growth process, *Growth Patterns in Physical Sciences and Biology* (J M Garcia-Ruiz *et al.*, ed.), Plenum Press, 1993, pp. 341–352.

[4] A.Ya. Khinchin, *Continued Fractions*, Dover Publications, 1997.

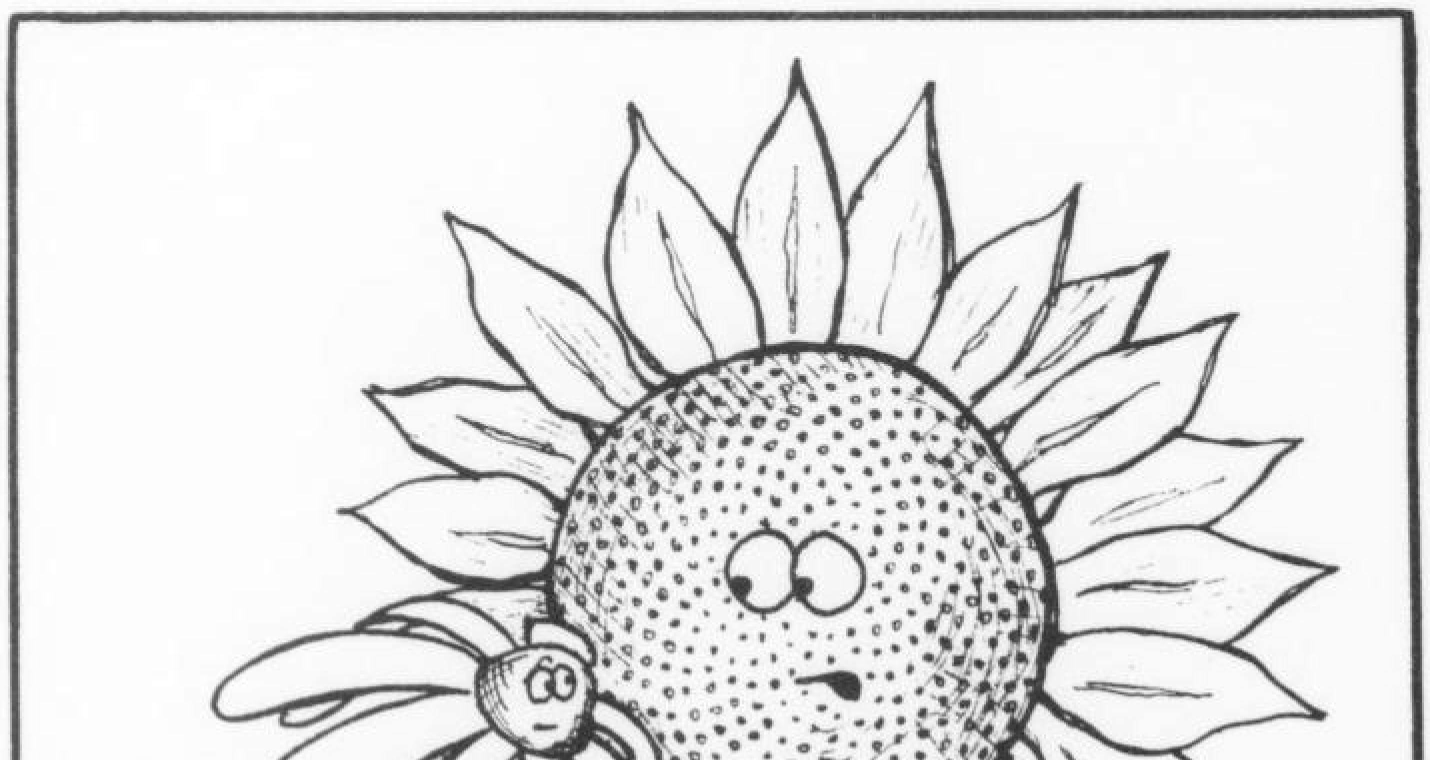
[5] Ron Knott, Fibonacci numbers and the golden section, http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fib.html.

[6] Gregory Kuperberg, A volume-preserving counterexample to the Seifert conjecture, *Comment. Math. Helv.* **71** (1996), 70–97.

[7] M. Naylor, Golden, $\sqrt{2}$, and π flowers: A spiral story. *Mathematics Magazine* **75** (June 2002), 163–172.

[8] M. Kunz and F. Rothen, Phyllotaxis or the properties of spiral lattices III. An algebraic model of morphogenesis. *J Phys I France*, **75** (1992), 2131–2172.

[9] Ravi Vakil, *A Mathematical Mosaic: Patterns & Problem Solving*, Brendan Kelly Publishing, 1997.



"Ever notice that the number of legs on an animal is always a number from the sequence {0, 2, 4, 6, 8,...}?"

Contributed by Judy Holdener, Kenyon College

My Lunch with Arnold

GÁBOR DOMOKOS

In Hungary I teach Civil Engineering. I lean more towards the mathematical side of the subject than to designing buildings. Just after the political changes swept over the country in the late 1980s I got a Fulbright Fellowship to visit America, to an engineering department known to have people with mathematical taste like mine. I had a good year writing papers with various American professors. One of them was Andy Ruina, with whom I became friends and had an infinite number of conversations on not quite as many topics. One recurrent theme was Andy's friend Jim Papadopoulos, a guy with academic taste but not an academic job. Through Andy, I came to respect the unseen Jim.

One day Andy told me that Jim had a simple conjecture but was too busy with his day job, designing machines to refill laser toner cartridges, to work on trying to prove it. Jim offered through Andy, as a gift of sorts, that I could work on the problem.

Jim imagined drawing a closed curve on a thick piece of plywood. A convex curve, meaning that it had no indented places. Now cut along that line with a jigsaw and balance the plywood piece, on edge, on a flat table. Gently keep it on edge so it doesn't fall flat on the table. In mathematical language, think of this as a two-dimensional (2D) problem. This plywood is stable only in certain positions. For example a square piece of plywood is stable on all four edges. In the positions where one diagonal or the other is vertical, the plywood is in equilibrium, but it is an unstable equilibrium. A tiny push and it will fall towards lying on one of the edges. An ellipse is in stable equilibrium when horizontal and resting on one of the two flatter parts. And the same ellipse is in unstable equilibrium when balanced on either end, like an upright egg. Jim conjectured that no matter what convex shape you draw and cut out, it has at least two orientations where it is stable. The ellipse has two such positions, a triangle has three (the three flat edges), a square has four, a regular polygon has as many stable equilibrium positions as it has edges. And a circle is a degenerate special case that is in equilibrium in every orientation (none of which is stable or unstable). Jim's conjecture was that *every* shape, but for a circle, has at least as many stable positions as an ellipse.

Jim's plywood conjecture was a simple idea, and it was true for every shape we could think of. Of course it is not true if you are allowed to add weights. For example, you can put a big weight in a plywood ellipse near one of the sharp ends, so the only stable configuration is standing upright, like a child's toy called the "comeback kid". We didn't allow that. We only allowed homogeneous shapes, uniform plywood.

After some days of thought and talk with Andy, Jim, and others, we found a proof that every convex piece of plywood has at least two orientations where it will stand stably. Then we generalized the idea to include things made from wire. We published the results in the respectable but not widely read *Journal of Elasticity*.

What kept bugging us was the 3D generalization. Imagine something made of solidified clay. Was it true that you could always find at least two orientations for such a thing where it would sit stably on a table? We couldn't prove this, and for good reason. Finally I found a counter-example: a shape that could balance stably on a table in just one position. Take a long solid cylinder and diagonally chop off one end, then at the opposite angle chop off the other end. This truncated cylinder is happy lying on the table with its long side down, but in no other position. Just one stable equilibrium. We never published this, and I stopped thinking about balancing plywood shapes, wire loops, and clay solids.

About five years later there was the International Congress on Industrial and Applied Mathematics in Hamburg. This was to be the biggest mathematics meeting ever, with over 2000 people attending. Coming from a second-world country I needed, applied for, and got a little first-world money so I could attend. The meeting had over 40 parallel sessions: at any one time of the day I had a choice of over 40 different talks I could listen to. My own talk was on something I thought was profound at that time. But it was put in the wrong session. To an outsider, math might seem like math. But either the subject is broad or mathematicians are narrow; the number of talks that any single conference attendee could hope to understand was small. Although my audience sat politely through my carefully practiced 15-minute presentation, I don't think any of the few who understood my English understood a word of my mathematics. Mine seems not to have been the only misplaced talk. I didn't understand any of the talks I went to, either. Besides thousands of these incomprehensible 15-minute talks, there were three simultaneous 45-minute invited talks each day.

But most centrally, there was one plenary talk with no simultaneous sessions. All 2000 mathematicians could attend without conflict. This plenary lecture was to be presented by no less a figure than Vladimir Igorevich Arnold, the man who solved Hilbert's thirteenth problem when he was a teenager and the author of countless famous articles, reviews, books, and theorems.

Like everyone else, I felt obligated to go, despite (again like everyone else) having little hope of understanding anything of this great man's work. There was a steady murmur in the room as Arnold began to speak; people chatting to their friends whom they understood rather than listening to Arnold whom they had no hope of understanding. When a talk is over my head I either switch off completely, as I did for most of the conference talks, or I try to catch a detail here or there that might fit together loosely in my mind somehow. I did the latter until my breath was taken away. Arnold's talk made excursions into various topics that I don't know about, like differential geometry and optics. But each topic ended with something about the number four. He said these topics were examples of a theorem created by the great nineteenth-century mathematician Jacobi. He said Jacobi's theorem had many applications, and that always something had to be bigger or equal to four. He covered one topic or another that would be familiar to each person in the audience, always coming back to the number four. After everyone in the audience had seen the number four appear in some problem that he or she knew something about, the murmurs of distracted conversation quieted. The giant auditorium became almost silent, with people practically holding their breath in attentiveness. Four in this problem, four in that, four in some problem or other that everyone could understand. Four, four, four.

My respect for Arnold grew. Being a brilliant mathematician is one thing. Riveting the attention of 2000 mathematicians, most of whom can't understand each other, is another. Although I didn't understand the lecture, I felt exhilarated and happy.

As I left the auditorium it suddenly struck me that Jim's plywood and wire problem might be related to Jacobi's theorem. We had proved that at least two stable equilibria existed, but this implies that there are at least four equilibria, two stable and two unstable. Like the ellipse. Arnold's four. I was so impressed with myself that I stopped dead for a minute, blocking the exit.

I had to tell this to Arnold. Maybe the number four was a coincidence, maybe not. He would know. But of course Arnold was mobbed after the talk. I realized that getting face to face with the great man might be impossible. But almost immediately I noticed a big poster. The conference organizers were advertising special lunches. For an exorbitant fee one could buy a ticket to eat with a math celebrity. Although my budget was tight and my mathematics is not at the level of Arnold, I could calculate that if I reduced my eating from two hotdogs a day to one I could afford a lunch ticket with the great Professor.

The lunch was a disaster, both from my point of view and Arnold's. The organizers had tried to maximize their profit rather than the ticket-buyers' pleasure. At the big round table with Arnold were ten eager young mathematicians. Each was carrying one or two "highly important" scientific papers which were full of "highly relevant" results they wanted to share with Arnold. He could not eat as they held out their papers and made claims about their great original contributions. And unless I was willing to butt into this noisy whining, as each of the people was doing to the others, I could not speak. I sat and tried to look attentive at the pathetic scene.

At the end of the meal Arnold finally asked me, "And what is *your* paper about?"

I said, "Nothing."

"Surely you have something to ask or say," he said.

But I was depressed by the fray and said no, I had just wanted to listen. The big meeting went on day after day. I ate one hotdog a day and I went to a hundred fifteen-minute talks that I didn't understand.

On the last day I packed my suitcase and headed for the airport. The main lobby of the conference center was deserted, maintenance people were taking down posters, the buffet was closed, people were fading out. As I strolled across the big hall I noticed, next to a young Asian man, leaning on a counter near the closed buffet, Professor V. I. Arnold. The young Asian man was talking excitedly in the tone I had noted at the disastrous lunch. As I walked closer, Arnold raised his voice slightly.

"As I told you already several times, there is nothing new in what you are telling me. I published this in 1980. Look it up. I do not want to discuss this further; moreover, I have an appointment with the gentleman carrying the suitcase over there. Good-bye."

The disappointed young mathematician got up to leave and Arnold turned to me. "You wanted to talk to me, right?" Stunned that he even remembered me, but aware of the part I suddenly was supposed to play, I pretended that the discussion was expected. "You sat at the lunch table, right? You must have had a reason. What is it about? Tell me fast. I have to catch my train."

We sat down, I collected my thoughts and explained about the plywood and the wire and how they gave the number two, which really meant four. He stared off without saying a word. After five minutes I asked him if he wanted to know how we proved that the plywood had at least four equilibria. He waved me away impatiently. "Of course I know how you proved it" and then he breezily outlined the proof in a few phrases. "That is not what I am thinking about. The question is whether your result follows from the Jacobi theorem or not."

He stared off again. I reminded him of his train but he waved me away again. Looking at his enormous concentration, and not knowing what I should be thinking about, the minutes went by slowly. Finally he said, "I think the Jacobi theorem and your problem are related, but yours is certainly not an example of the other. I think there is a third theorem that includes both Jacobi's theorem and your problem. I could tell better if I knew about the 3D version of your problem."

I proudly described the counter-example, the single stable equilibrium of the chopped-off cylinder, but he cut me off:

"You realize of course that this is not a counter-example! The main point of your 2D result was NOT to show that there are two or more stable equilibria, but to show that there are FOUR or more equilibria altogether." This was not the main point of our 2D result in my mind, or at least hadn't been. But now I realized that there was a higher level of thought going on here. Four and not two. "And your cylinder has four equilibria, three of which are unstable."

In a moment's pondering I realized he was right. The cylinder could also balance unstably when rotated 180 degrees on its axis and also on its two ends. Four. I was stunned. "A counter-example still may exist. Send me a letter when you find a body with less than four equilibria in the three-dimensional case," he said, "I have to catch my train. Good-bye, young man, and good luck to you!"

I returned to Hungary and my life of teaching and pretty little irrelevant problems, each important in my mind for a few months or years. It is possible that, besides the proofreader at the *Journal of Elasticity*, no-one's eyes have ever passed over every word in our paper on plywood and wire. Ten years later Arnold's conjecture turned out to be correct: the 3D counterexample not only existed but appeared to me as a mathematically most exciting object (see the following paper). I never saw Arnold again. Besides the number four, and four again, I still have no idea what the Jacobi theorem is about. So I will never understand the generalization of Jacobi's theorem that V. I. Arnold imagined to encompass also our balancing plywood and wire, cooked up there in the huge convention hall in Hamburg, Germany, sitting next to me at the deserted buffet.

ACKNOWLEDGEMENT

I thank Jim Papadopoulos for the pre-Arnold version of the problem described here, and Andy Ruina for encouraging me to write this story down and then editing it.

Mono-monostatic Bodies
The Answer to Arnold's Question

P. L. VÁRKONYI AND G. DOMOKOS

As V. I. Arnold conjectured: convex, homogeneous bodies with fewer than four equilibria (also called *mono-monostatic* bodies) may exist. Not only did his conjecture turn out to be true, the newly discovered objects show various interesting features. Our goal is to give an overview of these findings based on [7], as well as to present some new results. We will point out that mono-monostatic bodies are neither flat, nor thin, they are not similar to typical objects with more equilibria, and they are hard to approximate by polyhedra. Despite these "negative" traits, there seems to be an indication that these forms appear in Nature due to their special mechanical properties.

Do Mono-monostatic Bodies Exist?

In his recent book [11] V. I. Arnold presented a rich collection of problems sampled from his famous Moscow seminars. As Tabachnikov points out in his lively review [12], a central theme is geometrical and topological generalization of the classical Four-Vertex Theorem [2], stating that *a plane curve has at least four extrema of curvature.* The condition that some integer is *at least four* appears in numerous different problems in the book, in areas ranging from optics to mechanics. Being one of Arnold's long-term research interests, this was the central theme to his plenary lecture in 1995, Hamburg, at the International Conference on Industrial and Applied Mathematics, presented to more than 2000 mathematicians (see the accompanying article). The number of equilibria of homogeneous, rigid bodies presents a big temptation to believe in yet another emerging example of *being at least four* (in fact, the planar case was *proven* to be an example [1]). Arnold resisted and conjectured that, counter to everyday intuition and experience, the three-dimensional case might be different. In other words, he suggested that convex, homogeneous bodies *with fewer than four equilibria* (also called *mono-monostatic*) may exist. As often before, his conjecture proved not only to be correct but to open up an interesting avenue of mathematical thought coupled with physical and biological applications, which we explore below.

Why Are They Special?

We consider bodies resting on a horizontal surface in the presence of uniform gravity. Such bodies with just *one* stable equilibrium are called *monostatic* and they appear to be of special interest. It is easy to construct a monostatic body, such as a popular children's toy called "Comeback Kid" (Figure 1A). However, if we look for *homogeneous, convex* monostatic bodies, the task is much more difficult. In fact, in the 2D case one can prove [1] that among planar (slab-like) objects rolling along their circumference *no* monostatic bodies exist. (This statement is equivalent to the famous Four-Vertex Theorem [2] in differential geometry.)

The proof for the 2D case is indirect and runs as follows. Consider a convex, homogeneous planar "body" B and a polar coordinate system with origin at the center of gravity of B. Let the continuous function $R(\varphi)$ denote the boundary of B. As demonstrated in [1], non-degenerate stable/unstable equilibria of the body correspond to local minima/maxima of $R(\varphi)$. Assume that $R(\varphi)$ has only one local maximum and one local minimum. In this case there exists exactly one value $\varphi = \varphi_0$ for which $R(\varphi_0) = R(\varphi_0 + \pi)$; moreover, $R(\varphi) > R(\varphi_0)$ if $\pi > \varphi - \varphi_0 > 0$, and

$$R(\varphi) < R(\varphi_0) \text{ if } -\pi < \varphi - \varphi_0 < 0 \text{ (see Figure 2A).}$$

The straight line $\varphi = \varphi_0$ (identical to $\varphi = \varphi_0 + \pi$) passing through the origin O cuts B into a "thin" ($R(\varphi) < R(\varphi_0)$) and a "thick" ($R(\varphi) > R(\varphi_0)$) part. This implies that *O can not be the center of gravity,* i.e., it contradicts the initial assumption.

Not surprisingly, the 3D case is more complex. Although one can construct a homogeneous, convex monostatic body

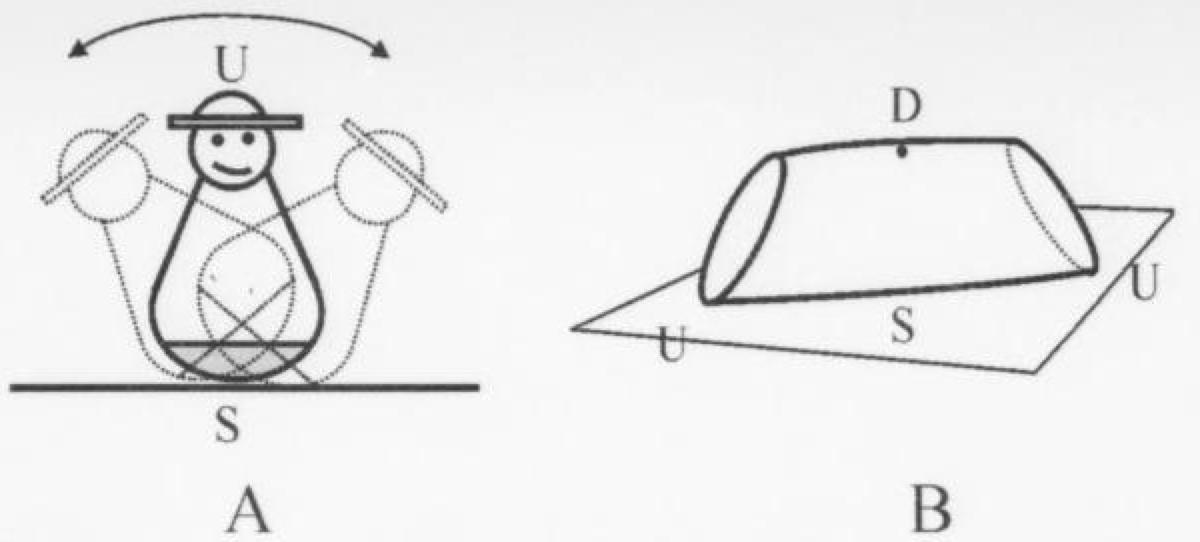


Figure 1. A. Children's toy with one stable and one unstable equilibrium (inhomogeneous, mono-monostatic body), also called the "comeback kid." B. Convex, homogeneous solid body with one stable equilibrium (monostatic body). In both plots, S, D, and U denote points of the surface corresponding to stable, saddle-type, and unstable equilibria of the bodies, respectively.

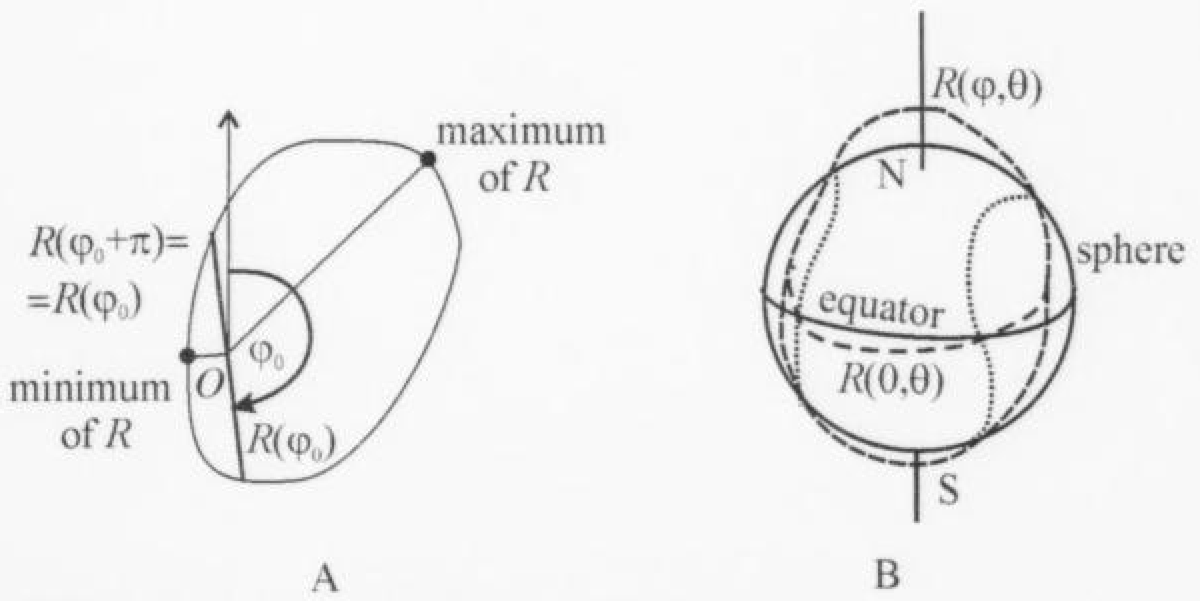


Figure 2. A. Example of a convex, homogeneous, planar body bounded by $R(\varphi)$ (polar distance from the origin O). Assuming $R(\varphi)$ has only two local extrema, the body can be cut to a "thin" and a "thick" half by the line $\varphi = \varphi_0$. Its center of gravity is on the "thick" side, in particular, it cannot coincide with O. B. 3D body (dashed line) separated into a "thin" and a "thick" part by a tennis ball-like space curve C (dotted line) along which $R = R_0$. Continuous line shows a sphere of radius R_0, which also contains the curve C.

(Figure 1B), the task is far less trivial if we look for a monostatic polyhedron with a *minimal number* of faces. Conway and Guy [3] constructed such a polyhedron with 19 faces (similar to the body in Figure 1B); it is still believed that this is the minimal number. It was shown by Heppes [6] that no homogeneous, monostatic tetrahedron exists. However, Dawson [4] showed that homogeneous, monostatic simplices exist in $d > 7$ dimensions. More recently, Dawson and Finbow [5] showed the existence of monostatic tetrahedra—but with inhomogeneous mass density.

One can construct a rather transparent classification scheme for bodies with exclusively non-degenerate balance points, based on the number and type of their equilibria. In 2D, stable and unstable equilibria always occur in pairs, so we say that a body belongs to class $\{i\}$ $(i > 0)$ if it has exactly $S = i$ stable (and thus, $U = i$ unstable) equilibria. As we showed above, class $\{1\}$ is empty. In 3D we appeal to the Poincaré-Hopf Theorem [8], stating for convex bodies that $S + U - D = 2$, S,U,D denoting the number of local minima, maxima, and saddles of the body's potential energy; so class $\{i,j\}$ $(i,j > 0)$ contains all bodies with $S = i$ stable, $U = j$ "unstable," and $D = i + j - 2$ saddle-type equilibria.

Monostatic bodies are in classes $\{1,j\}$; we will refer to the even more special class $\{1,1\}$ with just one stable and one unstable equilibrium as "*mono-monostatic.*" While in 2D being monostatic implies being mono-monostatic (and vice versa), the 3D case is more complicated: a monostatic body could have, in principle, any number of unstable equilibria (e.g., the body in Figure 1B belongs to class $\{1,2\}$ and has four equilibria altogether, as pointed out by Arnold, see story). *Arnold's conjecture was that class* $\{1,1\}$ *is not empty*, i.e., *that homogeneous, convex mono-monostatic bodies exist.* Before we outline the construction of such an object, we want to highlight its very special relation to other convex bodies.

Intuitively it seems clear that by applying small, local perturbations to a surface, one may produce additional lo-

AUTHORS

GÁBOR DOMOKOS earned his PhD ("Candidate" in Hungarian usage) in 1989 from the Hungarian Academy of Sciences. Since his election in 2004 he has been the youngest member of the Academy. In addition to his positions in Budapest, he has been visiting scholar at the University of Maryland and Cornell University. He has worked extensively on discrete and continuous dynamical systems with Philip Holmes. His leisure activities include classical music, freehand drawing, and hiking with his wife Réka.

Department of Mechanics, Materials, and Structures
Budapest University of Technology and Economics
H-1111 Budapest, Hungary
e-mail: domokos@iit.bme.hu

PÉTER VÁRKONYI was a medal-winner at the International Physics Olympiads in 1997. Later he wrote his doctoral thesis under Gábor Domokos at the Budapest University of Technology and Economics. His research is in optimization of structures, group representations, and adaptive dynamics for evolutionary models. He is currently a post-doctoral fellow working with Philip Holmes and Simon Levin. He is married with two children, and likes to spend spare time, if any, hiking or playing badminton.

Program in Applied and Computational Mathematics
Princeton University
Princeton, NJ 08544
USA
e-mail: vpeter@mit.bme.hu

cal maxima and minima (close to existing ones), similar to the "egg of Columbus." According to some accounts, Christopher Columbus attended a dinner which a Spanish gentleman had given in his honor. Columbus asked the gentlemen in attendance to make an egg stand on one end. After the gentlemen successively tried to and failed, they stated that it was impossible. Columbus then placed the egg's small end on the table, breaking the shell a bit, so that it could stand upright. Columbus then stated that it was "the simplest thing in the world. Anybody can do it, after he has been shown how!" In [7] we showed that in an analogous manner, one can *add* stable and unstable equilibria *one by one* by taking away locally small portions of the body. Apparently, the inverse is not possible, i.e., for a *typical* body one cannot decrease the number of equilibria via small perturbations.

This result indicates the special status of mono-monostatic bodies among other objects. For any given typical mono-monostatic body, one can find bodies in an arbitrary class $\{i,j\}$ which have *almost the same shape*. On the other hand, to any typical member of class $\{i,j\}$, $(i,j > 1)$, one can *not* find a mono-monostatic body which has almost the same shape. This may explain why mono-monostatic bodies do not occur often in Nature, also, why it is difficult to visualize such a shape. Next we will demonstrate such an object.

What Are They Like?

As in the planar case, a mono-monostatic 3D body can be cut to a "thin" and a "thick" part by a closed curve on its boundary, along which $R(\theta,\varphi)$ is constant. If this separatrix curve happens to be planar, its existence leads to contradiction, similar to the 2D case. (If, for example, it is the "equator" $\varphi = 0$ and $\varphi > 0/\varphi < 0$ are the thick/thin halves, the center of gravity should be on the upper $(\varphi > 0)$ side of the origin). However, in case of a generic spatial separatrix, the above argument no longer applies. In particular, the curve can be similar to the ones on the surfaces of tennis balls (Figure 2B). In this case the "upper" thick ("lower" thin) part is partially below (above) the equator; thus it is possible to have the center of gravity at the origin. Our construction will be of this type. We define a suitable two-parameter family of surfaces $R(\theta,\varphi,c,d)$ in the spherical coordinate system (r,θ,φ) with $-\pi/2 < \varphi < \pi/2$ and $0 \leq \theta \leq 2\pi$, or $\varphi = \pm\pi/2$ and no θ coordinate, while $c > 0$ and $0 < d < 1$ are parameters. Conveniently, R can be decomposed in the following way:

$$R(\theta,\varphi,c,d) = 1 + d \cdot \Delta R(\theta,\varphi,c), \tag{1}$$

where ΔR denotes the *type* of deviation from the unit sphere. "Thin"/"thick" parts of the body are characterized by negativeness/positiveness of ΔR (i.e., the separatrix between the thick and thin portions will be given by $\Delta R = 0$), while the parameter d is a measure of how far the surface is from the sphere. We will choose small values of d so as to make the surface convex. Now we proceed to define ΔR.

We will have the maximum/minimum points of ΔR $(\Delta R = \pm 1)$ at the North/South Pole $(\varphi = \pm\pi/2)$. The shapes of the thick and thin portions of the body are controlled by the parameter c: for $c >> 1$ the separatrix will approach the equator; for smaller values of c, the separatrix will become similar to the curve on the tennis ball.

Consider the following smooth, one-parameter mapping $f(\varphi,c)$: $(-\pi/2,\pi/2) \rightarrow (-\pi/2,\pi/2)$:

$$f(\varphi,c) = \pi \cdot \left[\frac{e^{\left[\frac{\varphi}{\pi c} + \frac{1}{2c}\right]} - 1}{e^{1/c} - 1} - \frac{1}{2} \right]. \tag{2}$$

For large values of the parameter $(c >> 1)$, this mapping is almost the identity; however, if c is close to 0, there is a large deviation from linearity. Based on (2), we define the related maps

$$f_1(\varphi,c) = \sin(f(\varphi,c)) \tag{3}$$

and

$$f_2(\varphi,c) = -f_1(-\varphi,c). \tag{4}$$

We will choose ΔR so as to obtain $\Delta R(\varphi,\theta,c) = f_2(\varphi,c)$ if $\theta = \pi/2$ or $3\pi/2$ (i.e., a big portion of these sections of the body lie in the thick part, cf. Figure 2B) and $\Delta R = f_2$ if $\theta = 0$ or π (the majority of these sections are in the thin part). The function

$$a(\theta,\varphi,c) = \frac{\cos^2(\theta) \cdot (1 - f_1^2)}{\cos^2(\theta)(1 - f_1^2) + \sin^2(\theta) \cdot (1 - f_2^2)} \tag{5}$$

$$= \frac{1}{1 + \tan^2(\theta) \dfrac{\cos^2(f(\varphi,c))}{\cos^2(f(\varphi,c))}} \quad (\text{where } |\varphi| < \pi/2)$$

is used to construct ΔR as a weighted average of f_1 and f_2 in the following way:

$$\Delta R(\theta,\varphi,c) = \left\{ \begin{array}{ll} a \cdot f_1 + (1 - a) \cdot f_2 & \text{if } |\varphi| < \pi/2 \\ 1 & \text{if } \varphi = \pi/2 \\ -1 & \text{if } \varphi = -\pi/2 \end{array} \right\}. \tag{6}$$

The choice of the function a guarantees, on the one hand, the gradual transition from f_1 to f_2 if θ is varied between 0 and $\pi/2$. On the other hand, it was chosen to result in the desired shape of thick/thin halves of the body (Fig. 2B). The function R defined by equations (1)–(6) is illustrated in Figure 3 for intermediate values of c and d. For $c >> 1$, the constructed surface $R = 1 + d\Delta R$ is separated by the $\varphi = 0$ equator into two unequal halves: the upper $(\varphi > 0)$ half is "thick" $(R > 1)$ and the lower $(\varphi < 0)$ half is "thin" $(R < 1)$. By decreasing c, the line separating the "thick" and "thin" portions becomes a space curve; thus the thicker portion moves downward and the thinner portion upward. As c approaches zero, the upper half of the body becomes thin and the lower one becomes thick (cf. Figure 4.)

In [7] we proved analytically that there exist ranges for c and d where the body is convex and the center of gravity is at the origin, i.e., it belongs to class $\{1,1\}$. Numerical studies suggest that d must be very small $(d < 5 \cdot 10^{-5})$ to satisfy convexity together with the other restrictions, so the created object is very similar to a sphere. (In the admitted range of d, the other parameter is approximately $c \approx 0.275$.)

Figure 3. Plot of the body if $c = d = 1/2$

What Are They *Not* Like?

Intuitively, it appears that mono-monostatic bodies can be neither very *flat* nor very *thin*; the former shape would have at least two stable equilibria; the latter, at least two unstable equilibria. To make this intuition more exact, we define the flatness F and thinness T of a body. Draw a closed curve c on the surface, traced by the position vector $R(s)$, $s \in [0,1]$ from the center of gravity O. Pick two points P_i ($i = 1, 2$) on *opposite* sides of c, with position vectors R_i ($i = 1, 2$), respectively. We define the flatness and thinness as

$$F = \sup_{\forall c, P_1, P_2} \left\{ \frac{\min_s(R(s))}{\max_i(R_i)} \right\}, \quad T = \sup_{\forall c, P_1, P_2} \left\{ \frac{\min_i(R_i)}{\max_s(R(s))} \right\}.$$

Although F and T are hard to compute for a general case, it is easy to give both a problem-specific and a general lower bound. For the latter, we have

$$F, T \geq 1, \qquad (7)$$

since $F = T = 1$ can be always obtained by shrinking the curve c to a single point. For "simple" objects, F and T can be determined, and the values agree fairly well with intuition in Table 1.

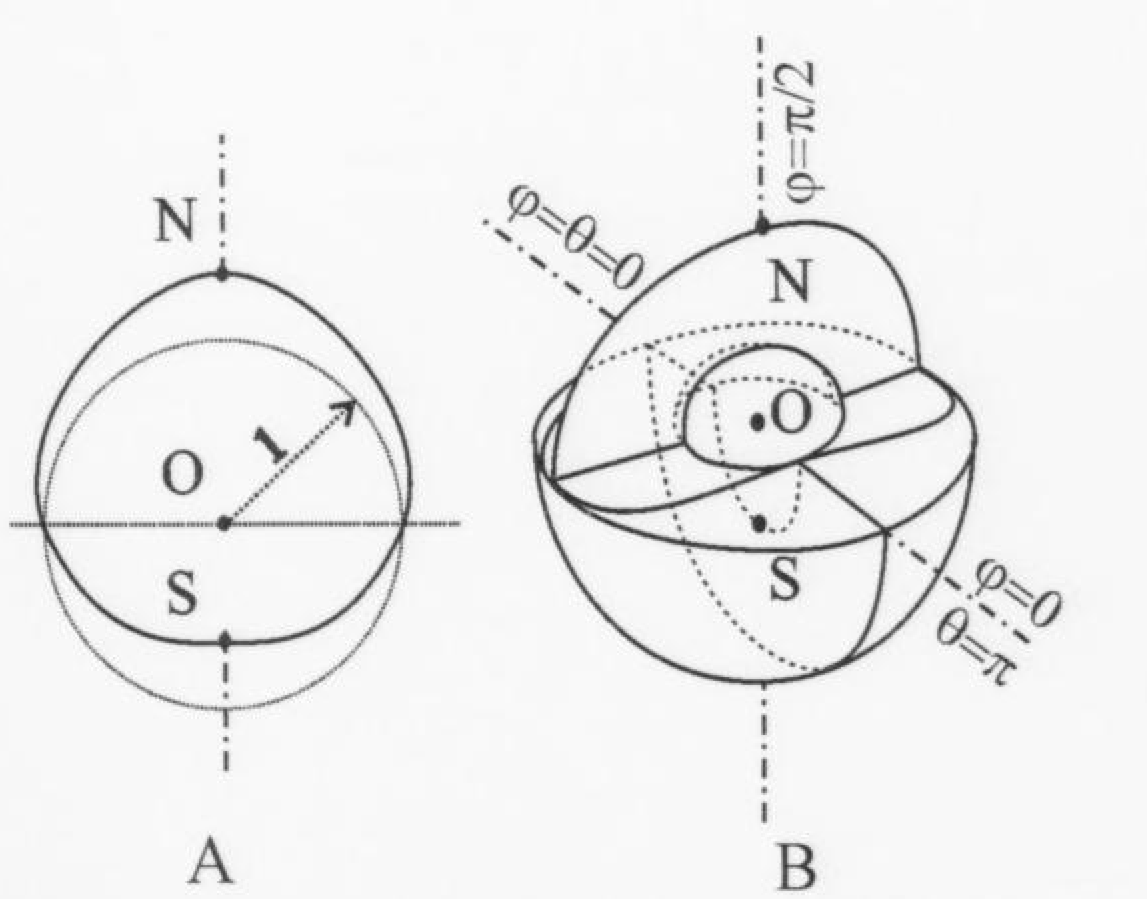


Figure 4. A. Side view of the body if $c >> 1$ (and $d \approx 1/3$). Note that $\Delta R > 0$ if $\varphi > 0$ and $\Delta R < 0$ if $\varphi < 0$. B. Spatial view if $c << 1$. Here, $\Delta R > 0$ typically for $\varphi < 0$ and vice versa.

Table 1. The flatness and thinness of some "simple" objects

Body	**Flatness *F***	**Thinness *T***
Sphere	1	1
Regular tetrahedron	$\sqrt{3}$	$\sqrt{3}$
Cube	$\sqrt{2}$	$\sqrt{(3/2)}$
Octahedron	$\sqrt{(3/2)}$	$\sqrt{2}$
Cylinder with radius r, height $2h$, $z = \sqrt{(r^2 + h^2)}$	z/h	z/r
Ellipsoid with axes $a < b < c$	b/a	c/b

Now we show that F and T are related to the number S of stable and U of unstable equilibria by

LEMMA 1: (a) $F = 1$ *if and only if* $S = 1$ *and*
(b) $T = 1$ *if and only if* $U = 1$.

We only prove (a); the proof of (b) runs analogously.

If $S > 1$, then there exists one global minimum for the radius R and at least one additional (local) minimum. Select c as a closed, $R = R_0 =$ *constant* curve, circling the local minimum very closely. Select the points P_1 and P_2 coinciding with global and local minima, respectively. Now we have $R_1 \leq R_2 < R_0$ and $\min(R(s)) = R_0$, $\max(R_i) = R_2$, so $S > 1$ implies $F > 1$.

If $S = 1$, then R has only one minimum, so it assumes only values greater than or equal to $\min(R(s))$ on one side of the curve c, so $F \leq 1$, but due to (7), we have $F = 1$. Q.e.d.

Lemma 1 confirms our initial intuition that mono-monostatic bodies can be *neither flat, nor thin*. In fact, they have simultaneously minimal flatness and minimal thinness; moreover, they are the only non-degenerate bodies having this property.

Another interesting though somewhat "negative" feature of mono-monostatic bodies is the apparent lack of any simple polyhedral approximation. As mentioned before, the existence of *monostatic* polyhedra with minimal number of faces has been investigated [3],[4],[5],[6]. One may generalize this to the existence of polyhedra *in class* $\{i, j\}$, with minimal number of faces. Intuitively it appears evident that polyhedra exist in each class: if we construct a sufficiently fine triangulation on the surface of a smooth body in class $\{i, j\}$ with vertices at unstable equilibria, edges at saddles and faces at stable equilibria, then the resulting polyhedron may—at sufficiently high mesh density and appropriate mesh ratios—"inherit" the class of the approximated smooth body. It also appears that if the topological inequalities $2i \geq j + 4$ and $2j \geq i + 4$ are valid, then we can have "minimal" polyhedra, where the number of stable equilibria equals the number of faces, the number of unstable equilibria equals the number of vertices, and the number of saddles equals the number of edges. Much more puzzling appear to be the polyhedra in classes *not satisfying* the above topological inequalities: a special case of these polyhedra are monostatic ones; however, many other types belong here as well. In particular, it would be of interest to know the minimal number of faces of a polyhedron in class $\{1,1\}$. We can imagine such a polyhedron as an ap-

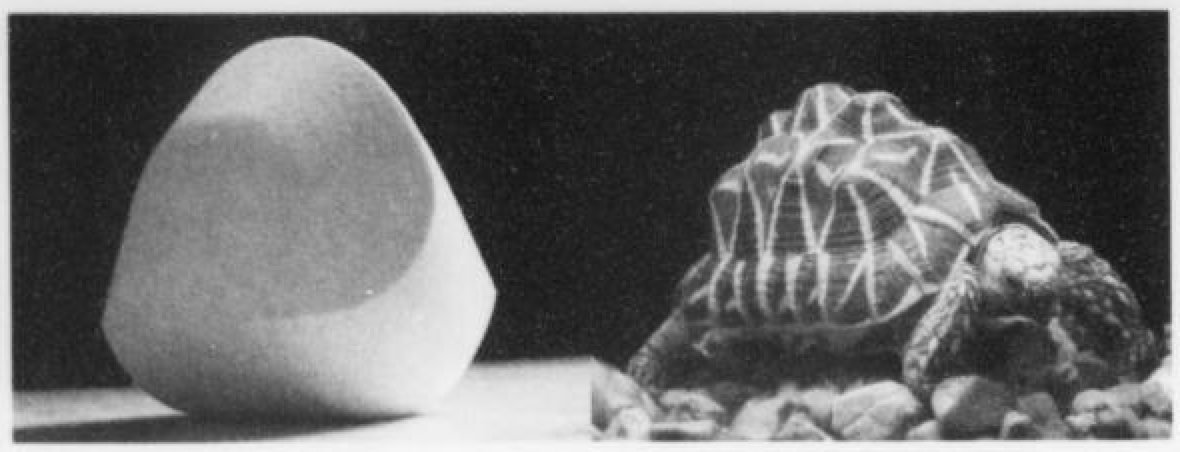


Figure 5. Mono-monostatic body and Indian Star Tortoise (*Geochelone elegans*).

proximation of a smooth mono-monostatic body. Since the latter are close to the sphere (they are neither flat nor thin), the number of equilibria is particularly sensitive to perturbations, so the minimal number of faces of a mono-monostatic polyhedron may be a very large number. Since we are most curious to see a homogeneous, 3D mono-monostatic polyhedron, we offer a prize at US \$10,000/N for the first such object; here N denotes the number of faces.

Mono-monostatic Bodies *do* Exist

Arnold's conjecture proved to be correct: there exist homogeneous, convex bodies with just two equilibria; we called these objects mono-monostatic.

Based on the results presented so far, one must get the impression that mono-monostatic bodies are *hiding*—that they are hard to visualize, hard to describe, and hard to identify. In particular, we showed that their form is not similar to any typical representative of any other equilibrium class. We also showed that they are *neither flat, nor thin*; in fact, they are the only non-degenerate objects having simultaneously minimal flatness and thinness. Imagining their polyhedral approximation seems to be a futile effort as well: the minimal number of faces for mono-monostatic polyhedra might be very large. The extreme physical fragility of these forms (i.e., their sensitivity to local perturbations due to abrasion) was also confirmed by statistical experiments on pebbles (reported in [7]); in a sample of 2000 pebbles not a single mono-monostatic object could be identified. Apparently, mono-monostatic bodies escape everyday human intuition.

They did not escape Arnold's intuition. Neither does Nature ignore these mysterious objects: being monostatic can be a life-saving property for land animals with a hard shell, such as beetles and turtles. In fact, the "righting response" (their ability to turn back when placed upside down) of these animals is often regarded as a measure of their fitness ([9],[10]). Although the example presented above under "Why Are They Special," proved to be practically indistinguishable from the sphere, rather different forms are also included in the mono-monostatic class. In particular, we identified one of these forms, which not only shows substantial deviation from the sphere, but also displays remarkable similarity to some turtles and beetles. We built the object by using 3D printing technology, and in Figure 5 it can be visually compared to an Indian Star Tortoise (*Geochelone elegans*).

Needless to say, the analogy is incomplete: turtles are neither homogeneous nor mono-monostatic. (They do not need to be exactly mono-monostatic; righting is assisted dynamically by the motion of the limbs.) On the other hand, being that close to a mono-monostatic form is probably not just a coincidence; as we indicated before, such forms are unlikely to be found by chance, either by us or by Evolution itself.

ACKNOWLEDGEMENT

The support of OTKA grant TS49885 and of the Dr. Imre Kurány'i fellowship are gratefully acknowledged.

REFERENCES

1. G. Domokos, J. Papadopoulos, A. Ruina: Static equilibria of planar, rigid bodies: is there anything new? *J. Elasticity* **36** (1994), 59–66.
2. M. Berger, B. Gostiaux: *Differential geometry: Manifolds, curves and surfaces*. Springer, New York (1988).
3. Conway, J. H. and Guy, R.: Stability of polyhedra, *SIAM Rev.* **11** (1969), 78–82.
4. Dawson, R.: Monostatic simplexes, *Amer. Math. Monthly* **92** (1985), 541–646.
5. Dawson, R. and Finbow, W.: What shape is a loaded die? *Mathematical Intelligencer* **22** (1999), 32–37.
6. Heppes, A.: A double-tipping tetrahedron, *SIAM Rev.* **9** (1967), 599–600.
7. Várkonyi, P. L. and Domokos G.: Static equilibria of rigid bodies: dice, pebbles and the Poincaré-Hopf Theorem. *J. Nonlinear Science* **16** (2006), 251–283.
8. Arnold, V. I.: *Ordinary Differential Equations*, 10th printing, 1998. MIT Press.
9. Freedberg, S. *et al.*: Developmental environment has long-lasting effects on behavioral performance in two turtles with environmental sex determination, *Evolutionary Ecology Res.* **6** (2004), 739–747.
10. Steyermark A. C., Spotila, J. R.: Body temperature and maternal identity affect snapping turtle (*Chelydra serpentina*) righting response, *COPEIA* **4** (2001), 1050–1057.
11. Arnold, V. I.: *Arnold's problems*. Springer, Berlin-Heidelberg-New York & PHASIS, Moscow, 2005.
12. Tabachnikov, S.: Review of *Arnold's problems, The Mathematical Intelligencer*, to appear.

Bubble Crawling in Dublin

Wiebke Drenckhan

Does your hometown have any mathematical tourist attractions such as statues, plaques, graves, the café where the famous conjecture was made, the desk where the famous initials are scratched, birthplaces, houses, or memorials? Have you encountered a mathematical sight on your travels? If so, we invite you to submit to this column a picture, a description of its mathematical significance, and either a map or directions so that others may follow in your tracks.

Please send all submissions to
Mathematical Tourist Editor,
Dirk Huylebrouck, Aartshertogstraat 42,
8400 Oostende, Belgium
e-mail: dirk.huylebrouck@ping.be

Caught in the rain you run for the nearest pub—which is never more than 50 metres away in Dublin. Recalling how the lady at the B&B had said "If you don't like the weather in Ireland, wait two hours!" you decide to do what the locals do: have a pint by the fireplace and wait.

As the bar man pours your Guinness, you admire the swirling motion of the many tiny gas bubbles as the famous stout settles in the glass. Your eyes get caught by the completely opaque froth that forms as the bubbles gather on the surface of the nearly black drink. A physicist friend once told you that because of the small size of the bubbles, a light ray gets reflected and deflected so many times, that it finds its way out again before it is absorbed, giving this froth the appearance of whipped cream. Also, the bubbles are filled with nitrogen (N_2), instead of carbon dioxide (CO_2), to slow down diffusion. Without this chemical trick, the creamy froth would quickly look like an ordinary dish-washing foam. But after watching the timescales within which the locals finish their pints, you doubt that they would ever notice this difference.

The rain continues, and your academic mind, despite being officially on holiday, wanders and wonders: "How many bubbles are in this froth and what are their shapes?" It goes without saying that as a mathematician you assume that all bubbles have the same volume. As long as they are round, which is the case if they are separated by a lot of liquid, this is the well-known problem of packing spheres. This has received attention by academics or greengrocers alike, aiming for the optimal packing of spherical objects.

Guinness bubbles—delicious food for the mathematical mind (by W. Drenckhan).

Rumour has it that this problem entered the realm of mathematics in the 1590s, when Sir Walter Raleigh, stocking his ship for an expedition, wondered if there was a quick way to calculate the number of cannon balls in a stack based on its height. His assistant, Thomas Harriot, not only supplied the appropriate equation but also mentioned the problem to Johannes Kepler a few years later. In 1611 Kepler proposed in his famous conjecture, that this was also the *most efficient* way of packing spheres—unfortunately without providing a proof. This had to wait until 1998, when T. Hales succeeded in providing a (debatably) rigorous proof of what the greengrocers had known all along: the face-centred cubic (fcc) packing (or its hexagonally close-packed cousin hcp) is the winner (Fig. 1).

Whilst you are staring at the froth, liquid drains out of it and the problem becomes more complicated. Bubbles

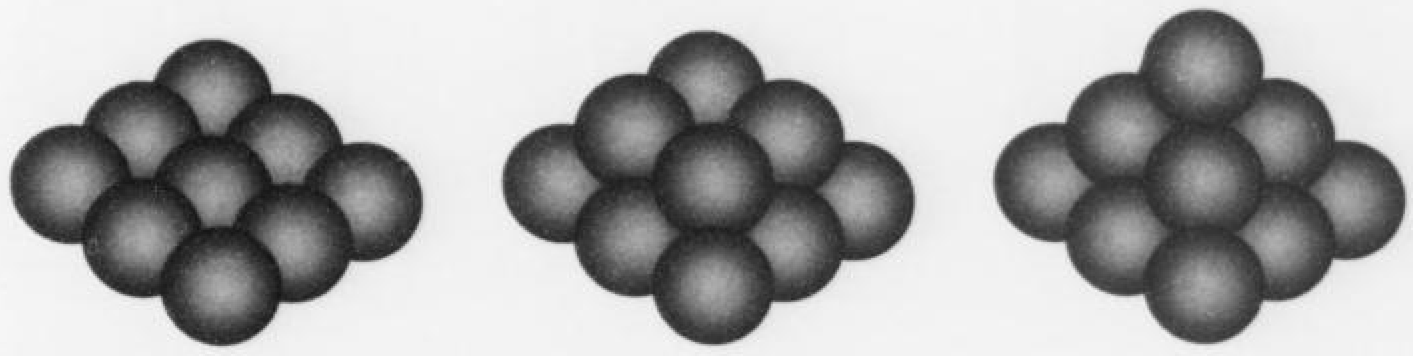


Figure 1. The way of stacking cannon balls or oranges; used in practice since we can remember, suggested as optimal by Kepler in 1611 and finally proven so by Hales in 1998.

now form polyhedra, which are separated by thin films whose mean curvature depends on the pressure difference between the bubbles they separate. With the amount of liquid being negligible, the packing problem takes a different form: the quantity to be minimised now is the "energy" of the system, which is represented by the total area of the gas/liquid interfaces. As the famous nineteenth-century Belgian scientist Joseph Plateau showed [1], this condition leads to very simple local equilibrium rules for the generally rather complex foam structures:

1. All films meet threefold at angles of 120°.
2. Lines of intersecting films always meet fourfold at angles of 109.6°.

However, like most people, you decide to think in more detail about the 2D analogue first. Here the close packing and interface minimisation problem produce the same solution, which can be found in many 2D and quasi-2D systems in nature, the most popular being the honeycomb (Fig. 2). But even though this solution has been widely accepted for a long time, it was only rigorously proven in 1999, once again by T. Hales [2].

The bee's challenge is actually slightly more complicated. Their hatcheries are stacked in two hexagonal layers (as shown in Figure 2), whose relative position and the resulting geometry of the separating wall has led to interesting academic debates [3, 4]. As a result, blind admiration for the intelligence of the small creatures was slightly dampened when the Hungarian mathematician Fejes Tóth proved that the bees could save 0.4% wax if they shifted the layers slightly [5].

Extending these considerations to true 3D systems turns out to be too complex for your back-of-a-beer-mat calculation. Just as you are about to give up you find yourself staring into the pub owner's delightful grin. "You know, I used to . . . " And before having time to remember the reputation of the Irish for story telling, you find yourself being drawn into one that sounds like a modern fairy tale about bubbles.

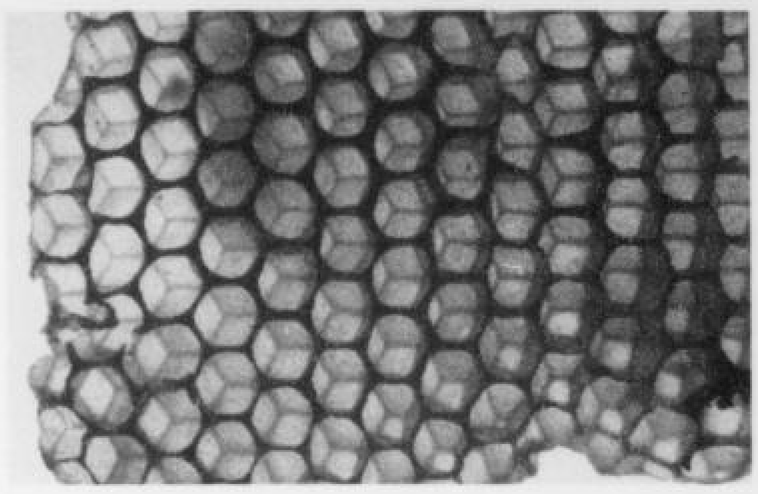


Figure 2. Masters of packing: a bee honeycomb (photograph by S. Hutzler).

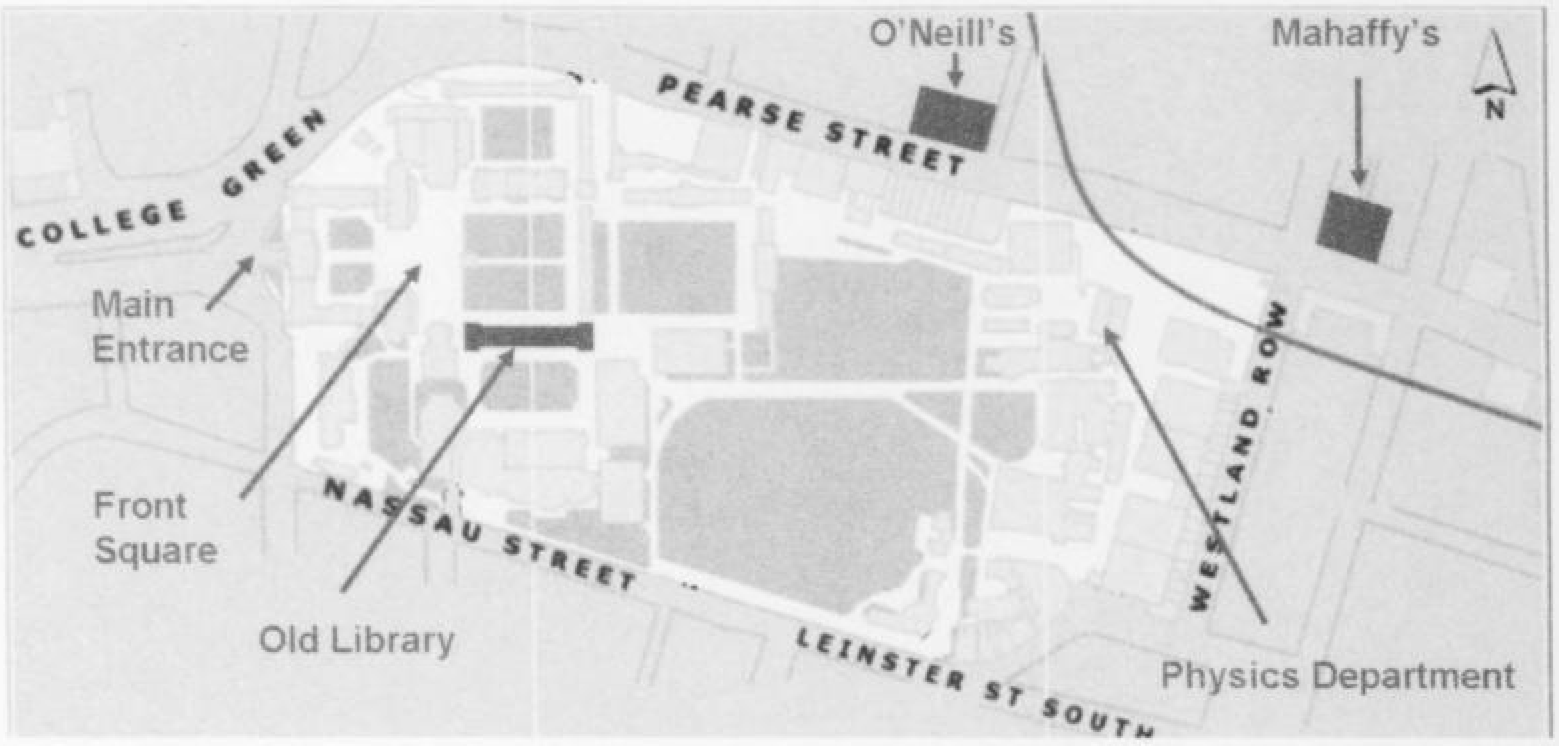


Figure 3. Map of Trinity College Dublin and surrounding area.

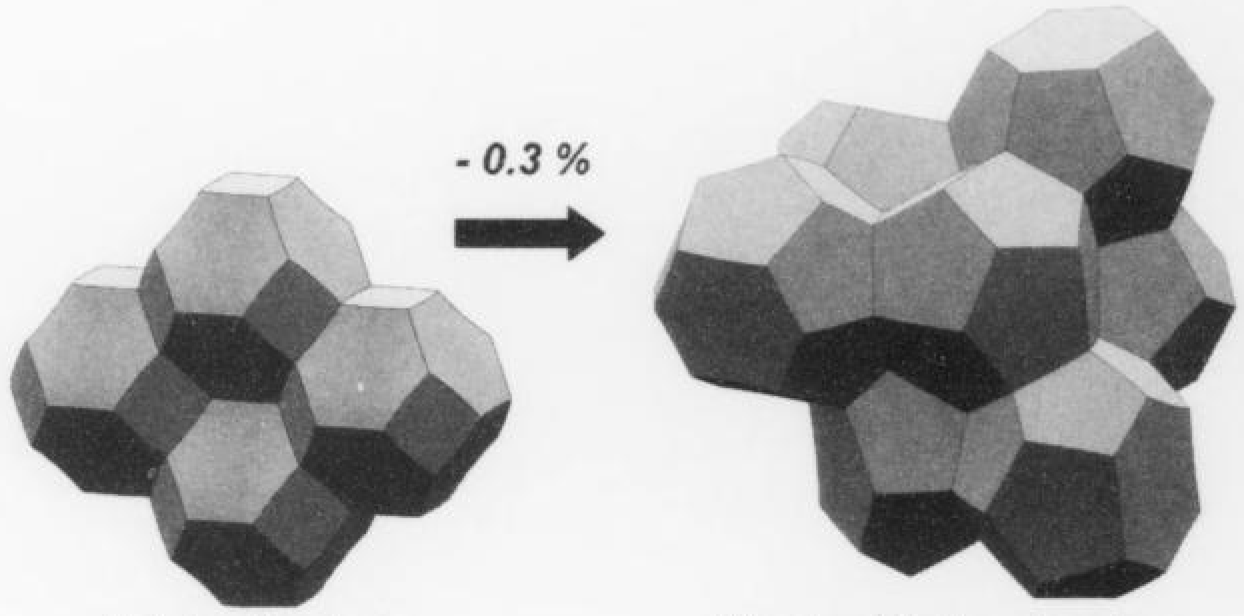


Figure 4. The Weaire-Phelan structure (right) decreases the total interface area by 0.3% in comparison to the structure conjectured by Lord Kelvin in 1887 (left).

AUTHOR

WIEBKE DRENCKHAN (29) is a postdoctoral fellow in Trinity College Dublin in Ireland. Having escaped the earnestness of Germany, she made her way to the Emerald Island with a physics degree from New Zealand. In the Irish environment she now finds ample inspiration for her research into the physics of "soft materials", specifically those composed of liquids and bubbles. Her more general perspective on the science community finds expression through the convincing powers of a drawing pen—her cartoons and illustrations regularly appear in books and journals.

Physics Department
Trinity College Dublin
Dublin 2
Ireland
e-mail: wiebke@phy.tcd.ie

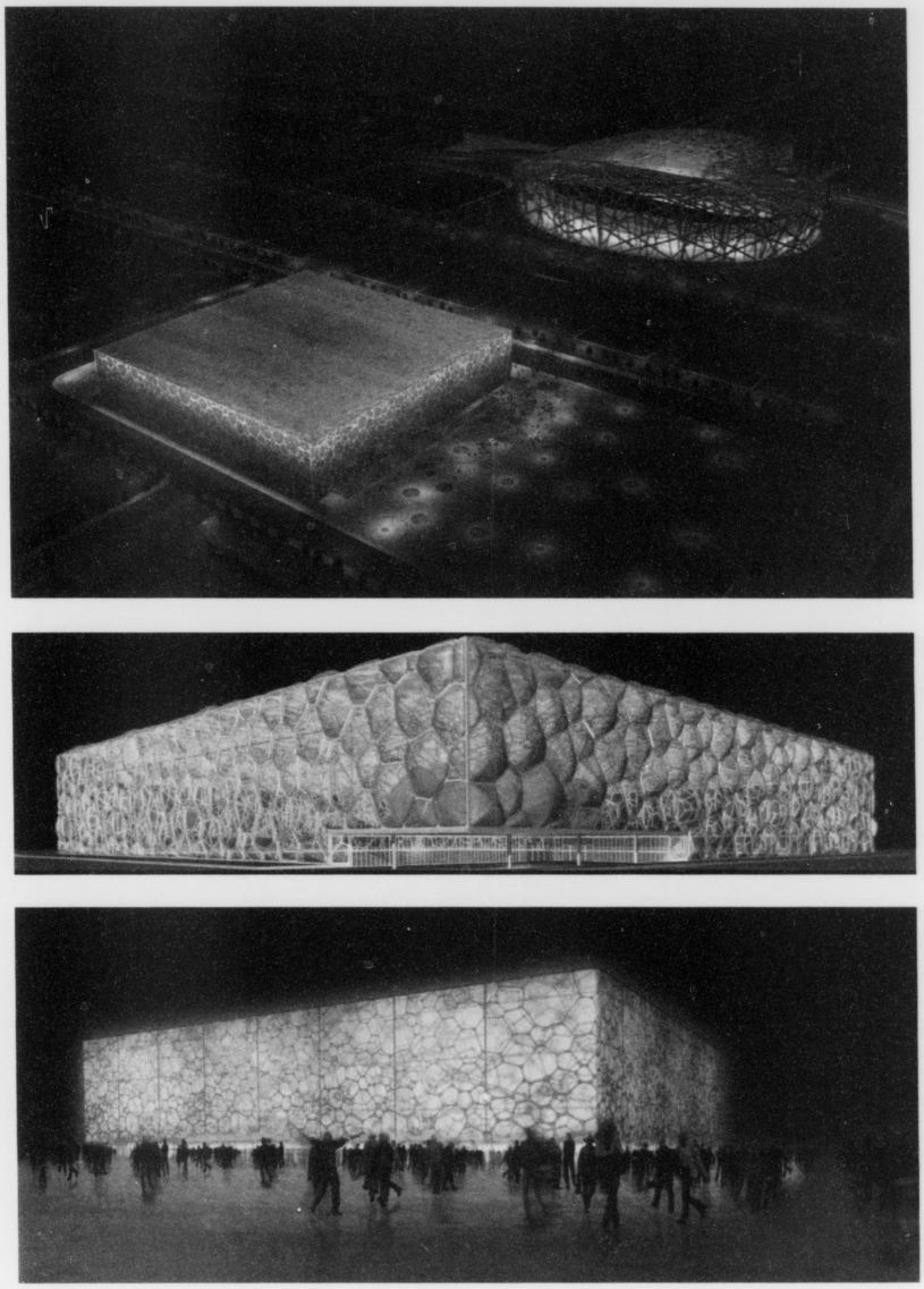


Figure 5. The architect's vision of the "WATER CUBE" (in the foreground) in front of the main stadium. The award-winning design of the National Swimming Centre for the Olympic Games in 2008 in Beijing is based on the Weaire-Phelan structure. It was developed by consulting engineers Arup and partners, PTW Architects, the China State Construction and Engineering Corporation, and the Shenzhen Design Institute, who provided the image.

The pub is "O'Neill's" beside Trinity College Dublin (TCD), the oldest of the Irish universities, founded by Queen Elizabeth in 1591 (Fig. 3). Since then it has been home to many great scientific minds, such as Sir William Rowan Hamilton, Georges Francis Fitzgerald, and Earnest Walton. The pub and its many local competitors have served TCD academics from all fields as a place to discuss ideas and relax over a drink with colleagues.

As you listen in amazement, you hear about a guy who kept coming into the pub with strangely shaped glass tubes to study the properties of the beer froth. People quickly stopped slagging[1] him when not much later an article in the *Irish Times* announced the discovery of the now world-famous "Weaire-Phelan structure". With this, physics Professor Denis Weaire and his PhD student Robert Phelan had found a specific cell arrangement, which brought them 0.3% ahead of a long historical hunt for the optimal packing of dry bubbles. Weaire's adventures with bubbles, and much else, are described in his book [4].

The previous best had been established in 1887 in a conjecture from no less a scientist than the great Scotsman Lord Kelvin, who had posed this question as part of his lifelong quest for a material structure of the ether of space. As an optimal solution he had suggested a structure of identical 14-sided cells, called tetrakaidecahedra (left side of Figure 4). And even though the ether disappeared from scientific debate, this question continued to pose a challenge to the mathematically minded.

The Kelvin conjecture remained questionable but unsurpassed for more than 100 years until the two TCD physicists tackled it by fruitfully combining familiarity with tetrahedrally bonded crystal structures and computer power. The crystal structure A15 or rather its dual, the Clathrate I structure, seemed a promising candidate. Using the "Surface Evolver" software developed by Ken Brakke [6, 7], they soon provided confirmation of their claim and also the precise shapes of the polyhedral bubbles (right side of Figure 4). Unlike the Kelvin structure, that of Weaire and Phelan consists of two different types of bubbles. In both structures some of the interfaces and edges are slightly curved to accommodate Plateau's laws.

The 1993 discovery of the Weaire-Phelan structure caused quite a stir among mathematicians and physicists; and nobody has yet come up with a better variation, despite the availability of ever-increasing computer power. The chances for seeing an analytical proof reasonably soon are slim. That's not to say that nobody has climbed the foothills of this mountain of proof [8]. Weaire says it won't happen in his lifetime, but perhaps in that of the redoubtable Hales.

[1]Second most popular Irish word, which can be translated as: "making fun of".

Figure 6. The sculpture "Throwing Shapes", which presents a section of the Weaire-Phelan structure. It was built by D. Grouse (left), and designed by W. Drenckhan (middle) and D. Weaire (right). A miniature version is owned by the London Science Museum.

That a small number like 0.3% is enough to become world famous is reflected beyond the scientific debate. The Weaire-Phelan structure recently transcended its status as a somewhat abstruse scientific problem when the Chinese authorities chose to make use of its artistic appeal and unique mechanical properties in the design of the National Swimming Centre for the Olympic Games 2008 in Beijing (Figure 5). This 86m award-winning "Water Cube" will be made of 4500 gigantic Weaire-Phelan bubbles enclosed in a plastic skin. The design is not only pleasing to the visitor's eye but also creates a greenhouse-effect to heat building and water. A convenient and well-considered side effect: the mechanical flexibility of the three-connected structure makes the building well suited to the seismic exposure of the area.

On a much smaller scale, but closer to home, Trinity College Dublin has recently acknowledged Weaire and Phelan's achievement by adorning its campus with a sculpture whose mirror-faced polyhedra represent a section of the famous bubbles (Fig. 6). Now in the

Figure 7. Panorama image of the Front Square of Trinity College Dublin (photograph by S. Wanja).

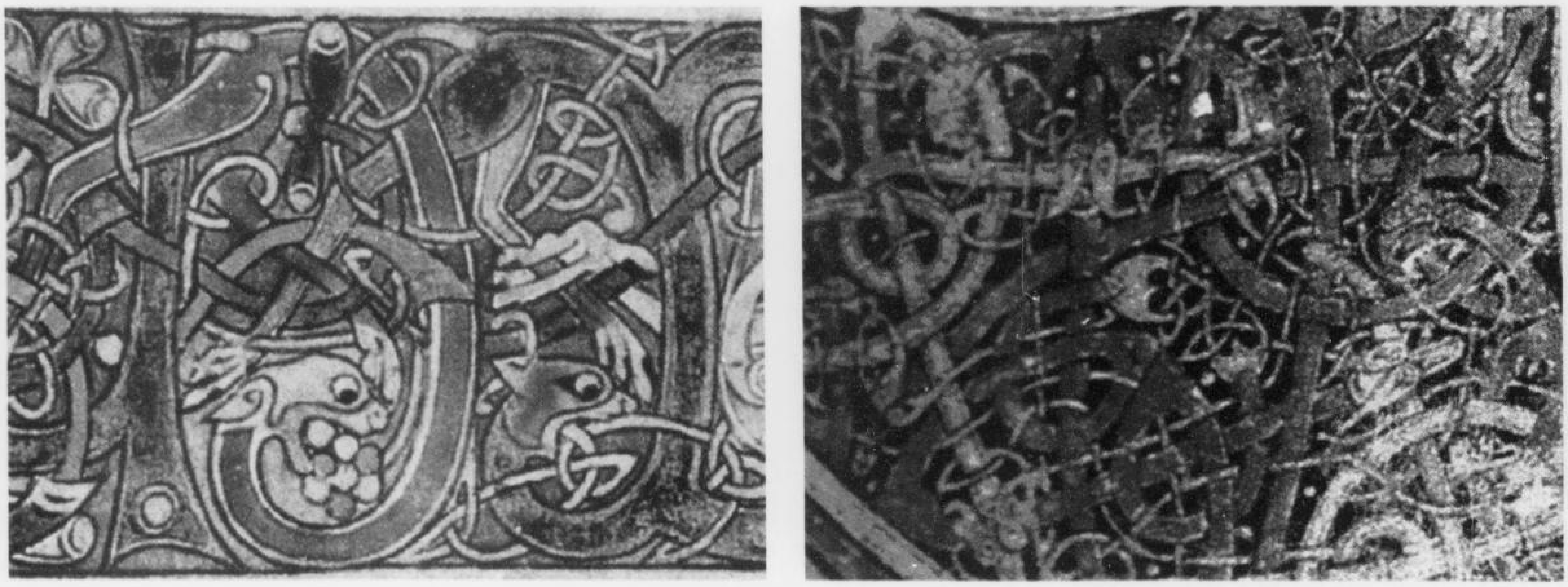


Figure 8. A dream for knot-theorists: sections of an infinite wealth of Celtic knots found in the "Book of Kells" in the Old Library of Trinity College Dublin.

hallway of the physics department, it will soon withstand the tempers of the Irish weather at the modern east-end of the historic campus.

As the pub owner finishes his fascinating story, you notice the reflection of the sunlight in your empty pint glass: the perfect time to go for a stroll through the historic campus. At the entrance of TCD you come face-to-face with a sculpture of the mathematician George Salmon (of conic sections fame). And you find more representations of mathematicians in the Long Room (the old library) and the many portraits that adorn the college. But it is still true that if you mention Synge in the pub, people will assume that you mean J. M. Synge, author of *Playboy of the Western World*, not his nephew J. L. Synge, famous in mechanics and relativity.

Trinity's families have excelled in both the sciences and the arts over successive generations. Some tried their hand at both—W. R. Hamilton is the most notorious example, combining sublime mathematics with lugubrious poetry! As part of the celebrations of the 200th anniversary of his birthday, a specific edition of 10 Euro coins can be purchased from the Central bank of Ireland; a special edition of stamps, from the Irish post.

The Old Library is the place to go for anybody with an interest in knot theory. On display here is the "Book of Kells", one of the most precious books in the world. The collection of the four ornately illustrated gospels of the bible, produced by Celtic monks around AD 800, is illuminated with a plethora of the famous Celtic knots (Fig. 8).

As you arrive at the bubble sculpture at the back end of Trinity, the rain returns. By now an experienced tourist in Ireland, you quickly make your way to the next pub. You are relieved to find out that this one is named after Mahaffy, the first professor of ancient Greek in TCD at the end of the 19th century, who had no interest in mathematics whatsoever!

But hang on a minute, was it not a Greek mathematician, Pappus of Alexandria, who first wrote about the "Sagacity of the bees" in his fifth book, about 300 AD? . . .

REFERENCES

1. Plateau, J. A. F. *Statique Experimentale et Theoretique des Liquides soumis aux seules Forces Moleculaires*. Gauthier-Villars, Paris, 1873.
2. Hales, T. C. The Honeycomb Conjecture. *Disc Comp Geom*, (2001), **25,** pp. 1–22.
3. Thompson, D'A. W. *On Growth and Form*. complete revised ed., New York, Dover Publications, 1992.
4. Weaire, D. and Aste, T. *The Pursuit of Perfect Packing*, Bristol and Philadelphia, Institute of Physics Publishing, 2000.
5. Fejes Toth, L. What the bees know and what they do not know. *Bull. Am. Math. Soc.* (1964), **70,** pp. 468–481.
6. Brakke, K. The surface evolver. *Exp Math*, (1992), **1,** pp. 141–156.
7. Brakke, K. [Available from: http://www.susqu.edu/facstaff/b/brakke/evolver/]
8. Kusner, R. and J. M. Sullivan. Comparing the Weaire-Phelan Equal-Volume Foam to Kelvin's Foam. 71–80 of *The Kelvin Problem*, Taylor & Francis (1996).

From Fractal Geometry to Fractured Architecture: The Federation Square of Melbourne

JOE HAMMER

Does your hometown have any mathematical tourist attractions such as statues, plaques, graves, the café where the famous conjecture was made, the desk where the famous initials are scratched, birthplaces, houses, or memorials? Have you encountered a mathematical sight on your travels? If so, we invite you to submit to this column a picture, a description of its mathematical significance, and either a map or directions so that others may follow in your tracks.

Please send all submissions to
Mathematical Tourist Editor,
Dirk Huylebrouck, Aartshertogstraat 42,
8400 Oostende, Belgium
e-mail: dirk.huylebrouck@ping.be

Melbourne, the second largest city in Australia, has a population of around four million. The capital of the state of Victoria, it lies on the southeast coast of the continent at the mouth of the Yarra River. Interestingly, Melbourne's location was determined by this modest river: in 1835 a city founder proclaimed, 'This is the spot for a village'. Some Melbournians still fondly call the city proper 'the village'. By 1900, the village had exploded to a populous vibrant city, thanks to the Gold Rush and excellent port facilities. Subsequently, an extensive railway network was built to transport people from the city, where they worked, to spreading suburbs.

Thoughtlessly, the builders of the bulk of the railway lines ran them along the Yarra and cut the city in two. By the beginning of the twentieth century, the riverside railway yard of over 50 lines became an eyesore that hampered the development of an otherwise vibrant metropolis. To remove this unsightly railway yard was out of the question; the only rational alternative was to roof it over. This immense engineering problem generated a second issue: any such development would create new plum real estate at the city's very gateway. What should arise on this deck?

Through the 20th century, countless ideas were proposed for the future of this site. Both private enterprise and government bodies held assorted competitions for the development. Most were just unrealistic dreams. For example, in 1975 some 2300 entries were submitted to a competition held by the Victorian State government. A member of the judging panel described entries of the finalists as 'a megalomania that makes the pyramids look like pimples'.

In the 1990s came the breakthrough. As Australia prepared to celebrate its centennial of Federation in 2001, the Federal Government sought projects appropriate to this celebration. Melbournians recognised in this an opportunity to turn their historical local village development dream into a federation project.

In 1997, with substantial financial support from the Federal Government, the State government of Victoria advertised an international architectural

The Main Components of the Project

1. A Civic Plaza (or The Square): capable of accommodating up to 15,000 people. The design of the square was the key component in the competition.
2. The Atrium: a covered public thoroughfare which complements the plaza, it can accommodate about 1,000 people.
3. An Art Gallery complex, which is an extension of the world-renowned National Gallery of Victoria (NGV), just over the river from Federation Square. It comprises more than 7000 square metres of exhibition space, and it houses about 22,000 Australian art objects.
4. Australian Centre for the Moving Image: includes several cinemas with the latest technology for preservation, exhibition, and education relating to moving images. This complex complements the Arts Centre across the Yarra, where theatres and a concert hall provide venues for live performances.
5. A variety of eateries, shops, and other service areas, including a riverside reception venue and multi-level car park.
6. All the above elements are built on a deck of 4 hectares (55,000-square metres) roofing the railway yard. The construction of the deck itself is a remarkable engineering achievement.

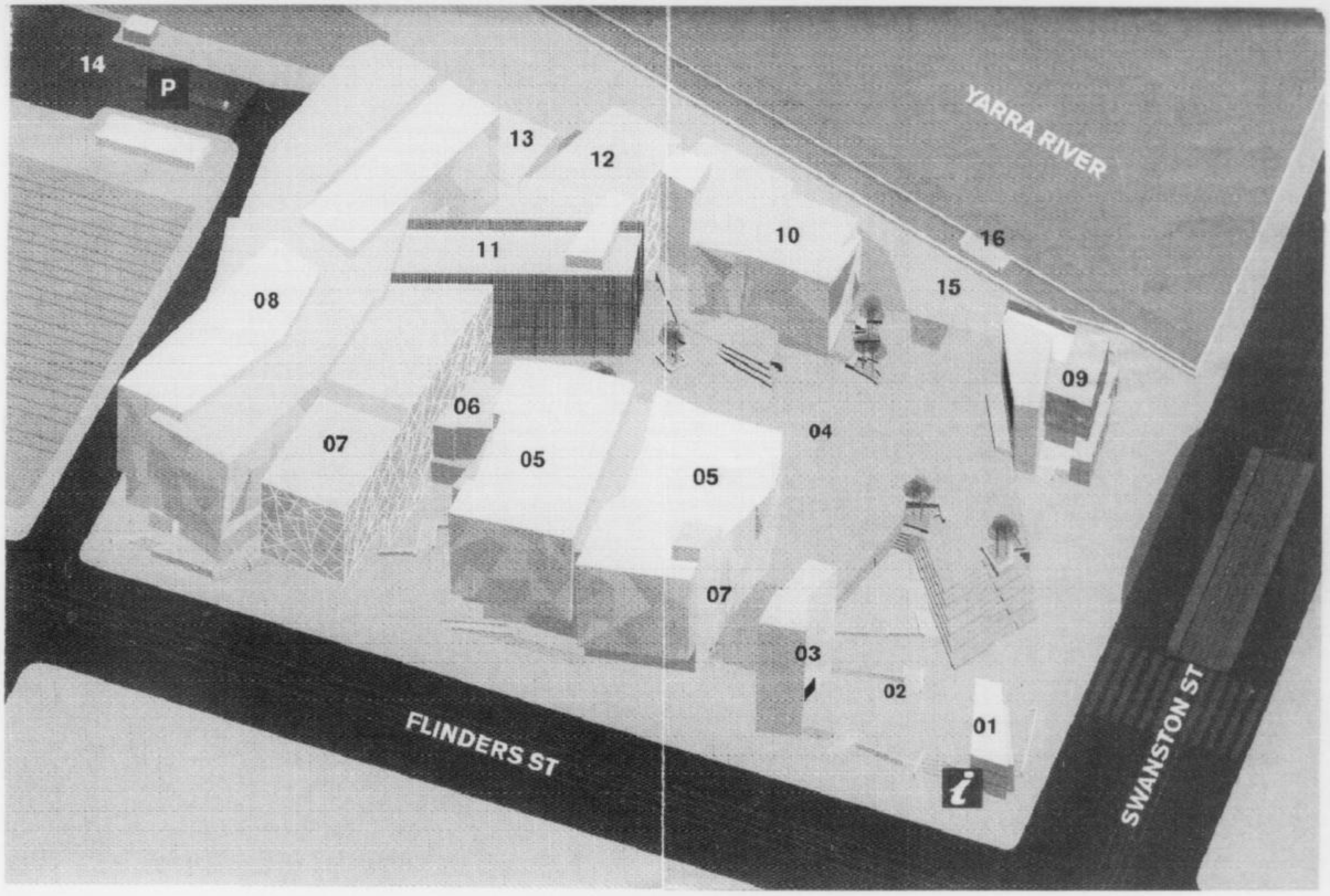


Figure 1. The layout of Federation Square. Key: 01, Information Centre; 02, St Pauls Court; 03, Retail; 04, The Plaza; 05, Cinema Centre; 06, Retail; 07, Atrium (Northern Part); 08, Art Gallery; 09, Hotel; 10, Museums; 11, Retail; 12, Atrium (Southern Part); 13, Function Centre; 14, Car Park; 15, River Terrace; 16, Wharf.

design competition for a centre of cultural activities. From 177 entries world wide, Lab Architectural Studio, based in London, was chosen unanimously by an international judging panel. Subsequently Lab, joined by Bates Smart Architects of Melbourne, was granted the job to develop the Federation Square complex. The principal architects were Peter Davidson and Donald L. Bates.

Before looking at the mathematics used in the project, it is helpful to read some selected comments from contemporary media which reflect the controversies and recognise the novelties in the design (Fig. 1). Most of the commenting authors are professionals in architecture or in art.

> Federation Square is as different and peculiar as if it came from another culture altogether [7]

> Contrary to the claims for its formal novelty and 'beauty', the design of this complex adheres to the conventions of humanism. [5]

> . . . [It] is the biggest example yet of a paradigm in architecture. [6]

> The façade is covered with the wonderful fractal patterning that is so arresting to the passersby. [4]

> It is a fusion of mathematics, art and architecture [4]

> It is a topological extension of the city. [5]

Geometry of the Façade—Art of Tilings

The façade of the three main buildings is a multi-faceted, multi-coloured composition, a trilogy in tiling design (Fig. 2). Visually it is comparable to some cubist paintings where the objects appear fragmented, faceted, and broken. So it is quite a surprise to learn that the geometric seed of this complicated-ap-

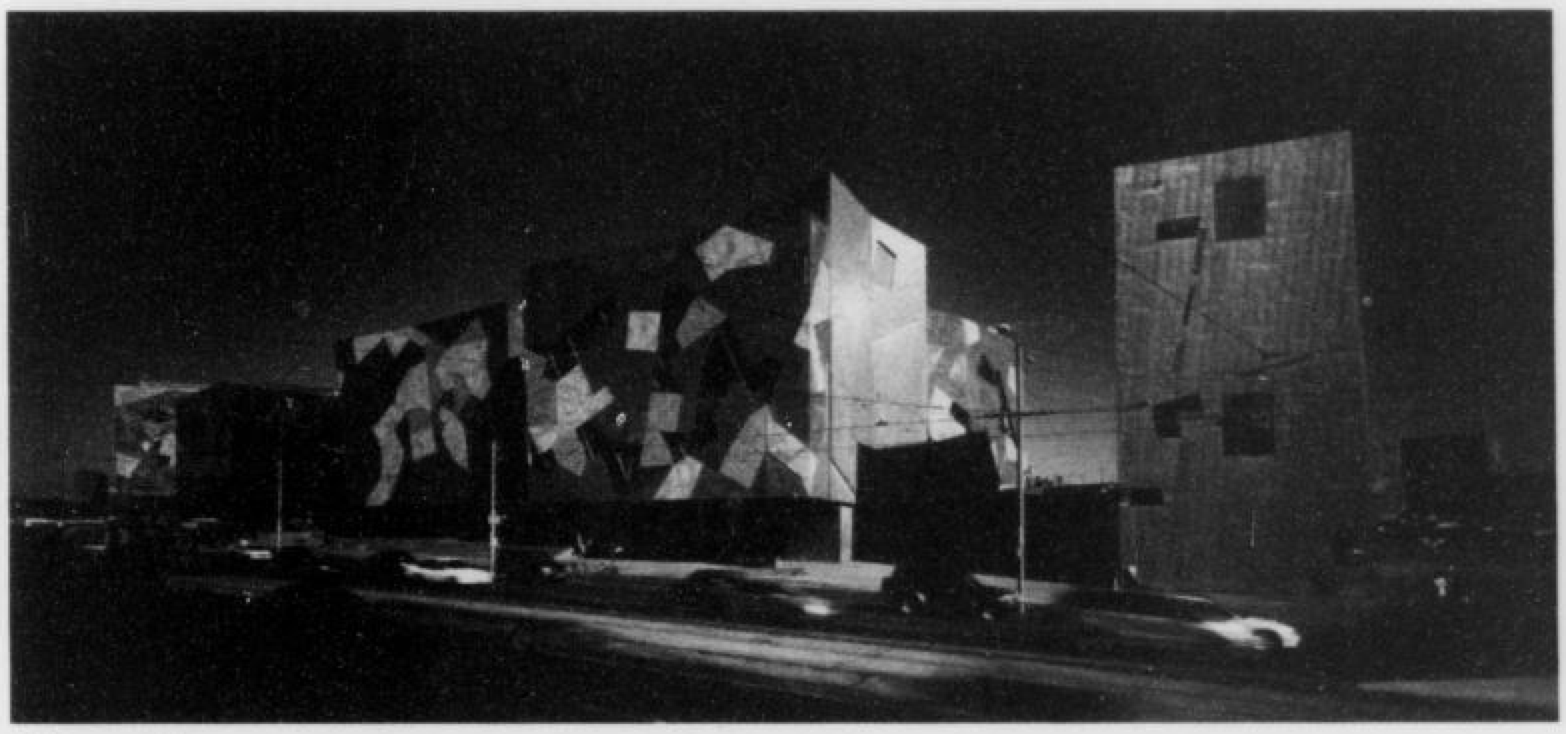


Figure 2.

pearing structure is a single triangle, namely a right-angled triangle with proportion of the perpendicular sides 1:2. (It is to be noted that the ratio of the sides of the largest face of a standard brick is also 1:2)

To see how the tiling was developed, let us play with a simple jigsaw puzzle. From a piece of cardboard, cut five identical right-angled triangles whose perpendicular sides measure, say, one centimetre and two centimetres. Assemble the five triangles in such a way as to obtain a larger version of the initial right-angled prototriangle. (This can be obtained by applying: rotations through one of the angles of the prototriangle, vertical reflection, and

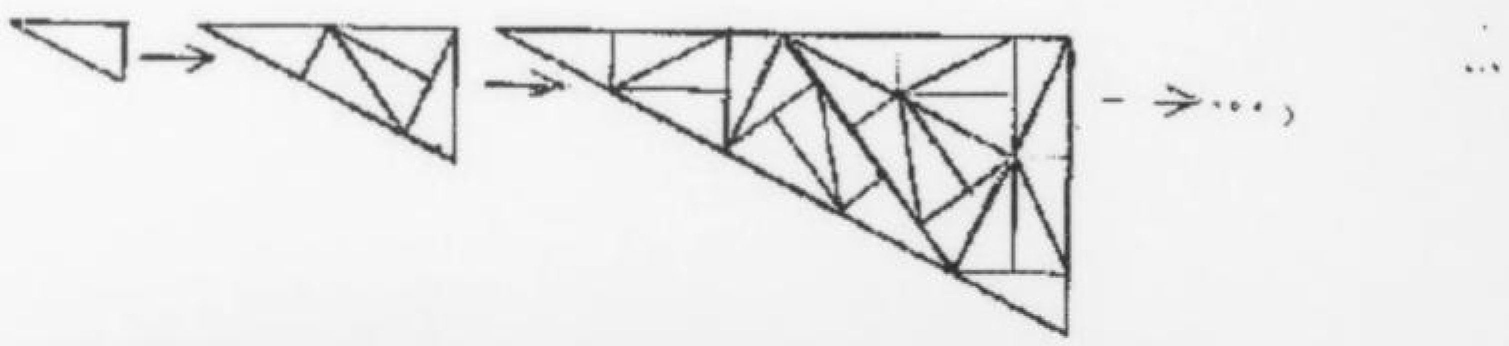
The façade of the three main buildings is a trilogy in tiling design.

appropriate translation, so that no two adjacent triangles are in the same alignment.) Call this new triangle a panel (Fig. 3). Next, assemble five panels the same way. Again we obtain a right-angled triangle similar to the prototriangle. Call this new triangle a megapanel. Subdivide each panel of the megapanel into prototriangles and observe that in the megapanel, the vertices of eight prototriangles meet at one point so that their sides are in a star-like, pinwheel formation (Fig. 3). The architects used such megapanels as cladding units to assemble the façade, where the pinwheel grid is the characteristic feature. The measurements of the actual building prototile are 0.6 metre, 1.2 metre, and $\sqrt{1.8}$ metre.

Mathematically we can continue this construction process to obtain an incremental sequence system of self-similar triangles. Radin [8], who designed the pinwheel tiling, also proved that the procedure provides an aperiodic tiling of the euclidean plane. Notice that periodic tiling, obviously possible for two panels, or two of any element of the sequence, can abut along the diagonals to form a rectangular tile.

Notice also, that we can reverse the above procedure: instead of assembling, sub-divide a right-angled triangle of ratio 1:2 into five congruent triangles similar to the initial triangle. Keep on sub-dividing the new smaller triangles or just some of them, and you obtain a decreasing sequence of fractal configurations. For further interesting problems regarding Radin tiling, consult Radin [8, 9] and the references therein.

We perceive that the façade is a finite part of an infinite iterative procedure. The pinwheel tiling starts and ends within the boundaries of the



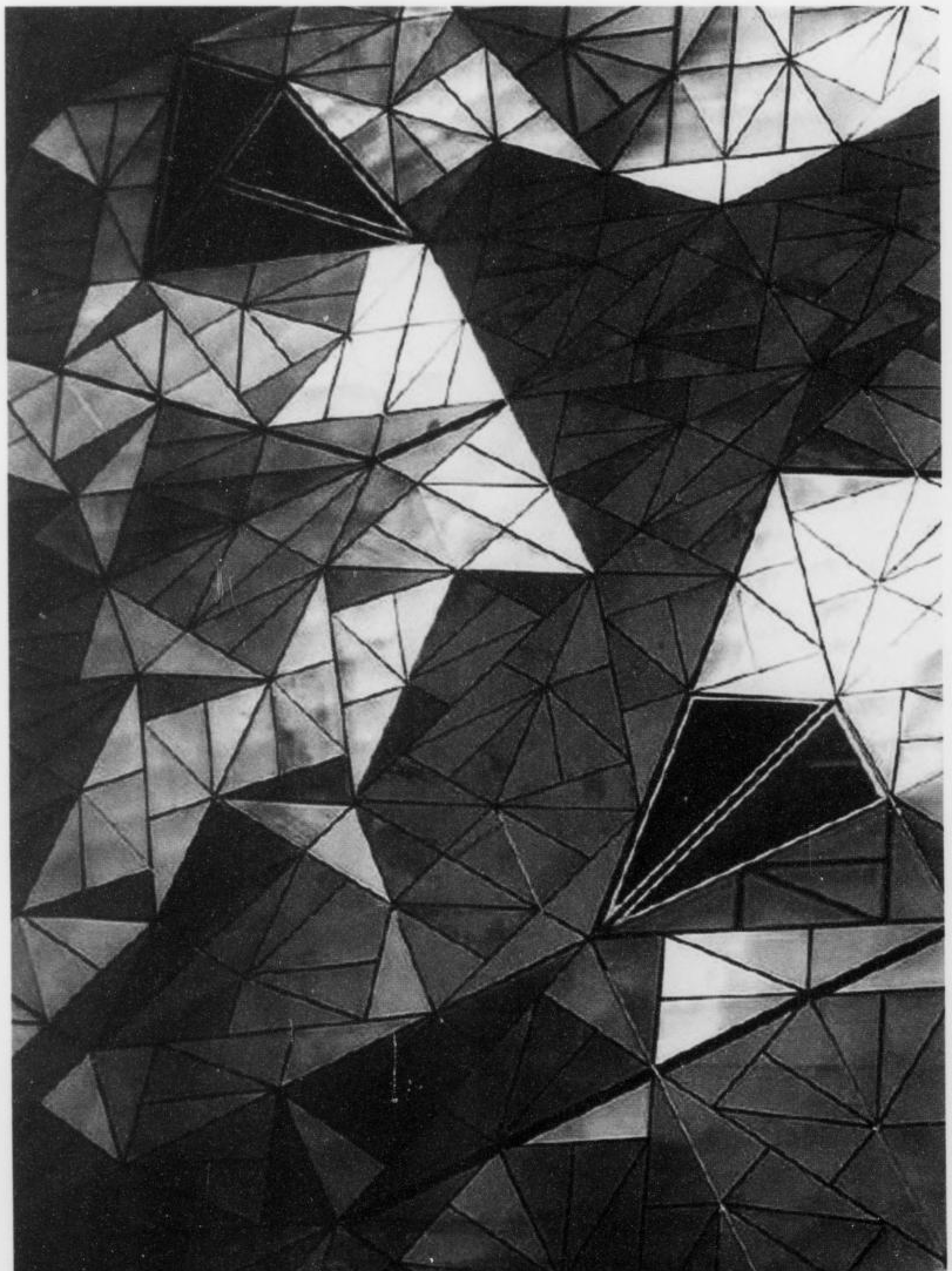


Figure 3.

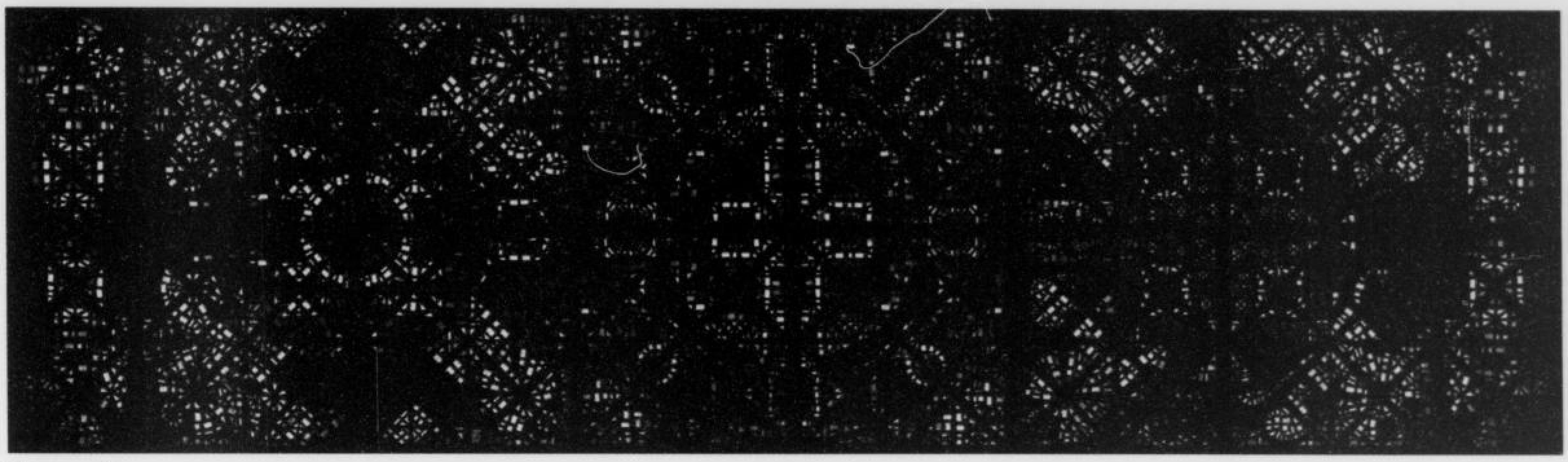


Figure 4.

façade. This project shows how mathematically motivated designs in architecture are restricted and determined by many factors such as construction, function, and material.

Three materials were used for the tiling: glass, sandstone, and zinc. The glass comes in opaque and clear, the sandstone is rough or polished, and the zinc is perforated or solid. The façade contains approximately 22,000 prototiles, about 2000 of them of glass, 8000 of sandstone, and 12,000 of zinc. However, there is a fourth material which contributes to the total picture. Through the glass panels, and elsewhere, we can see sections of the corrugated steel structure (Fig. 2) on which the panels are mounted—the skeleton of the façade. The geometry of the behind-the-scenes skeleton gives the facade its 'angled tangled' surface. The architects succeeded in selecting the materials and their colours to refer to the façades of the neighbouring buildings, despite its unorthodox fractured surface deviating from the usual façades.

> *Radin proved that the procedure provides an aperiodic tiling of the euclidean plane.*

Interestingly, even the basic geometric theme, the triangle itself, is featured in many places nearby. I will just mention two examples. One is the landmark Roy Grounds designed 162 metre-high spire of the Victorian Arts Centre (Fig. 2) almost opposite Federation Square. The entire structure is made up of a triangular mesh of sculptural beauty. The second example is the stunning stained glass ceiling of the Great Hall (Fig. 4) in the National Gallery of Victoria, designed and executed by Leonard French. It measures 51×15 metres and is constructed of approximately 10,000 pieces of multi-coloured hand-cut glass, embedded in a triangular grid. The twelve supporting columns are topped like the spikes of inverted umbrellas, a three-dimensional pin-wheel formation. (You can best enjoy this wonderful ceiling by lying supine in the finely carpeted hall, relaxing with all the other tourists doing the same! The effect is to make you feel that you are under branches of trees with budding coloured crowns.) In both projects the shape and translation are very different from that of the façade and, in fact, from each other.

The Civic Plaza—The Square

The Civic Plaza known as The Square. It is not a 'square' at all but rather an irregularly shaped, fractured square of area 7500 square metres containing several interconnected bays, each of which is designed as a mini-square suitable for intimate gatherings, such as family parties and open-air video showings. This is a fine example of the application of fractal self-similarity in architecture. It is apparently novel to design a plaza in this way. The ground of the plaza has a sloping topography, demanded by the structural need to roof the railway yard. However, the archi-

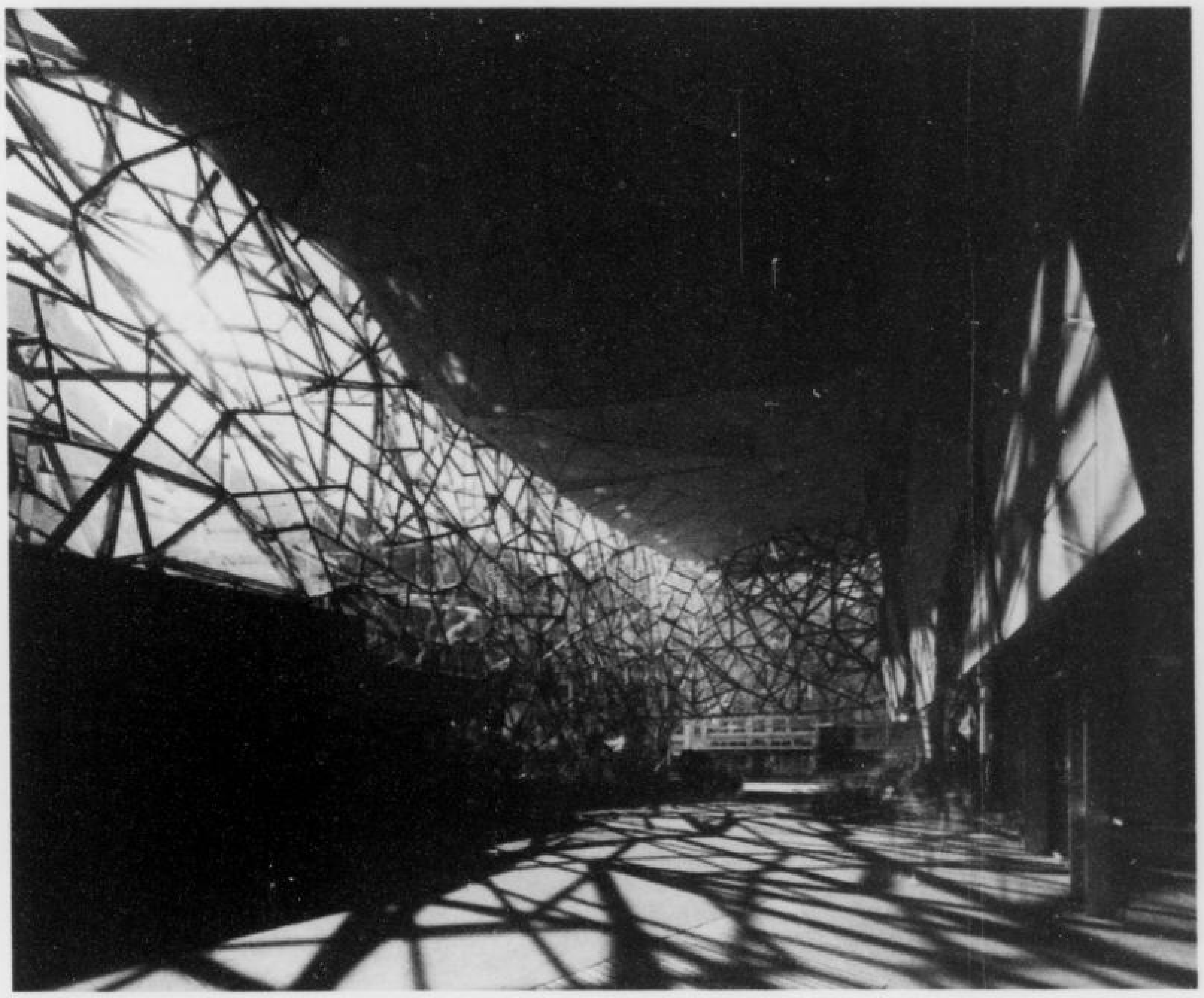


Figure 5

tect took advantage of this slope by transforming its surface to design a spectacular open amphitheatre. The ground inclines from street level by steps and slopes to a height of about 6 metres at the façade, allowing a panorama towards the river.

The paving of the ground is a remarkable work of art designed by Paul Carter [2]. It is surfaced with sandstone cobblestones laid in seemingly sinusoidal patterns of varying amplitudes or periods. There is a striking contrast between the undulating cobbled ground surface and the all-linear triangulated façade, a backdrop to the plaza. The many-coloured crystal-like cobbles emit a variety of sparkling colours, which change depending on how the sunlight strikes them. There is a delightful interplay between this colour spectrum and the multicoloured façade in the background.

These two elements of the plaza—the ground surface and the façade behind—make a coherent unit. If the artist Christo were to wrap the façade, the ground would definitely appear odd and different. Arguably, Lab Architectural Studio may have won the competition because of this imaginative design concept. Some critics noted that the façade and the interiors of the buildings are not related to each other. It seems these people failed to notice that the façade was designed to relate to the surface of the plaza.

The Atrium—Tourist in Wonderland

The atrium is the spine of the complex. It consists of three areas. The northern part (Fig. 5) is an entry from Flinders Street leading to the art gallery and cinemedia centre. In the middle section is a covered amphitheatre with seating capacity of 450, complementing the open amphitheatre of the square. The southern section slopes towards the river, accessed by an impressive stairway. The atrium cuts through the entire complex with a north–south axis providing a covered footbridge over the railway yards, linking the commercial quarters of the city with the Yarra River. It complements the open-air Princes Bridge. Probably necessity will eventually demand a covered footbridge over the river directly from the atrium.

In addition to the atrium's important functional facilities, it is the geometry of the steel network of the surrounding walls that is of particular interest for us. The network is a three-dimensional generalisation of the façade's pinwheel grid. With your megapanel

> *The design is a unique avenue of abstract sculptures composed of geometric themes.*

cardboard model, you can visualise this generalisation. Cut out a number of strips of elastic webbing and place them on the spokes of a pinwheel on your model. Fasten the ends of the elastic to the cardboard and then pull the centre ends up from the cardboard plane to the space in arbitrary directions. You will obtain skeletons of conical bodies. In the atrium's structure, the elastic strips are replaced by 200-mm-square hollow sections of corrugated steel tube segments. The tube segments are joined by sophisticated star-like steel connectors. The segments emanate from the connectors in up to six different directions, providing the three-dimensional version of the pinwheel. The intricate connectors made it possible to design a network of skeletons of a variety of three-dimensional configurations, such as tetrahedrons and prisms. The design is a unique avenue of abstract sculptures composed of geometric themes.

The steel structure is enveloped inside and out by tinted glass walls with sandblasted decoration. The glass consists of polygonal panes of nine different shapes that are derived from the façade's right-angled prototriangle. The actual measurement of the perpendicular sites of the 'protopane' triangles is 1 metre and 2 metres. Both the steel structure that supports the glass walls inside and the frames which connect the panes contribute to the sculptural effect of the enclosed main structure. Notable is the angled ceiling, made of acoustically effective translucent plastic material, which carries out the polygonal pattern of the glass walls. The geometry of the compound steel network is especially 'transparent' through the huge glass walls surrounding the amphitheatre. It appears as if Escher-type four-dimensional fantasy configurations come alive in our 'real' 3-space.

For a mathematical tourist, an extra night time visit will be rewarding. This is not only to enjoy the intricate theatrical lighting illuminating the atrium with an 'Alice in Wonderland' spell, but also to notice on the ground and elsewhere, pinwheel formations in the shadows projected from the 3-space skeletons branching out of the connectors (Fig. 4). As a bonus, the ever-changing lighting of the Cultural Centre spire is especially dramatic.

ACKNOWLEDGEMENTS

For much information and photographs, I thank the Federation Square management and the Art Gallery of Victoria. Thanks are also due to an anonymous referee for useful comments.

REFERENCES

1. Bovill, C., *Fractal Geometry in Architecture and Design,* Birkhäuser, 1996
2. Carter, P., *Mythform,* The Miegunyah Press, 2005
3. Crist, G., "Urbanity and Federation", *Architecture Australia* Vol. 92 (2003) pp. 56–57
4. Day, N., *Federation Square,* Hardie Grant Books 2003
5. Hartoonian, G., "The Tectonic of Camouflage", *Architecture Australia* Vol. 92 (2003) pp. 60–62
6. Jencks, C., "A New Heart of Melbourne", *Domus* 2003, Vol. 856 p. 675
7. Macarthur, J., "The Asthetics of Public Space", *Architecture Australia* Vol. 92 (2003) pp. 48–49
8. Radin, C., "The Pinwheel Tilings of the Plane", *Annals of Math* (2) Vol. 139 (1994) pp. 257–264
9. Ibid, *Miles of Tiles,* American Mathematical Society 1999
10. Szalpaj, P., *Contemporary Architecture and the Digital Design Process,* Elsevier, 2005. For further information visit the website: www.fedsq.com

School of Mathematics and Statistics
University of Sydney
NSW 2006
Australia
e-mail: hammerg@tpg.com.au

Principle of Continuity
A Brief History

Israel Kleiner

The Principle of Continuity was a very broad law, often not explicitly formulated, but used widely and importantly throughout the seventeenth, eighteenth, and nineteenth centuries. In general terms, the Principle of Continuity says that what holds in a given case continues to hold in what appear to be like cases. Specifically, it maintains that

(a) What is true for positive numbers is true for negative numbers.
(b) What is true for real numbers is true for complex numbers.
(c) What is true up to the limit is true at the limit.
(d) What is true for finite quantities is true for infinitely small and infinitely large quantities.
(e) What is true for polynomials is true for power series.
(f) What is true for a given figure is true for a figure obtained from it by continuous motion.
(g) What is true for ordinary integers is true for (say) Gaussian integers $\{a + bi: a, b \in Z\}$.

Each of these assumptions was used by mathematicians at one time or another, as we shall see. No doubt they realized that not *all* properties holding in a given case carry over to what appear to be like cases; they chose the properties that suited their purposes. Moreover, these purported analogies, even when they failed to materialize, were often starting points for fruitful theories.

André Weil, in his essay "From metaphysics to mathematics", gives poetic expression to some of the above thoughts [35, p. 408]:

> Mathematicians of the eighteenth century were accustomed to speak of "the metaphysics of the calculus", or "the metaphysics of the theory of equations". They understood by this a vague set of analogies, difficult to grasp and difficult to formulate, which nonetheless seemed to them to play an important role at a given moment in mathematical research and discovery. . . .
>
> All mathematicians know that nothing is more fertile than these obscure analogies, these troubled reflections of one theory in another, these furtive caresses, these inexplicable misunderstandings; also nothing gives more pleasure to the investigator. A day comes when . . . the metaphysics has become mathematics, ready to form the material whose cold beauty will no longer know how to move us.

We may begin our story with Kepler, although the Principle of Continuity, in one form or another, was used implicitly earlier, as we shall see. In the early seventeenth century Kepler enunciated a Principle of Continuity in connection with his study of conics. All conics, he claimed, are of the same species. For example, a parabola may be regarded as a limiting case of an ellipse or a hyperbola, in which one of the foci has gone to infinity. And "a straight line goes over into a parabola through infinite hyperbolas, and through infinite ellipses into a circle" [33]. (Desargues and Pascal thought along similar lines.) See also [24].

It was Leibniz, however, who made the Principle of Continuity—he called it *lex continui*—into an all-embracing law. (He owed some of his ideas to Descartes.) It appears throughout his work—in mathematics, philosophy, and science. Here are several ways in which he expressed it [15, pp. 291–294]:

(i) Nature makes no leaps. . . . We pass from the small to the great, and the reverse, through the medium.
(ii) When the essential determinations of one being approximate those of another, all the properties of the former should also gradually approximate those of the latter.
(iii) Since we can move from polygons to a circle by a continuous change and without making a leap, it is also necessary not to make a leap in passing from the properties of polygons to those of a circle, otherwise the law of continuity would be violated.

Leibniz's rationale for this encompassing principle was that "the sovereign wisdom, the source of all things, acts as a perfect geometrician. . . . [And geometry is] but the science of the continuous" [15, p. 292].

In this article, I will focus on examples from several areas of mathematics—analysis, algebra, geometry, and number theory—to illustrate the Principle of Continuity "in action", in its various guises. I will also highlight in each case the transition from the metaphysics to the mathematics, from vague analogies to fruitful theories.

Analysis

(a) The seventeenth century saw the rise of calculus/analysis, one of the great intellectual achievements of all time. It was founded independently by Newton and Leibniz during the last third of that century, although practically all of the prominent mathematicians of Europe around 1650 could solve many of the problems in which elementary calculus is now used. At the same time, it took another two centuries to provide the subject with rigorous foundations. The immediate task of Newton and Leibniz—the "basic problem"—was this:

BASIC PROBLEM: To devise general methods for discovering and deriving results in analysis.

It is in response to "basic problems" that Principles of Continuity were devised.

Central to Leibniz's approach in dealing with this problem was the notion of "differential", the difference between two infinitesimally close points. He computed with differentials as if they were real numbers, although he at times had to make "adjustments". Here is an example:

Leibniz searched for some time to find the rules for differentiating products and quotients. When he found them, the "proofs" were easy. Here is his discovery/derivation of the product rule:

$$\begin{aligned} d(xy) &= (x + dx)(y + dy) - xy \\ &= xy + xdy + ydx + (dx)(dy) - xy = xdy + ydx. \end{aligned}$$

Leibniz omits $(dx)(dy)$, noting that it is "infinitely small in comparison with the rest" [11, p. 255].

The dx and dy are the differentials of the variables x and y, respectively. The notions of derivative and of function—used nowadays to formulate the product rule—were introduced only in the following century (though Newton's "fluxion" is a derivative with respect to time). Note that Leibniz has here both discovered and derived the product rule. Discovery and derivation ("proof") often went hand-in-hand. Of course Leibniz's demonstration would not be acceptable to us, but standards of rigor have changed, and in any case contemporaries of Leibniz were, for the most part, not looking for rigorous proof. (But see [23] for an example of rigorous proofs given by Leibniz. The article was first published in 1993, so its contents might not have been known to his contemporaries.) They were satisfied with what Pólya would call "plausible reasoning" [30] and what Weil would describe as "metaphysics".

THE METAPHYSICS (1670s–): What holds for the real numbers also holds for the "hyperreal" numbers (essentially, the reals and the infinitesimals/differentials), *with some exceptions* (in this case, ignoring higher differentials).

BASIC PROBLEM: To determine which concepts and results of the calculus are transferable from the reals to the hyperreals. Put another way, to give precise meaning to the exceptions.

It took 300 years to fix the problem, to turn the metaphysics into mathematics. The fixing was done by Robinson.

THE MATHEMATICS (1960): Robinson's nonstandard analysis.

Robinson and Keisler explain the long delay:

> What was lacking at the time [of Leibniz] was a formal language which would make it possible to give a precise expression of, and delimitation to, the laws which were supposed to apply equally to the finite numbers and to the extended system including infinitely small and infinitely large numbers [32, p. 266].
>
> The reason Robinson's work was not done sooner is that the Transfer Principle for the hyperreal numbers is a type of axiom that was not familiar in mathematics until recently [17, p. 904].

The "formal language" was model theory, and the "Transfer Principle" was a law that decreed the conditions under which transferability of concepts and results between the reals and hyperreals was permissible.

Robinson saw nonstandard analysis as a vindication of Leibniz's (and Euler's) use of infinitesimals) [32, p. 2]. As he put it: "Leibniz's theory of infinitely small and infinitely large numbers, . . . in spite of its inconsistencies, . . . may be regarded as a genuine precursor of the theory in the present book" [32, p. 269]. He argued, moreover, that the history of the calculus had to be rewritten in light of nonstandard analysis [27, pp. 260–261]. Bos, in a spirited rejoinder, objected to these views [5, pp. 81–86]. As far as the Principle of Continuity goes, we do not claim that the Leibnizian calculus marched inexorably towards its natural resolution in nonstandard analysis, only that Robinson's work provided a rigorous justification of Leibniz's use of differentials. The same comment applies, *mutatis mutan-*

AUTHOR

ISRAEL KLEINER Originally a specialist in ring theory (Ph.D. McGill), **Israel Kleiner** has for the past twenty years concentrated on history of mathematics in relation to the teaching of mathematics. He served recently as Vice President of the Canadian Society for the History and Philosophy of Mathematics. He is a winner of the Pólya, Allendoerfer, and Ford prizes of the Mathematical Association of America.

Department of Mathematics & Statistics
York University
Toronto, ON M3J 1P3
Canada
e-mail: kleiner@rogers.com

dis, to our other examples. All that we claim in each case of transition from the metaphysics to the mathematics is that the latter was a suitable rigorous formulation of the former, not that it was the only way one could have gone.

(b) Already in the seventeenth century, but especially in the eighteenth, power series became a fundamental tool in analysis. They were usually treated like polynomials, with little concern for convergence (but see [23]). The operative (and philosophical) principle, even if not explicitly stated in general form, was that the rules applicable to polynomials could also be applied to power series. Newton, Euler, and Lagrange (among others) subscribed to this view.

An excellent example of Euler's use of these ideas is his discovery/derivation of the formula

$$1 + 1/2^2 + 1/3^2 + 1/4^2 + \ldots = \pi^2/6.$$

This is how he argues: The roots of sin x are 0, $\pm\pi$, $\pm 2\pi$, $\pm 3\pi$, . . . These, then, are also the roots of the "infinite polynomial" $x - x^3/3! + x^5/5! \ldots$, which is the power-series expansion of sin x. Dividing by x, hence eliminating the root $x = 0$, implies that the roots of $1 - x^2/3! + x^4/5! - \ldots$) are $\pm\pi$, $\pm 2\pi$, $\pm 3\pi$, ...

Now, the infinite polynomial obtained by expansion of the infinite product

$$[1 - x^2/\pi^2][1 - x^2/(2\pi)^2][1 - x^2/(3\pi)^2] \ldots$$

has precisely the same roots and the same constant term as $1 - x^2/3! + x^4/5! - \ldots$, hence the two infinite polynomials are identical (cf. the case of "ordinary" polynomials):

$$1 - x^2/3! + x^4/5! - \ldots = [1 - x^2/\pi^2][1 - x^2/(2\pi)^2][1 - x^2/(3\pi)^2] \ldots .$$

Comparing the coefficients of x^2 on both sides yields $-1/3! = -[1/\pi^2 + 1/(2\pi)^2 + 1/(3\pi)^2 + \ldots]$. Simplifying, we get $1 + 1/2^2 + 1/3^2 + \ldots = \pi^2/6$.

What a *tour de force*! One stands in awe of Euler's wizardry. The result was quite a coup for him: Neither Leibniz nor Jakob Bernoulli was able to find the sum of the series $1 + 1/2^2 + 1/3^2 + 1/4^2 + \ldots$. Note that, as in the previous example, discovery and demonstration went hand-in-hand, although even some of Euler's contemporaries objected to his demonstration.

The Metaphysics: What holds for polynomials also holds for power series.

Basic Problem: Justification of "algebraic analysis" (a term coined by Lagrange). That is, how do we justify analytic procedures by using formal algebraic manipulations?

What made seventeenth- and especially eighteenth-century mathematicians put their trust in the power of symbols? First and foremost, the use of such formal methods led to important results. Moreover, the methods were often applied to problems, the reasonableness of whose solutions "guaranteed" the correctness of the results and, by implication, the correctness of the methods. In an interesting article on eighteenth-century analysis, Fraser puts the issue thus [12, p. 331]:

> The 18th-century faith in formalism, which seems to us today rather puzzling, was reinforced in practice by the success of analytical [algebraic] methods. At base it rested on what was essentially a philosophical conviction.

Those attitudes gradually began to change. Two very important "practical" problems—the vibrating-string problem and the heat-conduction problem (of the eighteenth and early nineteenth centuries, respectively)—raised questions about central issues in calculus that could no longer be addressed by algebraic analysis. They necessitated, in particular, the clarification of the concepts of function, convergence, continuity, and the integral. This Cauchy proceeded to do. Thus,

The Mathematics: Cauchy (1820s) provided rigorous foundations for analysis by *eliminating* algebra as a foundational basis for calculus. He put it thus [14, p. 6]:

> As for my methods, I have sought to give them all the rigor which exists in [Euclidean] geometry, so as never to refer to reasons drawn from the generalness of algebra. Reasons of this [latter] type, though often enough admitted, especially in passing from convergent series to divergent series, and from real quantities to imaginary expressions, can be considered only . . . as inductions, sometimes appropriate to suggest truth, but as having little accord with the much-praised exactness of the mathematical sciences. . . . Most [algebraic] formulas hold true only under certain conditions, and for certain values of the quantities they contain. By determining these conditions and these values, and by fixing precisely the sense of all the notations I use, *I make all uncertainty disappear.*

Cauchy accomplished the task by selecting a few fundamental concepts, namely limit, continuity, convergence, derivative, and integral, establishing the limit concept as the one on which to base all the others, and deriving by fairly modern and rigorous means the major results of calculus. That this sounds commonplace to us today is in large part a tribute to Cauchy's program—a grand design, brilliantly executed.

Cauchy's new proposals for the rigorization of calculus generated their own problems and enticed a new generation of mathematicians to tackle them. He, too, was not immune to occasional metaphysical reasoning. For example, he believed that every continuous function is differentiable, except possibly at isolated points, and he "proved" the following

Theorem (1821): *An infinite sum (a convergent series) of continuous functions is a continuous function.*

The Metaphysics: Continuity of functions carries over from finite to infinite sums.

Cauchy's proof of the above theorem relied on infinitesimals; this masked the distinction between pointwise and uniform convergence of a series of functions. For an analysis of where Cauchy went wrong see [6]. Laugwitz [28] argues that with an appropriate interpretation of Cauchy's use of infinitesimals his proof can be made rigorous.

In 1826, Abel gave a counterexample to the above theorem. He put it delicately [6, p. 113]:

> But it seems to me that this [Cauchy's] theorem admits exceptions. For example, the series $\sin x - 1/2 \sin 2x + 1/3 \sin 3x - \ldots$ is discontinuous for every value $(2m+1)\pi$ of x, m being a whole number. There are, as we know, many series of this kind. [Note: $\sin x - 1/2 \sin 2x + 1/3 \sin 3x - \ldots = x/2$ for $x \in (-\pi,\pi)$, but if $x = \pi$, $\pi/2 \neq \sin \pi - 1/2 \sin 2\pi + \ldots = 0$.]

We should keep in mind that the concept of continuity is very subtle and was not very well understood in Cauchy's time. Moreover, "the fact that a statement has been refuted does not mean that it will be clear where the incriminating point lies" [6, p. 202]. The fact that there are different ways to consider convergence of series of functions emerged only gradually over the next several decades.

The Mathematics (Late 1840s): Seidel and Weierstrass introduced (independently) *uniform convergence* [6]. It is, of course, a *uniformly* convergent series of continuous functions that must be continuous.

Algebra

For about three millennia, until the early nineteenth century, "algebra" meant solving polynomial equations, mainly of degree four or less. This is now known as *classical algebra*. By the early decades of the twentieth century, algebra had evolved into the study of axiomatic systems, known collectively as *abstract algebra*. The transition occurred in the nineteenth century. In the first example, I focus on one aspect of this transition: English contributions to algebra in the first half of that century.

(a) The study of the solution of polynomial equations inevitably leads to the study of the nature and properties of various number systems, for of course the solutions of the equations are numbers. Thus the study of number systems constituted an important aspect of classical algebra.

The negative and complex numbers, although used frequently in the eighteenth century (the Fundamental Theorem of Algebra made them indispensable), were often viewed with misgivings and were little understood. For example, Newton described negative numbers as quantities "less than nothing," and Leibniz said that a complex number is "an amphibian between being and nonbeing." Although rules for the *manipulation* of negative numbers, such as $(-1)(-1) = 1$, had been known since antiquity, no mathematical *justification* for these rules had been given in the past.

During the late eighteenth and early nineteenth centuries, mathematicians began to ask *why* such rules should hold. Members of the Analytical Society at Cambridge University made important advances on this question. In the early nineteenth century Mathematics at Cambridge was part of liberal arts studies, and was viewed as a paradigm of absolute truths employed for the logical training of young minds. It was therefore important, these mathematicians felt, to base algebra, and in particular the laws of operation with negative numbers, on firm foundations [31].

Basic Problem: To justify the laws of manipulation with negative numbers. For example, why is $(-1)(-1) = 1$?

The most comprehensive work on this topic was Peacock's (1791–1858) *Treatise of Algebra* of 1830 (improved edition, 1845). (Peacock and other members of the Analytical Society were building on the ideas of seventeenth-century continental mathematicians [25].) His main idea was to distinguish between "arithmetical algebra" and "symbolical algebra." The former referred to laws and operations on symbols that stood only for *positive* numbers and thus, in Peacock's view, needed no justification. For example, $a - (b - c) = a - b + c$ is a law of arithmetical algebra when $b > c$ and $a > b - c$. It becomes a law of symbolical algebra if no restrictions are placed on a, b, and c. In fact, *no interpretation of the symbols is called for*. Thus *symbolical algebra* was the subject, newly founded by Peacock (and others), of operations with symbols that need not refer to specific objects, but that obey the laws of arithmetical algebra. (Cf. Newton's designation of algebra as "universal arithmetic".)

Peacock justified his identification of the laws of symbolical algebra with those of arithmetical algebra by means of his Principle of Permanence of Equivalent Forms. It said that [31, p. 38]

> Whatever form is Algebraically equivalent to another, when expressed in general symbols, must be true whatever those symbols denote. Conversely, if we discover an equivalent form in Arithmetical Algebra or any other subordinate science, when the symbols are general in form though specific in their nature [i.e., referring to positive numbers], the same must be an equivalent form, when the symbols are general in their nature [i.e., not referring to specific objects] as well as in their form.

In short, *the laws of arithmetic shall be the laws of algebra*. What these laws were was not made explicit at the time. The laws were clarified in the second half of the nineteenth century, when they turned into axioms for rings and fields [20], [21].

It is noteworthy that what we do in trying to clarify the laws that numbers obey is not very different from what Peacock did: we too decree what the laws of the various number systems shall be. These decrees we call axioms.

The Metaphysics: What holds for positive numbers holds for negative numbers.

Peacock's Principle of Permanence turned out to be very useful. For example, it enabled him to prove the following

Theorem (1845): $(-a)(-b) = ab$.

Proof. Since

$$(a - b)(c - d) = ac + bd - ad - bc \qquad (**)$$

is a law of arithmetical algebra whenever $a > b$ and $c > d$, it becomes, by the Principle of Permanence, a law of symbolical algebra, which holds without restriction on a, b, c, d. Letting $a = 0$ and $c = 0$ in (**) yields $(-b)(-d) = bd$.

Peacock's work, and that of others, signalled a fundamental shift in the essence of algebra from a focus on the *meaning* of symbols to a stress on their *laws of operation*. Witness Peacock's description of symbolical algebra [31, p. 36]:

In symbolical algebra, the rules determine the meaning of the operations . . . we might call them arbitrary assumptions, in as much as they are arbitrarily imposed upon a science of symbols and their combinations, which might be adapted to any other assumed system of consistent rules.

This was a very sophisticated idea, well ahead of its time. In fact, however, Peacock paid only lip service to the arbitrary nature of the laws. In practice, they remained the laws of arithmetic. In the next several decades English mathematicians put into practice what Peacock had preached by introducing algebras with properties which differed in various ways from those of arithmetic. In the words of Bourbaki [7, p. 52],

The algebrists of the English school bring out first, between 1830 and 1850, the abstract notion of law of composition, and enlarge immediately the field of Algebra by applying this notion to a host of new mathematical objects: the algebra of Logic with Boole, vectors, quaternions and general hypercomplex systems with Hamilton, matrices and non-associative laws with Cayley.

Thus, whatever its limitations, symbolical algebra provided a positive climate for subsequent developments in algebra. Laws of operation on symbols began to take on a life of their own, becoming objects of study in their own right rather than a language to represent relationships among numbers.

The Mathematics: Advent of abstract (axiomatic) thinking in algebra.

(b) Here is a sixteenth-century application of the Principle of Continuity: For centuries mathematicians adhered to the following view concerning square roots of negative numbers: since the squares of positive as well as of negative numbers are positive, square roots of negative numbers do not—in fact, can not—exist. All this changed in the sixteenth century, following the work on the solution of equations of several Italian mathematicians.

A solution by radicals of the cubic was first published by Cardano (1501–76) in *The Great Art* (referring to algebra) of 1545. What came to be known as Cardano's formula for the solution of the cubic $x^3 = ax + b$ was given by

$$x = \sqrt[3]{b/2 + \sqrt{(b/2)^2 - (a/3)^3}} + \sqrt[3]{b/2 - \sqrt{(b/2)^2 - (a/3)^3}}.$$

Square roots of negative numbers arise "naturally" when Cardano's formula is used to solve cubic equations. For example, application of the formula to the equation $x^3 = 9x + 2$ gives $x = \sqrt[3]{1 + \sqrt{-26}} + \sqrt[3]{1 - \sqrt{-26}}$.

What was one to make of this solution? Since Cardano was suspicious of negative numbers, he certainly had no taste for their square roots, so he regarded his formula as inapplicable to equations such as $x^3 = 9x + 2$. He concluded that such expressions are "as subtle as they are useless". Judged by past experience, this was not an unreasonable assumption.

The crucial breakthrough was achieved by Bombelli (1526–72). In his important book *Algebra* (1572) he applied Cardano's formula to the equation $x^3 = 15x + 4$ and obtained $x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$. But he could not dismiss this solution, for he noted (by inspection) that $x = 4$ is also a root of this equation. (Its other two roots, $-2 \pm \sqrt{3}$, are also real.) This gave rise to a paradox: while all three roots of the cubic $x^3 = 15x + 4$ are real, the formula used to obtain them involved square roots of negative numbers—meaningless at the time.

Basic Problem: How was one to resolve this paradox?

Bombelli adopted the rules for real quantities to manipulate "meaningless" expressions of the form $a + \sqrt{-b}$ ($b > 0$), and thus managed to show that $\sqrt[3]{2 + \sqrt{-121}} = 2 + \sqrt{-1}$ and $\sqrt[3]{2 - \sqrt{-121}} = 2 - \sqrt{-1}$, hence that $x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = (2 + \sqrt{-1}) + (2 - \sqrt{-1}) = 4$ [29, p. 18].

The Metaphysics: What holds for real numbers also holds for complex numbers.

Bombelli had given meaning to the "meaningless" by thinking the "unthinkable", namely that square roots of negative numbers could be manipulated in a meaningful way to yield significant results. As he put it [29, p. 19],

It was a wild thought in the judgment of many; and I too was for a long time of the same opinion. The whole matter seemed to rest on sophistry rather than on truth. Yet I sought so long, until I actually proved this to be the case.

This was the birth of complex numbers. But birth did not entail legitimacy. For the next two centuries complex numbers were shrouded in mystery, little understood, and often ignored. Only after their geometric representation in 1831 by Gauss as points in the plane were they accepted as *bona fide* elements of the number system. (The earlier works of Argand and Wessel on this topic were not well known among mathematicians.)

The Mathematics: Complex numbers are admitted as legitimate mathematical entities.

Geometry

For several millennia, until the early nineteenth century, "geometry" meant euclidean geometry. The nineteenth century witnessed an explosive growth in the subject, both in scope and in depth. New geometries emerged: projective geometry (Desargues's 1639 work on this subject came to light only in 1845), hyperbolic geometry, elliptic geometry, Riemannian geometry, and algebraic geometry. Poncelet (1788–1867) founded (synthetic) projective geometry in the early 1820s as an independent subject, but he lamented its lack of general principles. The proof of each result had to be handled differently. Thus, the following:

Basic Problem: To develop tools for the emerging subject of projective geometry.

This Poncelet did by introducing a Principle of Continuity in his 1822 book *Traité des propriétés projectives des figures*.

The Metaphysics: Poncelet's Principle of Continuity [8, p. 136]:

A property known of a figure in sufficient generality also holds for all other figures obtainable from it by continuous variation of position.

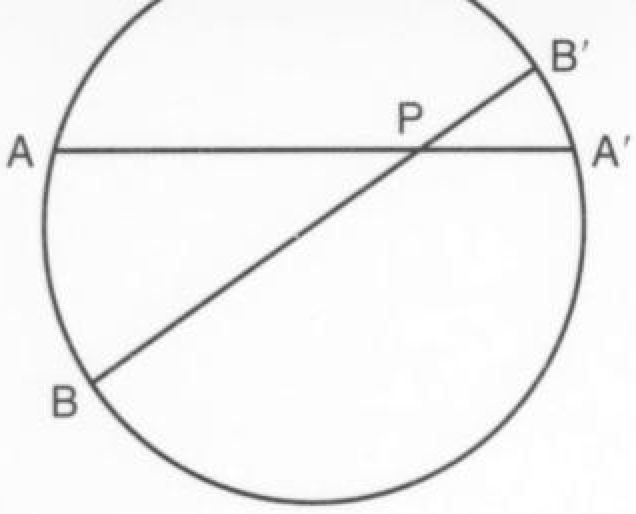


FIGURE 1

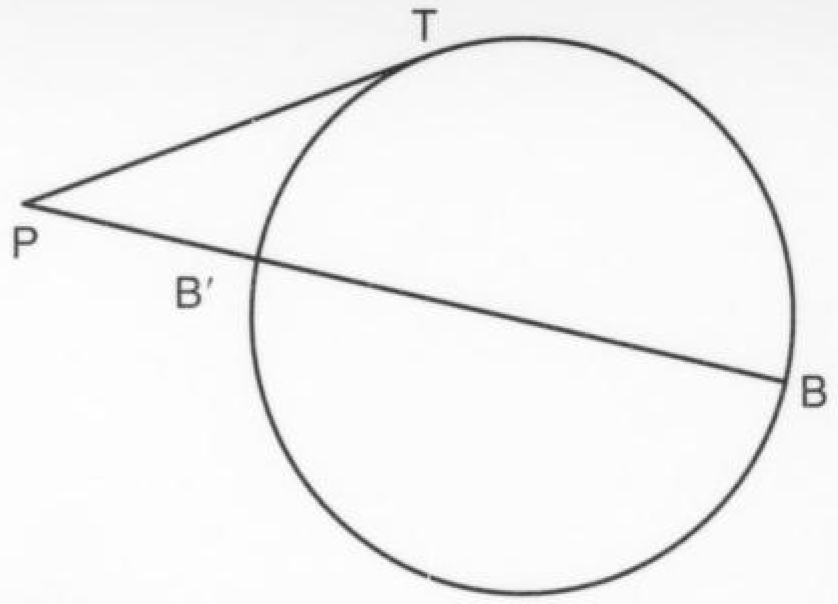


FIGURE 3

Note that here we have a formulation of the Principle of Continuity in which "continuity" in the traditional sense enters. (This aspect of the Principle of Continuity goes back to Kepler and Leibniz.) Of course, this formulation also falls under the rubric of the Principle of Continuity that I have given, namely that "what holds in a given case continues to hold in what appear to be like cases". Here the transition from a given case to what appears to be a like case is achieved via continuity in the traditional sense, whereas in the other examples we have considered (and in the ones that follow in the next section), the transition is achieved by extending the domain under consideration.

As an elementary illustration of his Principle, Poncelet cited the well-known (and easily established) theorem about the equality of the products of the segments of intersecting chords in a circle: $PB \times PB' = PA \times PA'$ (Fig. 1). The Principle of Continuity then implies that also $PB \times PB' = PA \times PA'$ (Fig. 2) and $PB \times PB' = (PT)^2$ (Fig. 3).

A much more substantial result that Poncelet proved using his Principle of Continuity was the so-called Closure Theorem.

CLOSURE THEOREM: Let C and D be two conics. Let P_1 be a point of C and L_1 a tangent to D through P_1. Let P_1, L_1, P_2, L_2, P_3, L_3, . . . be a "Poncelet transverse" between C and D, that is, P_i is on C, L_i is tangent to D and P_i is the intersection of L_{i-1} and L_i (Fig. 4). We say that the Poncelet transverse closes after n steps if $P_{n+1} = P_1$. The closure theorem says that if a transverse, starting at P_1 on C, closes after n steps, then a Poncelet transverse from *any* point on C will close after n steps (Fig. 5). Thus, if there is one inscribed n-gon between C and D, then there are infinitely many such n-gons. (Poncelet's formulation of this result is somewhat different [4]).

Bos *et al.* give three different proofs of the Closure Theorem: Poncelet's, in 1813–1814, using the Principle of Continuity, Jacobi's, in 1828, using elliptic functions, and Griffiths's, in 1976, using elliptic curves [4].

The Principle of Continuity was criticized (by, among others, Cauchy) for being vague, but it was a powerful tool, used by Poncelet to great effect to establish projective geometry as a central discipline. (It was he who coined the term "principle of continuity".)

A natural question arose: What *is* projective geometry? Two major issues emerged: the relationship of projective to euclidean geometry and the validity of the principle of duality. For Poncelet, the major problem of projective geometry was the determination of all properties of geometric figures that do not change under projections. In his development of the subject he used notions from euclidean geometry (length and angle). To him, projective geometry was a subgeometry of euclidean geometry. Other geometers began to believe that projective geometry is more basic than euclidean geometry. In 1859 Cayley showed that, in fact, euclidean geometry is a subgeometry of projective geometry.

Poncelet and others formulated the principle of duality in projective geometry. Although it appeared to be a working principle, its validity was in question. A vigorous debate raged in the early decades of the nineteenth century about the relative merits of the synthetic versus the ana-

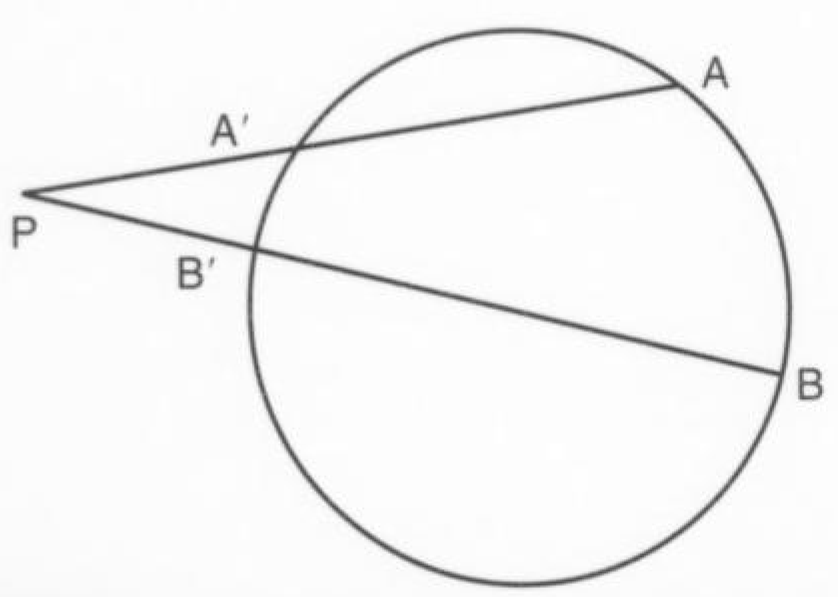


FIGURE 2

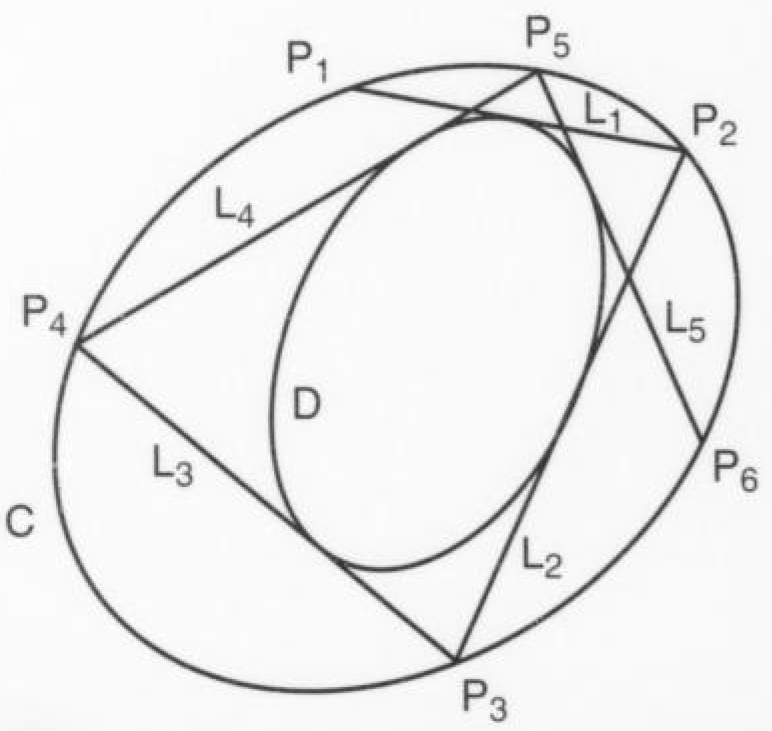


FIGURE 4

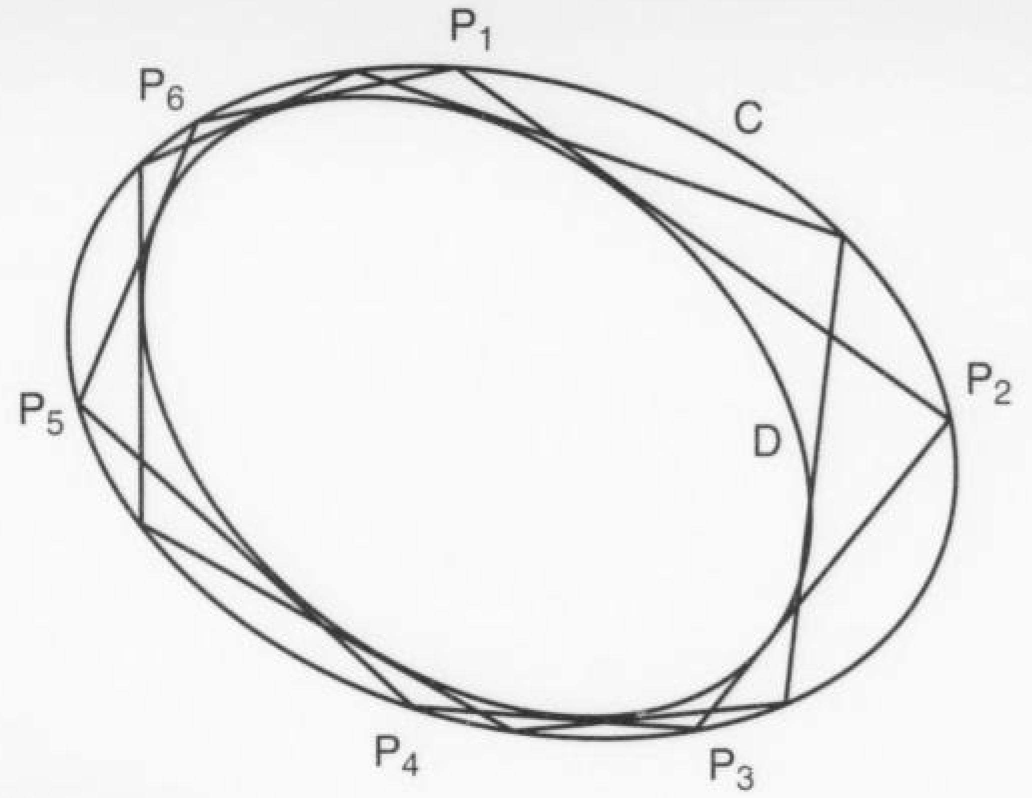


FIGURE 5

lytic approaches to geometry. The principle of duality seems to have been a test case for the two schools of thought. Poncelet, as we noted, developed projective geometry synthetically. Gergonne and Plücker were fervent proponents of the analytic approach. Both introduced homogeneous coordinates; this made the principle of duality (analytically) transparent. In 1882 Pasch supplied an axiomatic treatment of projective geometry, which made the principle of duality synthetically transparent.

In the second half of the nineteenth century the question about the nature of projective geometry was incorporated in a broader question: What is geometry? There were good reasons to pose this question.

The nineteenth century was a golden age in geometry. New geometries arose (as I have mentioned). Geometric methods competed for supremacy: the metric versus the projective, the synthetic versus the analytic. And important new ideas entered the subject: elements at infinity (points and lines), use of complex numbers (e.g., complex projective space), the principle of duality, use of calculus, extension of geometry to n dimensions, Grassmann's calculus of extension (this involved important geometric ideas), invariants (e.g., the Cayley-Sylvester invariant theory of forms), and groups (e.g., groups of the regular solids). An important development was Klein's proof that not only euclidean, but also noneuclidean geometry (both hyperbolic and elliptic) are subgeometries of projective geometry. For a time it was said that "projective geometry is all geometry". A broad look at the subject of geometry was in order.

THE MATHEMATICS: Klein's definition of geometry: the *Erlangen Program* (1872) [18].

In a lecture at the University of Erlangen, titled A Comparative Review of Recent Researches in Geometry, Klein classified the various geometries using the unifying notions of group and invariance. He defined a geometry of a set S and a group G of permutations of S as the totality of properties of the subsets of S that are invariant under the permutations in G. This conception of geometry, although not all-encompassing (for example, it excluded Riemannian geometry, of which Klein seems to have been unaware in 1872), had considerable influence on the development of the subject [3].

Under Klein's view of geometry, projective geometry (say of the plane) is the totality of properties of the projective plane left invariant under collineations (those transformations that take lines into lines). As for Poncelet's Principle of Continuity, its "mathematical content is today reduced to the identity theorem for analytic functions and the fundamental theorem of algebra" [8, p. 136]—which sounds like a declaration that today it must not be applied as a principle.

Number Theory

The study of number theory goes back several millennia. Its two main contributors in ancient Greece were Euclid (ca 300 BC) and Diophantus (ca 250 AD). Their works differ fundamentally, both in method and in content. Euclid's comprises Books VII–IX of the *Elements* and is in the "theorem–proof" style. Here Euclid introduced some of the subject's main concepts, such as divisibility, prime and composite integers, greatest common divisor, and least common multiple, and established some of its main results, among them the elucidean algorithm, the infinitude of primes, results on perfect numbers, and what some historians consider to be a version of the Fundamental Theorem of Arithmetic. (Much of the number-theoretic work in the *Elements* is due to earlier mathematicians.)

Diophantus's work is contained in the *Arithmetica*—a collection of about 200 problems, each giving rise to one or more diophantine equations, many of degree two or three. These are equations in two or more variables, with integer coefficients, for which the solutions sought are integers or rational numbers. Diophantus found rational solutions for these equations, often by ingenious methods. Their study has, since Diophantus, become a central topic in number theory [2], [34].

BASIC PROBLEM: To develop tools for solving diophantine equations.

Let us consider two celebrated examples.

(a) $x^2 + 2 = y^3$. This is a special case of the Bachet equation, $x^2 + k = y^3$ (k an integer), which is an important example of an elliptic curve. The case $x^2 + 2 = y^3$ appears already in the *Arithmetica* (Problem VI.17). Fermat gave its positive solution, $x = 5$, $y = 3$, but did not publish a proof of the fact that this is the *only* such solution. It was left for Euler, over 100 years later, to do that.

Euler introduced a fundamental new idea to solve $x^2 + 2 = y^3$. He factored its left-hand side, which yielded the equation $(x + \sqrt{2}i)(x - \sqrt{2}i) = y^3$. This was now an equation in a domain D of "complex integers", where $D = \{a + b\sqrt{2}i: a, b \in Z\}$. Here was the first use of complex numbers—"foreign objects"—in number theory.

Euler proceeded as follows: If a, b, c are integers such that $ab = c^3$, and $(a,b) = 1$, then $a = u^3$ and $b = v^3$, with u and v integers. This is a well-known and easily established result in number theory. (It holds with the exponent 3 replaced by any integer, and for any number of factors a, b, . . .) Euler carried it over—*without acknowledgment*—to the domain D.

Since $(x + \sqrt{2}i)(x - \sqrt{2}i) = y^3$, and $(x + \sqrt{2}i, x - \sqrt{2}i) = 1$ (Euler claimed, without substantiation, that $(m,n) = 1$ implies $(m + n\sqrt{2}i, m - n\sqrt{2}i) = 1$)), it follows that

$$x + \sqrt{2}i = (a + b\sqrt{2}i)^3 = (a^3 - 6ab^2) + (3a^2b - 2b^3)\sqrt{2}i$$

for some integers a and b. Equating real and imaginary parts, we get $x = a^3 - 6ab^2$ and $1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2)$. Since a and b are integers, we must have $a = \pm 1$, $b = 1$, hence $x = \pm 5$, $y = 3$. These, then, are the only solutions of $x^2 + 2 = y^3$ [34]. Now to our second example.

(b) $x^p + y^p = z^p$, p an odd prime. In 1847 Lamé claimed before the Paris Academy to have proved Fermat's Last Theorem, the unsolvability in integers of this equation, as follows:

Assume that the equation $x^p + y^p = z^p$ *has* integer solutions. Factor its left-hand side to obtain

$$(x + y)(x + yw)(x + yw^2) \ldots (x + yw^{p-1}) = z^p \quad (**),$$

where w is a primitive p-th root of 1 (that is, w is a root of $x^p = 1$, $w \neq 1$). This is now an equation in the domain $D_p = \{a_0 + a_1w + \ldots + a_{p-1}w^{p-1}: a_i \in Z\}$ of so-called *cyclotomic integers*.

Lamé claimed, not unlike Euler, that since the product on the left-hand side of (**) is a p-th power, each factor must be a p-th power. (By multiplication by an appropriate constant, he was able to make the factors relatively prime in pairs.) He then showed that there are integers u, v, w such that $u^p + v^p = w^p$, with $0 < w < z$. Continuing this process *ad infinitum* leads to a contradiction. So Fermat's Last Theorem was proved.

Both Euler's and Lamé's proofs were essentially correct, on the assumption—which they both implicitly made—that the domains under consideration (D and D_p) possess unique factorization.

The Metaphysics: The unique factorization property, which holds for the domain of ordinary integers, also holds for various domains of "complex integers" (e.g., D and D_p).

Of course, this is not always the case. While unique factorization holds in D, and in D_p for $p < 23$, it fails in D_p for all $p \geq 23$. So Euler's proof was essentially correct, while Lamé's failed for all $p \geq 23$. But it was a driving force behind important developments. Mathematicians began to address the questions: For which "integer domains" (such as D and D_p above) does unique factorization hold? What *is* an "integer domain"? When unique factorization fails, can it be restored in some way?

The Mathematics: The study of unique factorization in various domains. This led in the second half of the nineteenth century to the introduction of fundamental algebraic concepts, such as ring, ideal, and field, and to the rise, in the hands of Dedekind and Kronecker, of *algebraic number theory* [19].

Concluding Remarks

Underlying the use of the Principle of Continuity is the tension between rule and context. In the final analysis, context is of course all-important, but the rule took centre-stage in the mathematical breakthroughs I have discussed. Even the cases in which the Principle of Continuity was inapplicable—the cautionary tales, if you will—were often starting points for fruitful developments (cf. Lamé's "proof" of Fermat's Last Theorem).

The interplay between rule and context, between computation and conceptualization, between algorithm and proof, is central in mathematics—both in research and in teaching. Whitehead and Freudenthal give expression to some of these thoughts:

> It is a profoundly erroneous truism, repeated by all copybooks and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in battle—they are strictly limited in number, they require fresh horses, and must only be made at decisive moments [36, pp. 41–42]:
>
> I have observed, not only with other people but also with myself . . . that sources of insight can be clogged by automatisms. One finally masters an activity so perfectly that the question of how and why is not even asked any more, cannot be asked any more, and is not even understood any more as a meaningful and relevant question [13, p. 469].

The Principle of Continuity is of course not a universal law. In particular, there are many important instances in which progress was made by disregarding it, bucking what appeared to be immutable laws. Three examples follow.

(a) Ignoring the commutative law of multiplication (which had been a *sine qua non* for number systems) in attempts to extend the multiplication of complex numbers to triples enabled Hamilton in the 1840s to invent/discover quaternions [16].

(b) Ignoring the law that the whole is greater than any of its parts (one of Euclid's "common notions") overcame a major obstacle in Cantor's introduction of infinite cardinals and ordinals in the 1870s [10].

(c) Ignoring the received wisdom that a function must be given by a formula or a curve (the seventeenth- and eighteenth-century view of functions) enabled the introduction of "pathological" functions (e.g., everywhere continuous and nowhere differentiable functions) and the rise of mathematical analysis [22].

The Principle of Continuity can be thought of as an argument by analogy. I have only scratched the surface of this vast topic. See, for example, Pólya's superb *Mathematics and Plausible Reasoning*, which is addressed to students and teachers [30]. In this article I have considered a rather restricted notion of analogy, in which mathematical arguments, objects, or theories are carried over from given cases to what appear to be like cases (e.g., from positive to negative numbers, real to complex numbers, polynomial to power series, and ordinary integers to "complex integers"). And—most important—in the examples I have given, mathematicians *assumed* that the analogies were valid.

The power of analogy in mathematics often stems from seeing similarities between theories not readily visible to the "naked eye". And, of course, nowadays we would have to *prove* that the analogies held. The following is an important example of analogy (a Principle of Continuity, if you will) in this broader sense:

In the 1850s Riemann introduced the fundamental notion of a Riemann surface to study algebraic functions. But his methods were nonrigorous. Dedekind and Weber, in an important 1882 paper, set themselves the task of "justify[ing] the theory of algebraic functions of a single variable . . . from a simple as well as rigorous and completely general viewpoint" [27, p. 154]. To accomplish this, they carried over to algebraic functions the ideas that Dedekind had introduced in the 1870s for algebraic numbers. This was a singular achievement, pointing to what was to become an important analogy between algebraic number theory and algebraic geometry [21].

See [9, Chapter 4], [26], and [35] for further remarks on analogy in mathematics.

I conclude with a quotation from Michael Atiyah's 1975 Bakerian Lecture on Global Geometry [1, p. 717]:

> Mathematics can, I think, be viewed as the *science of analogy*, and the widespread applicability of mathematics in the natural sciences, which has intrigued all mathematicians of a philosophical bent, arises from the fundamental role which comparisons play in the mental process we refer to as 'understanding'.

REFERENCES

1. M. F. Atiyah, Bakerian lecture, 1975: global geometry, *Amer. Math. Monthly* **111** (2004) 716–723.
2. I. G. Bashmakova, *Diophantus and Diophantine Equations*, Mathematics Association of America, Washington, DC, 1997.
3. G. Birkhoff and M. K. Bennett, Felix Klein and his "Erlanger Programm", in *History and Philosophy of Modern Mathematics*, ed. by W. Aspray and P. Kitcher, University of Minnesota Press, Minneapolis, 1988, pp. 145–176.
4. H. J. M. Bos, C. Kers, F. Oort, and D. W. Raven, Poncelet's closure theorem, *Expositiones Math.* **5** (1987) 289–364.
5. ——, Differentials, higher-order differentials and the derivative in the Leibnizian calculus, *Arch. Hist. Ex. Sc.* **14** (1974) 1–90.
6. U. Bottazzini, *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*, Springer-Verlag, New York, 1986.
7. N. Bourbaki, *Elements of the History of Mathematics*, Springer-Verlag, New York, 1991.
8. E. Brieskorn and H. Knörrer, *Plane Algebraic Curves*, Birkhäuser, Boston, 1986.
9. D. Corfield, *Towards a Philosophy of Real Mathematics*, Cambridge University Press, Cambridge, 2003.
10. J. W. Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, Mass., 1979.
11. C. H. Edwards, *The Historical Development of the Calculus*, Springer-Verlag, New York, 1979.
12. C. G. Fraser, The calculus as algebraic analysis: some observations on mathematical analysis in the 18th century, *Arch. Hist. Ex. Sci.* **39** (1989) 317–335.
13. H. Freudenthal, *The Didactic Phenomenology of Mathematical Structures*, Reidel, New York, 1983.
14. J. V. Grabiner, *The Origins of Cauchy's Rigorous Calculus*, MIT Press, Cambridge, Mass., 1981.
15. H. Grant, Leibniz and the spell of the continuous, *College Math. J.* **25** (1994) 291–294.
16. T. L. Hankins, *Sir William Rowan Hamilton*, Johns Hopkins University Press, Baltimore, 1980.
17. J. Keisler, *Elementary Calculus: An Infinitesimal Approach*, 2nd ed., Prindle, Weber & Schmidt, Boston, 1986.
18. F. Klein, A comparative review of recent researches in geometry, *New York Math. Soc. Bull.* **2** (1893) 215–249.
19. I. Kleiner, From Fermat to Wiles: Fermat's Last Theorem becomes a theorem, *Elem. der Math.* **55** (2000) 19–37.
20. ——, Field theory: from equations to axiomatization, Parts I and II, *Amer. Math. Monthly* **106** (1999) 677–684 and 859–863.
21. ——, From numbers to rings: the early history of ring theory, *Elem. der Math.* **53** (1998) 18–35.
22. ——, Evolution of the function concept: a brief survey, *College Math. J.* **20** (1989) 282–300.
23. E. Knobloch, Leibniz's rigorous foundation of infinitesimal geometry by means of Riemannian sums, *Synthèse* **133** (2002) 59–73.
24. ——, Analogy and the growth of mathematical knowledge, in *The Growth of Mathematical Knowledge*, ed. by E. Grosholz and H. Breger, Kluwer Academic Publishers, Amsterdam, 2000, pp. 295–314.
25. E. Koppelman, The calculus of operations and the rise of abstract algebra, *Arch. Hist. Ex. Sc.* **8** (1971/72) 155–242.
26. M. H. Krieger, Some of what mathematicians do, *Notices Amer. Math. Soc.* **51** (2004) 1226–1230.
27. D. Laugwitz, *Bernhard Riemann, 1826–1866*, Birkhäuser, Boston, 1999.
28. ——, On the historical development of infinitesimal mathematics I, II, *Amer. Math. Monthly* **104** (1997) 447–455, 660–669.
29. P. J. Nahin, *An Imaginary Tale: The Story of $\sqrt{-1}$*, Princeton University Press, Princeton, 1998.
30. G. Pólya, *Mathematics and Plausible Reasoning*, 2 vols., Princeton University Press, Princeton, 1954.
31. H. M. Pycior, George Peacock and the British origins of symbolical algebra, *Hist. Math.* **8** (1981) 23–45.
32. A. Robinson, *Non-Standard Analysis*, North-Holland, Amsterdam, 1966.
33. B. A. Rosenfeld, The analytic principle of continuity, *Amer. Math. Monthly* **112** (2005) 743–748.
34. A. Weil, *Number Theory: An Approach through History*, Birkhäuser, Boston, 1984.
35. ——, De la métaphysique aux mathématiques, *Collected Papers*, Vol. 2, Springer-Verlag, New York, 1980, pp. 408–412. (Translation into English in J. Gray, Open University Course in History of Mathematics, Unit 12, p. 30.)
36. A. N. Whitehead, *An Introduction to Mathematics*, Oxford University Press, Oxford, 1948.

Feel like writing a review for The Mathematical Intelligencer? You are welcome to submit an unsolicited review of a book of your choice; or, if you would welcome being assigned a book to review, please write us, telling us your expertise and your predilections.

Column Editor: **Osmo Pekonen,**
Agora Centre, 40014
University of Jyväskylä, Finland
e-mail: pekonen@mit.jyu.fi

Geometrical Landscapes: The Voyages of Discovery and the Transformation of Mathematical Practice

by Amir R. Alexander

STANFORD, CALIFORNIA, STANFORD UNIVERSITY PRESS, 2002, 336 PP. US $70.00, ISBN: 0804732604

REVIEWED BY KIM WILLIAMS

The history of science—which includes the history of mathematics—is a dodgy business. The historian has to wear two different hats. He must have a firm grasp of the mathematical concepts he is dealing with, and he must be able to properly insert them into their historical context. Physicist David Speiser addressed this duality thus:

> . . . science and history are two radically different endeavors of the human spirit.
>
> The essence of science lies in its property of being systematic since science ultimately always wishes to grasp the laws of nature, which it strives to uncover and to formulate in the simplest and most transparent form. But human history, and thus also the history of science, is the complete opposite of this: it is totally unsystematic, always complex and never simple nor transparent. So, for writing the history of science two different, indeed totally opposite, endeavors must simultaneously be at work in the same man. Thus, from the same man two almost irreconcilable gifts are requested—gifts from his intellect as well as from his heart. This confrontation, one might say 'clash', of the endeavor to systematize and to extract the universally valid from the documents which the historian finds before him, with the aim to determine the conditions under which this, always unique, discovery was made, under very special circumstances and by one distinct individual different from all others, and then to interpret its significance for the development of science, is the character of the history of science. It is its very essence, even its unique prerogative and also its characteristic charm [spe, pp. 39–40].

In other words, a special kind of mental aperture is required to undertake and accomplish the necessary inquiry into science and its historical context. But the gifts of the historian of science are often underappreciated. In the past, an historian of mathematics risked being considered something of a pariah in the eyes of his mathematical colleagues: if he were really good at mathematics, he wouldn't do history, would he? Many like to think that this mentality is obsolete today, but it lurks in the shadows still.

Interdisciplinary studies, for all the lip service given to them, are no less dodgy. It takes an extraordinary mental aperture to look outside the limits of one's own discipline to find inspiration or clarity in another. The simple question of professional language or vocabulary can be one barrier, perhaps particularly in the case of mathematics. But territorial protectionism can also be a barrier to understanding the viewpoints of other disciplines. The scholars who attempt to cross disciplinary borders are often harshly criticised by practitioners of the other discipline as unprepared to address the subject, the implication being that we should stick to what we know.

So Amir Alexander is to be commended for daring to undertake *Geometrical Landscapes: The Voyages of Discovery and the Transformation of Mathematical Practice*, a study that is both historical and interdisciplinary. But the risks of such an undertaking are not to be underrated. This brings to mind Samuel Johnson's quip about women preachers: "Sir, a woman preaching is like a dog's walking on his hinder legs. It is not done well; but you are surprised to find it done at all" [bos, p. 327]. The "clash" mentioned by David Speiser is evident here: Alexander delves with his heart into the age of exploration, but he forces his intellect to rationalize and follow.

Alexander relegates the explanation of why he approached his subject in the way he did to an appendix at the end of the book, but he would have done well to bring it to the fore, because his theory about how to connect effectively the history of mathematics to cultural history is the guiding force of the book. He has certainly given the issue of how mathematical history should be approached careful consideration, and he has constructed his book to reflect his conclusions:

> . . . [W]hile it is always difficult to point to the presence of social ideas and practices within mathematics, mathematical work does, I argue, contain a narrative. Once this narrative is identified, it can be related to other, nonmathematical cultural tales that are prevalent within the mathematician's social circles. A clear connection between the "mathematical" and the "external" stories would place a mathematical work firmly within its historical setting [ale, p. 209].

The argument that Alexander presents is a case study of his method.

The book addresses the relationships between the geographical explorations that permeated the age from the mid-1500s to the mid-1600s and mathematical explorations of the same time. But Alexander's brand of both historicism and interdisciplinarianism are not standard stock. One might expect, for instance, that he would discuss how mathematics was put into the service of exploration, such as the development of mathematical instruments to aid navigation. He does discuss this briefly, but it isn't the thrust of his study. Rather he looks at how the spirit of geographical exploration influenced the spirit of mathematical exploration and injected it with the same point of view, the same vocabularly to describe the discoveries. Essentially he describes the macrocosm of exploration in general, and then zooms in to view the microcosm of a particular mathematical exploration, that of Thomas Hariot and the development of infinitesimals as a means of exploring the hidden inner structures of geometrical figures.

Hariot is an interesting, if minor, figure, a true Renaissance man in the breadth of his interests. An acclaimed mathematician in his time, though largely unheard of today, he took part in the 1585 expedition of Sir Walter Raleigh to Roanoke, Virginia. Out of this came his one publication, *A briefe and true report of the new found land of Virginia*, published in 1588. Practical matters influenced the nature of some of his work: in addition to celestial navigation, for instance, he studied the piling and trajectories of bullets. He explored techniques of producing terrestrial globes, worked on map projections with Mercator, and corresponded with Kepler about the refraction of light and the dispersion of light in the rainbow. He developed his own telescopes and discovered sunspots at almost the same time as Galileo.

Alexander makes the argument that Hariot's strictly mathematical studies were influenced by his experiences in exploration:

> . . . [I]n Hariot's mathematics, just as much as in his physical investigations, the structure . . . revealed is by no means random. It is, rather, shaped by the standard narrative of geographical exploration. The confusing array of difficulties to be overcome, the hidden mysteries of the interior, and the eventual detection of clear passages through the obstacles are clearly in evidence in Hariot's mathematical studies. The structure of the mathematical continuum, much like the optical medium, thus came to resemble the geography of foreign lands [ale, p. 131].

The similarity of language is used as a key to demonstrating that, for Hariot, exploring a vast new physical world was the same as exploring the inner reaches of a purely mental mathematical world:

> Hariot . . . utilize[s] the basic elements of the tale of exploration in hinting at the solution to the paradox: the rhumb is presented as a confusing and imponderable "labyrinth" withholding its secret and defying mathematical analysis. Hariot then breaks through the difficulties. . . . Much like the distant shores of America, or the problematic medium of refraction, the confusing, paradoxically smooth curve is mastered by carving it up into discrete fundamental parts [ale, p. 154].

Alexander wants to convince us that Hariot's studies of infinitesimals couldn't have taken place at any other time in history, that the spirit of the age of exploration was a determining force in Hariot's aims and points of view as he approached the questions of the equiangular spiral and the continuum, and the nature of geometric figures themselves.

While I appreciate Alexander's goals, and the care with which he has prepared the argument, I don't find the end result very convincing: it feels forced. There are several reasons for this. First of all, I am unconvinced that the culture of exploration was as compelling as Alexander argues. I don't doubt that it was the spirit of the age; I doubt that it provided as irresistible a force as Alexander claims. A comparison might be made with the culture of Catholicism that ruled the age of Galileo, during which the threat of accusation of heresy carried such force that it certainly conditioned studies in science. Today, on the other hand, we live in the age of capitalism and the culture of success, yet it is conceivable that one could choose to live outside that culture (refusing grants and tenures) and still produce excellent mathematics. The question here is whether, given the culture of exploration, one could have lived in that age, ignored its spirit, and yet produced great mathematics. I believe the answer is yes.

The other great weakness I find is that Hariot himself is a minor and idiosyncratic figure, and does not seem to me to support the weight of the argument that Alexander places on his shoulders.

On the other hand, I felt that the single section of chapter six, titled "The Hidden Troves of Mathematics: Indivisibles are the Mathematics of Exploration", in which the work of Hariot is placed in perspective against the works of Galileo, Cavalieri, and Torricelli on indivisibles and Wallis's adaptation of indivisible into an algebraic "arithmetic of infinites", did more to place Hariot in his proper context than the narrative of exploration. Yes, there exists a linguistic similarity between Hariot's mathematics and exploration (the simile of labyrinths); yes, there is an affinity of approach (the need to see and describe clearly the shape of things); but these affinities are less compelling (to this reviewer) than the relationships between Hariot's mathematical ideas and those of men who were working on similar ideas. It seems to me simply that Hariot's mathematics are much more meaningful against a mathematical background than against a cultural background.

Alexander's research on Hariot, which represents an exploration in itself, as much of Hariot's mathematical work is unpublished and had to be reconstructed from manuscripts and letters, is admirable. Similarly admirable is his venturing outside mathematics—like a true explorer—to construct the narrative of geographical exploration. I am unsure how successful the final result is. I found it difficult to shift between the narrative of geographical exploration and the narrative of the exploration of infinitesimals. I am reminded of how Jonathan Swift was inspired by the deformed images of microscope and magnifying glass to create his mythical nations of Lilliput and Brobdingnag: each realm merits exploration in itself; going back and forth between the two to find overarching unifying principles might overtax the armchair explorer.

REFERENCES

[ale] Alexander, Amir R. *Geometrical Landscapes: The Voyages of Discovery and the Transformation of Mathematical Practice.* Stanford, CA, Stanford University Press.

[bos] Boswell, James. 1980. *Life of Johnson.* Oxford, Oxford University Press.

[spe] Speiser, David. 2003. The Importance of Concepts for Science. *Meccanica* **38**, 483–492.

The Equations: Icons of Knowledge

by Sander Bais

CAMBRIDGE, MASSACHUSETTS, HARVARD UNIVERSITY PRESS, 2005
96 PP. $US 18.95. ISBN 0-674-01967-9
HARDCOVER.

REVIEWED BY OSMO PEKONEN

The word 'icon' has many meanings. One of them, which has been gaining frequency in use, is defined by Oxford English Dictionary (2001) as "a person or thing regarded as a representative symbol, esp. of a culture or movement; a person, institution, etc., considered worthy of admiration or respect". Certainly, Einstein is an icon of physics, and even of science at large, and so is his equation $E = mc^2$, undoubtedly the most famous equation in the world. This one-liner is typically portrayed in many T-shirt designs, together with Einstein's straggly hairstyle and a looming nuclear mushroom cloud in the background.

Sander Bais, a Dutch physicist at the University of Amsterdam, has compiled in a small book 17 fundamental equations, or systems of equations, of mathematical physics. The beautiful layout, designed by Gijs Matthijs Ontwerpers, represents the chosen formulas as color plates of white on red. Each equation is accompanied with a non-technical historical summary. The following equations are included: (1) the logistic equation, or Verhulst equation, (2) Newton's laws, (3) the Lorentz force, (4) the continuity equation, (5) the Maxwell equations, (6) the wave equations, (7) the Korteweg–De Vries equation, (8) the three laws of Thermodynamics, (9) the Boltzmann equation, (10) the Navier-Stokes equations, (11) the kinematics equations of Special Relativity, (12) the Einstein field equations, (13) the Schrödinger equation, (14) the Dirac equation, (15) the Lagrangian of Quantum Chromodynamics, (16) the Lagrangian of the Glashow–Weinberg–Salam model, and (17) the superstring action principle.

No one can doubt the importance of these equations, yet how many of them have really acquired standing as "icons of knowledge" in popular culture? Newton's and Maxwell's laws are certainly famous enough, and so are the equations of Einstein and Schrödinger. You occasionally see them on posters, mugs, and T-shirts, and on the walls of scientific institutions. The Korteweg–De Vries equation (KdV, for short) is a rare marvel as it contains a *third*-order derivative: $u_t + u_{xxx} + 6uu_x = 0$. It might well be a cultural icon, at least in Bais's home country, the Netherlands. But all the rest of the formulas are either too little known, or too messy, to gain actual iconic status. The laws of thermodynamics are usually not expressed as equations at all, but rather in words. For his 17th entry, Bais cheats a bit. No one really knows whether Superstring Theory will become a fundamental theory of Nature and what its mathematical structure will ultimately look like. (Bais admits this in his commentary.)

Which equations of pure mathematics, if any, are so famous that they have become icons of science? Surely, the Pythagorean Theorem which is expressed by the equation $a^2 + b^2 = c^2$. No less famous is its generalization $a^n + b^n = c^n$, which has to do with Fermat's Last Theorem. The somewhat artificial statement $e^{i\pi} + 1 = 0$, reputed to be "the most beautiful formula of mathematics", must be included. But beyond these, one is hard pressed to claim that any equation of mathematics has gained the status of a cultural icon. For instance, the Riemann Hypothesis is famous enough but cannot be summarized as a single easily memorable formula. The Continuum Hypothesis, on the other hand, can be written as a simple-looking formula, but few people will grasp its meaning without lengthy explanations. The Fibonacci sequence 1, 1, 2, 3, 5, 8, 13 . . . often occurs in the environmental art of Marzio Merz and has gained iconic status thanks to him, but can you call it an equation?

Among the more famous physical equations, Bais misses the equation of entropy $S = k \log W$, which has been carved on Boltzmann's tombstone in the Central Cemetery of Vienna. Boltzmann rests in good company there, together with Beethoven, Brahms, Mozart, Schubert, and four composers named Strauss. In Westminster Abbey, the Dirac equation is said to be the only mathematical equation present among the commemorative plaques of great Britons.

Other famous equations carved in stone include the quaternion algebra $i^2 = j^2 = k^2 = ijk = -1$ inscribed by Hamilton on the 16th of October 1843 on a stone of Brougham Bridge, not far from Dublin. The original inscription is gone, yet a commemorative plaque carrying the same formula was unveiled in 1958 by Eamon De Valera, then president of Ireland. The quaternion equations have appeared on Irish stamps twice (in 1983 and in 2005), so they have certainly acquired iconic status, at least in Ireland. Other equations that have appeared on stamps and on tombstones have been discussed, respectively, in the Stamp Corner and Mathematical Tourist sections of this magazine over the years.

Sander Bais's book is a plea for inserting more equations into public places, starting from popular books of science. "There is a fashionable dogma about popularizing science, which imposes a veto on the use of equations in any popular exposé of science. Some people hate equations, while others love them. The veto is like asking somebody to explain art without showing pictures. In this book, we override the veto and basically turn it around; for a change, the equations themselves will be the focus of attention."

Bais's book is also available in Dutch [1] and in German [2].

REFERENCES

[1] S. Bais, *De natuurwetten. Iconen van onze kennis*, Amsterdam University Press, 2005.

[2] S. Bais, *Die Gleichungen der Physik. Meilensteine des Wissens*, Birkhäuser Verlag, 2005.

Agora Centre
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
e-mail: pekonen@mit.jyu.fi

Collected Works, Volume 6

by Michael Atiyah

OXFORD, CLARENDON PRESS, 2004
xxv + 1030 PP. £100.00. ISBN 0-19-853099-4
THE FULL 6-VOLUME SET OF *COLLECTED WORKS*, £395.00. ISBN 0-19-852094-8

REVIEWED BY OSMO PEKONEN

There is nothing more exasperating than a set of collected works with one missing volume. Therefore, librarians everywhere: please notice that a sixth volume of Sir Michael Atiyah's Collected Works has been separately published and should be added to your collection.

The original Collected Works appeared in 1988 as a five-volume set: Volume 1: Early Papers; General Papers, Volume 2: K-Theory, Volumes 3 and 4: Index Theory, Volume 5: Gauge Theories. These volumes already contained a total of 3206 pages. At that point, a lesser scientist might have been content and thrown in the towel. Not so with Sir Michael. The sixth, 1030-page, volume reaps a rich second harvest of mathematical research still at its best. Atiyah's books on magnetic monopoles (with Nigel Hitchin) and on knot theory are reprinted. Research papers, e.g., on equivariant Euler characteristics (with Graeme Segal), on topological Lagrangians (with Lisa Jeffrey), on skyrmions (with Nick Manton), on configurations of points on a sphere (with Paul Sutcliffe), and on a variant of K-theory (with Michael Hopkins) are included.

A personal favorite of mine is a beautiful paper of 1987 where, in one formula, a certain relationship arises between the topological η-invariant of Atiyah, Patodi, and Singer, on the one hand, and the number-theoretic Dedekind η-function, on the other. This is an exhilarating coincidence, indeed!

Another paper ponders the mystery of the Neolithic carved stone balls discovered in Scotland which provide perfect models of the Platonic solids one thousand years before Plato or Pythagoras (Marshall 1977).

During the period covered in this volume, Sir Michael occupied important positions in the administration of science: He was Master of Trinity College, Cambridge, in 1990–97, Director of the Isaac Newton Institute for Mathematical Sciences, Cambridge, in 1990–96, President of the Royal Society in 1990–95, and Chancellor of Leicester University in 1995–2005, among many other duties. Such positions generate the burden and pleasure of pronouncing laudatory and commemorative talks, many of which have been included in the Volume 6. We have got, for instance, appreciations of the work of Atiyah's contemporaries Raoul Bott, Simon Donaldson, Friedrich Hirzebruch, Kunihiko Kodaira, Roger Penrose, John Todd, and Edward Witten.

"A Personal History" summarizes the main events of Atiyah's mathematical career but tells disappointingly little about the man's human side. Atiyah's contributions to the "theoretical mathematics" debate launched by Jaffe and Quinn in 1993 are reprinted, and so are many other philosophical reflections on the past and future of mathematics, with special emphasis on the interaction of geometry and physics. Atiyah is a careful reader of the classics; he is often able to detect the seeds of the future of mathematics in its past. Many of his comments have proved insightful if taken as predictions of the development of mathematical physics. I don't find Atiyah particularly profound as a philosopher of science, though, judging by the technical standards of that profession. His attitude is a prosaic and pragmatic one, typical of a creative mind more keen on conceiving new results than waffling for too long on their metaphysical or other hidden meanings. The reader may want to compare, for instance, Atiyah's review of Jean-Pierre Changeux and Alain Connes's philosophical book *Conversations of Mind, Matter and Mathematics* with that by Jean Petitot published in the Fall 2005 issue of this magazine. Atiyah rejects Platonic ideal worlds dear to Connes as too difficult to grasp and more modestly locates mathematical reality in the "collective consciousness of mankind." He seems to agree more with Changeux, the neurologist. "The brain evolved in order to deal with the physical world, so it should not be too surprising that it has developed a language, mathematics, that is well suited for the purpose." I am afraid, however, that the mystery of the "unreasonable effectiveness" of mathematics in the natural sciences is not dispelled this easily.

In 2004, Atiyah was awarded the Abel Prize, together with his most important American collaborator I. M. Singer. Pictures of Atiyah and Singer posing with the King and Queen of Norway are included, and so is Atiyah's prize speech delivered in Oslo on 25 May, 2004. In his speech, Atiyah cited Bott, Hirzebruch, Hitchin, and Segal among his most eminent collaborators, paying special credit also to Witten.

Among the many highlights of the volume are two astonishing papers of 2001 that Atiyah wrote on M-theory together with the physicists Maldacena, Vafa, and Witten. In these papers, the exceptional Lie groups play a crucial role. Amusingly, in an interview with *The Mathematical Intelligencer* (Winter 1984 issue), Sir Michael had declared:

> For example, the classification of Lie groups is a bit peculiar. You have this list of groups, both classical and exceptional. But for most practical purposes, you just use the classical groups. The exceptional groups are just there to show you that the theory is a bit bigger; it is pretty rare that they ever turn up.

This must be a statement he has since repented many times. Indeed, M-theory—supposed to become the much-acclaimed Theory of Everything—hinges upon one exceptional Lie group, E_8. There may be a lesson in this: Never say "never" in mathematics. This brings to mind a scene from the James Bond classic *Never Say Never Again*. The superagent in Her Majesty's Secret Service has retired, but M wants to call him into duty again. He tosses M's harbinger into a swimming pool yelling:

–M sent you?

–Only to plead for your return, Sir. M says that without you in the service, he fears for the security of the civilized world.

–Never again.

–Never?

REFERENCES

Marshall, D.N., Carved stone balls, *Proceedings of the Society of Antiquaries of Scotland* 108 (1977), pp. 40–72.

Agora Centre
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
e-mail: pekonen@mit.jyu.fi

Cinq siècles de mathématiques en France

by Marcel Berger

PARIS, ASSOCIATION POUR LA DIFFUSION DE LA PENSÉE FRANÇAISE, 2005
288 PP. 17, ISBN 2-914935-38-2

REVIEWED BY OSMO PEKONEN

Marcel Berger, the former director of Institut des Hautes Études Scientifiques (IHES) of Buressur-Yvette, France, is an eminent differential-geometer who also has a keen interest in the history of mathematics. The book under review was commissioned from him by the French Ministry of Foreign Affairs, which distributes free copies of it through French embassies and consulates throughout the world. Therefore: *Aux livres, citoyens!* Rush to collect your personal copy of this precious book.

Berger's book is an astonishing feat of popularization: in fewer than 300 pages, he manages to cover 500 years of history of French mathematics, taking into account all the major figures up till today and covering essentially all the fields. He illustrates the most complex theories with judiciously chosen examples and figures, occasionally spending a formula, as well. While some of the more abstruse passages of the book will necessarily remain beyond the scope of the non-mathematician reader, anyone interested in French culture will enjoy the numerous historical anecdotes and links with art and literature which Berger has incorporated. The book achieves the often-sought *unité de doctrine* between Art and Science. The author explains the essence of mathematical research by quoting the following lines of Paul Valéry:

> Patience, patience,
> Patience dans l'azur!
> Chaque atome de silence
> Est la chance d'un fruit mûr!

Berger highlights three glorious periods in the history of French mathematics: (1) the years 1550–1650, characterized by Viète, Fermat, Descartes, Desargues, and Pascal; (2) the revolutionary period, whose great names include the "three L's" Laplace, Lagrange, and Legendre, together with d'Alembert, Condorcet, Monge, and Lazare Carnot; (3) the Bourbaki period from 1935 on. These epochs of *grandeur* have alternated with long stages of eclipse of France in the mathematical firmament. In the general history of France, the 17th century is often referred to as the *Grand siècle*. However, in the history of mathematics, the latter half of the 17th century was not so great for France because calculus was invented elsewhere, by Leibniz and Newton. Another long period of French decline was 1840–1935 when Germany, England, and other countries dominated the scene; there were brilliant individuals but not a school in France.

Some patterns in the lives of famous French mathematicians have endured through the centuries. Many have been involved with politics. Laplace and Monge were nominated ministers by Napoleon, and Fourier served for a time as the governor of Lower Egypt. Henri Poincaré was the cousin of Raymond Poincaré, the President of the Republic. Paul Painlevé served as Minister of War during WWI, and served briefly as Prime Minister, as well.

Mathematics and war have been intimately connected in France for five centuries. When Viète, in the service of Henri IV in a war against Spain, used his mathematical knowledge to crack the secret codes of the Spanish army, the King of Spain, convinced that the French had acquired some Satanic powers, complained to the Pope. Descartes and Desargues, it seems, joined forces in the siege of La Rochelle in 1627–28. The École Polytechnique, founded in 1794, has produced legions of mathematician-officers: Cauchy, Chasles, Liouville, Poincaré, Poisson, and Poncelet, to name just a few. Poncelet was made prisoner of war in Napoleon's 1812 Russian campaign; he developed the essentials of projective geometry during his two-year captivity in Saratov on the Volga. André Weil, a reserve officer, in his turn, was held captive in Finland in 1939. Berger cannot resist adding, for the sake of a good anecdote, that Weil surely thought out his famous conjectures of 1949 while in jail in Finland ten years earlier. However, Weil's captivity in Finland only lasted a fortnight (from 30 November till 12 December 1939, to be precise [1]), so I am skeptical about this suggestion.

In the author's estimate, André Weil and his philosopher sister Simone Weil were "among the most extraordinary pairs of the entire history of humankind". Simone Weil has got a street named after her in the 13th arrondissement of Paris whereas André Weil still awaits a similar honor. André Weil has often been described as "the greatest mathematician of the 20th century". In this book, only Jean Leray competes for this title (Alexander Grothendieck is called a "giant"). Henri Cartan, Jacques-Louis Lions, and Laurent Schwartz are cited among the "great teachers" who renovated the mathematical thinking of a generation.

Berger has many good stories to tell about the French Fields medalists. After all, he brought some of them to IHES, a very special place whose atmosphere has something of the seraglio. People living there have interesting hobbies, like tightrope walking in the case of Mikhael Gromov. Berger applauds Grothendieck, Laurent Lafforgue, Jean-Pierre Serre, and others who have kept on publishing in French, thus upholding the language of Molière as an important scientific language, at least in the field of mathematics.

The author grudgingly admits that the United States is the leading power of mathematics in today's world. He hastens to add, however, that among mathematical cities Paris arguably still ranks as number one. Indeed, nowhere else is there such a concentration of great mathematicians living and working in such a vibrant community.

Rarely has a country promoted itself abroad by praising the achievements of its mathematicians. The propaganda aim of Berger's book is to attract more talented overseas students to France. For recruiting purposes, it contains a list of useful addresses and websites.

Paul Appell once said: "France produces mathematicians like an apple-tree produces apples." My only objection to Berger's book is that he could have doubled the glorious history of French mathematics by another 500 years with the inclusion of Gerbert of Aurillac (ca 945–1003), also known as Pope Sylvester II [2].

REFERENCES

[1] Osmo Pekonen, L'affaire Weil à Helsinki en 1939, *Gazette des mathématiciens* **52** (avril 1992), pp. 13–20.

[2] Osmo Pekonen, Gerbert of Aurillac: Mathematician and Pope, *The Mathematical Intelligencer* **22:**4 (2000), pp. 67–70.

Agora Centre
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
e-mail: pekonen@mit.jyu.fi

The Equation that Couldn't Be Solved: How Mathematical Genius Discovered the Language of Symmetry

by Mario Livio

NEW YORK, SIMON & SCHUSTER, 2005, 353 PP., $26.95 ISBN-13: 978-0-7432-5820-3

REVIEWED BY ERIC GRUNWALD

Want to write a book? Here are some ideas for you. You could write about the non-solubility of the quintic equation by radicals. Or you could trace the development of the modern idea of symmetry through to its current pre-eminence in physics. Or perhaps you could write a biographical account of some of the key architects of modern mathematics. Or you could even write a more speculative book about the influence of symmetry on our notion of beauty and on psychology in general.

The title of Mario Livio's book suggests he has plumped for the first option. His subtitle implies the second. In fact, his book is a rather meandering mishmash of all four. Does this matter? Why not write a book weaving these four related threads together? Well, maybe, but the trouble is that the target readership of the four books is so different.

The first book, a reasonably non-technical account of Galois theory, would be aimed at people with a specific interest in mathematics who want to get a feeling for some important ideas. When I was thirteen or so, I read a delightful book by W. W. Sawyer called *A Concrete Approach to Abstract Algebra*, which showed the impossibility of trisecting the angle with straight edge and compass. It contained the deepest mathematical ideas I had come across at the time, explained clearly and simply without excessive technical detail and (from memory) without too much fudging of the main points. I got a real thrill from getting to the end and understanding most of it. Livio might have done the same thing for the quintic. But he hasn't. There is a boring introduction to groups, with lots of multiplication tables, and then a three-page section called "Galois's Brilliant Proof," only two of which are devoted to a discussion of the proof, much of the rest being worthy quotes about how terrifically clever it all is. I had been looking forward to a really clear heuristic narrative on the relationship between the solutions of an equation and symmetry groups, and I was disappointed and frustrated.

The second book would be the most interesting of the four, I think, and I did enjoy Livio's chapter "Symmetry Rules," about the importance of symmetry and groups in physics. Livio, an astronomer, seems at his happiest in this area, and he writes well. But, as with the first book, this book requires a readership prepared to roll up its metaphorical sleeves and plunge into more technicalities than Livio allows himself here. I wish Livio had written this book. Perhaps he will.

The third, the biographical book, would have to be aimed at people who haven't read E. T. Bell. Big chunks of Livio's book are devoted to biographies of Abel and Galois, which are pretty much exactly what I had remembered from Bell, whom I read some forty years ago. Livio seems proud of his original research: "I have put a tremendous effort into trying to solve the two-centuries-old mystery of the death of . . . Évariste Galois. I believe that I have come closer to the truth than was ever possible before." He appears to be writing for historians of mathematics rather than for people who have never heard of Galois, whom I suspect to be the more likely readership. As for me, I'm not terribly interested in whether Galois was betrayed by the Duc de Merde or by Mademoiselle Cul de Sac; I want to know how such an intelligent man could have been so dumb as to get himself killed in a duel, and I still don't get it. The confusion over the intended readership is exemplified by the sort of adjectival biographies that pepper the book. We read about "the witty playwright George Bernard Shaw." Is the book aimed at people who haven't heard of Shaw or don't know he was witty? In a similar vein, we are told about "Clive Bell (1881–1964), an art critic and member of the Bloomsbury group (which, by the way, included novelist Virginia Woolf)." There is no extra charge for the parenthetical information. Then in two successive sentences we hear about "the famous Harvard mathematician George David Birkhoff (1884–1944)" and "the French mathematician Henri Poincaré." So Birkhoff is more famous than Poincaré? Or perhaps Poincaré is too famous to be called 'famous'?

As for the fourth book, I think there is scope for an updated version of Weyl's *Symmetry* without the technical bits, together with an account of recent work in psychology. Livio neatly points out that while animals have up-down asymmetry (gravity) and front-back asymmetry (locomotion), "*there is nothing major in the sea, on the ground, or in the air, to distinguish between left and right*" (his italics). But I felt the focus on bilateral symmetry in psychology sat uneasily with the rest of the book, which is about much more sophisticated forms of symmetry. It almost felt like cheating to give it the same name. Anyway, I am sceptical about the Weyl tendency to conflate mathematical ideas of beauty, i.e., symmetry, with the everyday concept of beauty. What's symmetrical about a beautiful landscape, for instance? Or a beautiful love story? And as for the much-publicised work about bilateral symmetry having a key role in sexual attractiveness, I need a lot more convincing. I have heard men discuss every conceivable aspect of a woman, but never her symmetry.

This isn't a bad look at all. It's a bit unfocused, and frustrating in parts, but it is competently done. I am the wrong reader, because I wanted more mathematical ideas and less E. T. Bell. Who is the right reader, then? I think it would be a *smokoi* (School Mathematics Only, Keen On Ideas). So I asked my son, an exemplary smokoi, to have a look at it. He enjoyed it more than I did. He was moved by the biographies of Abel and Galois, whom he knew nothing about, and he liked the parts about physics. He felt (like me) that the psychology belonged in a different book, and commented, "Humans are symmetrical, therefore symmetrical things provoke an emotional reaction in humans. Yawn." And he "got lost with subgroups." (I think I would have done too in his place.) If you need to buy a present for a smokoi, try this book.

Perihelion Ltd.
187 Sheen Lane
London SW14 8LE
England, U.K.
e-mail: ericgrunwald@aol.com

Incompleteness: The Proof and Paradox of Kurt Gödel

by Rebecca Goldstein

NEW YORK AND LONDON, W. W. NORTON & COMPANY, 2005, 296 PP. US $22.95 HARDCOVER, ISBN 0-393-05169-2

REVIEWED BY AMIR ALEXANDER

In the final session of a meeting on metamathematics held in Königsberg in October, 1930, 23-year-old Kurt Gödel reported on the results of his work on mathematical logic. It was possible, he said, that there might be true though unprovable arithmetical propositions; furthermore, he added, he himself had proved that there are such propositions. That was all. In a meeting that featured the leading philosophers and logicians of the day, the report by the unknown young Viennese went largely unnoticed. When Rudolph Carnap and Hans Reichenbach, who attended the meeting, reported on the proceedings in their journal *Erkenntnis*, they did not even mention Gödel or his statement.

And yet, Gödel's announcement far outlived anything else that was said in that meeting three quarters of a century ago. Though unrecognized by those present, Gödel's "Incompleteness theorem" went to the very heart of the meeting's topic, metamathematics, undermining the most basic assumptions that the participants took for granted. At the same time, like a stone dropped in a tranquil lake, the implications of Gödel's theorem kept expanding far beyond metamathematics and its parent field, mathematical logic, into philosophy and our very understanding of the nature of knowledge and reality. In the end, Gödel's Königsberg announcement proved to be the epicenter of an intellectual earthquake that would reshape the landscape of 20th century thought.

In her beautifully written and captivating *Incompleteness*, Rebecca Goldstein succeeds admirably well in explaining why this was so, walking the reader through both the sources and the radical implications of Gödel's theorem.

The fundamental issue addressed by Gödel's theorem, according to Goldstein, is nothing less than the question of truth itself. What is truth, and how can it be known? Broadly but persuasively, Goldstein points to two competing schools of thought on the matter, whose opposition dates to classical antiquity. On the one side are ranged the "Sophists" and their heirs, whose outlook is encapsulated in Protagoras's dictum, "Man is the measure of all things." In this view, truth is never free of its human seekers but is always dependent on the specific conditions and beliefs of individuals or broader society. What is "true" for me may not be "true" for you, and what is "true" in the West may not be so in the East. Truth, for the Sophists and their successors, is never absolute, but always relative to "Man" and the physical, intellectual, and cultural attributes of the human condition.

The Sophists' greatest critic was Plato, who founded the most influential school in the history of philosophy on the opposite tenet—that truth is real, eternal, and independent of human conditions. The transient physical reality we see around us, he famously argued, is but a pale reflection of the dazzling beauty and perfection of eternal unchanging true "forms." The way to approach these pure forms, furthermore, is through careful logical reasoning, of which Geometry provides the best exemplar.

The great divide between the relativist Sophists on the one hand and Platonists on the other, Goldstein argues, has persisted to our own day. The old Sophist outlook now appears in new and sophisticated garb as postmodernism, positivism, and formalism, but it retains the underlying dictum: "Man is the measure of all things." Platonism, while rarely advocated explicitly, has also retained its grip on the modern mind, which has never relinquished its ambition to gain access to pure unadulterated truths. Gödel's theorem, says Goldstein, weighs in precisely on the fundamental question dividing these two streams of Western thought: are there absolute truths independent of human knowledge? Yes, said Gödel, who believed he had settled the ancient philosophical dispute for all time. The Platonists, Gödel said, were right, the "Sophists" and their heirs wrong. Most shockingly, he said he could prove it.

As Goldstein makes clear, Gödel developed his views in an intellectual atmosphere that was distinctly hostile to his Platonist views. While a student at the University of Vienna in the 1920s, Gödel was a regular member of Moritz Schlick's "Vienna Circle," to which much of 20th-century philosophy of science traces its roots. The Circle was an informal but regular gathering of like-minded intellectuals, and its members included leading lights such as Schlick, Otto Neurath, and Rudolph Carnap, as well as unknown students, of which Gödel was one. Most famously, the Vienna Circle was the birthplace of "logical positivism," a philosophy of science Goldstein describes as "a severe theory of meaning that makes liberal use of the term *meaningless*" (p. 44).

For logical positivists, meaningful statements are only those which are based on the verifiable empirical evidence of our senses. Any statements that go beyond this are by definition "metaphysical," and therefore meaningless. The exception to this general rule is mathematics, which, the logical positivists acknowledged, is based on *a-priori* assumptions rather than on empirical evidence. Mathematical statements can therefore be considered meaningful, but only in the sense that they logically and necessarily follow from the assumptions made. Mathematical truths are essentially endless restatements of the initial assumptions, and do not possess any independent reality in and of themselves.

Mathematics, nevertheless, does play a crucial role in logical positivist thinking, for mathematics has the unique capacity to move correctly and systematically from one set of statements to another. Mathematics therefore allows for long chains of deductions, moving from one meaningful statement to the next. It provides, in essence, the syntax for correct logical deduction, making it possible to move from the simplest empirical observations to the higher complex truths of science.

> *If mathematics possesses truths that are unprovable, they are completely independent of human assumptions and deductions.*

All of this, according to Goldstein, places the logical positivists firmly in the "Sophist" camp of philosophy. All possible statements, after all, are dependent on human empirical observations, and any attempt to move beyond this is meaningless. The sciences furthermore, are nothing but sophisticated restatements of these empirical facts, combined and elaborated with the aid of mathematical syntax. In particular, mathematics refers to no independent reality, and, in fact, has no subject matter of its own. It is merely the syntax of correct deduction, which makes possible the systematic elaboration of empirical truths. At the core, for the logical positivists, all meaning was dependent on specific empirical observations, and all knowledge was an elaboration of these primitive building blocks. Any attempt to go beyond these limits to an external independent truth was pointless, and in fact meaningless. Man, indeed, was the measure of all things.

For several years in the late 1920s Gödel regularly attended the meetings of the Vienna Circle, listening intently but apparently saying very little. By this time, according to Goldstein, Gödel was already a committed Platonist who fundamentally disagreed with the Circle's logical positivist views. Gödel believed that mathematics investigated the relationships of pre-existing mathematical objects such as numbers and sets. Far from being a logical syntax, mathematics, for Gödel, discovers the characteristics and true relations between mathematical objects that exist on their own Platonic plane, independent of human knowledge or understanding. Nevertheless, he was so discreet about his Platonist beliefs that his fellow Circle members had no clue that there was an apostate in their midst. Gödel, as Goldstein makes clear, had no interest in entering into the lengthy and inconclusive debate, which would undoubtedly have followed from announcing his views to his friends. His aim was to prove his position logically and conclusively, state it briefly, and so settle the matter for all time. This was precisely what he was attempting to accomplish when he announced his results to a largely apathetic audience in the Königsberg conference of 1930.

As cited by Goldstein from the *Encyclopedia of Philosophy*, Gödel's results can be roughly stated as follows:

> In any formal system adequate for number theory there exists an undecidable formula—that is, a formula that is not provable and whose negation is not provable (p. 23).

As Gödel himself explained in his Königsberg announcement, this means that mathematical results such as Fermat's last theorem and Goldbach's conjecture can be true in all cases, but nevertheless unprovable (though of course it turned out that Fermat's last theorem, at least, was provable after all). For Gödel, this clearly demonstrated that mathematics could not be the empty logical syntax suggested by the logical positivist. If mathematics possesses truths that are unprovable, this means that they are completely independent of human assumptions and logical deductions. Such statements are simply true, in and of themselves, accurately describing relationships between eternal mathematical objects. To Gödel, at least, this meant that Platonism was true, and its critics wrong. Q.E.D.

The first to glimpse the potential implications of Gödel's announcement was John von Neumann, who was present at Königsberg and questioned Gödel for more details on his proof. When he returned to the Institute for Advanced Studies in Princeton he spread the word about Gödel's work. In 1934 this led to Gödel's first visit to the Institute that would become his home six years later. Reflecting on Gödel's theorem, von Neumann arrived at a corollary, which Gödel himself confirmed:

> The consistency of a formal system adequate for number theory cannot be proved within the system (p. 23).

As von Neumann and others were quick to realize, this profoundly undermined the ongoing effort to provide mathematics with secure logical foundations.

Since the 1890s David Hilbert had been promoting his formalist program to establish the unity and consistency of mathematics. "Mathematics," Hilbert wrote, "is a game played according to simple rules with meaningless marks on paper." For a true formalist (and some question whether Hilbert himself was consistently one), mathematics is the empty manipulation of symbols according to pre-specified rules. Mathematics does not study pre-existing objects, and it does not possess any independent truths. It is "true" only in so far as each empty statement correctly follows from another according to the rules of mathematical deduction.

Hilbert's ambition was to prove that this was indeed so—that mathematics did, in fact, conform to his ideal of a consistent logical construct, founded on first principles. In 1899 he published his *Grundlagen der Geometrie* [*Foundations of Geometry*], which proved that geometry was consistent and could be captured by a formal system. There was, however, a catch: the proof was contingent on *arithmetic*'s being similarly consistent and complete. Proving the consistency and completeness of arithmetic was therefore crucial for the formalist program. When in 1900 Hilbert ranked the 23 most important problems in mathematics, the consistency of arithmetic was second on the list.

Goldstein is quick to point to the unmistakable affinities between Hilbert's views and the logical positivists' approach to mathematics. For both, mathematics was meaningless in itself, and lacked any descriptive content. For both, mathematics was an empty system of relationships, whose only truth value lay in its internal consistency. And to both, Gödel's theorem seemed to deal a serious blow. Not only is the logical structure of mathematics incomplete, since there are mathematical truths which completely escape the systematic logical deductions of proof; but the very consistency of mathematics cannot be proved. The second problem on Hilbert's list, in other words, can never be resolved.

The formalist effort to unify mathematics and provide it with secure logical foundations never recovered from the blow dealt it by Gödel. In that sense one can say that Gödel's proof had much the effect that he expected. As a Platonist, he was sure to draw satisfaction from the dismantling of a program which sought to prove mathematics meaningless and devoid of descriptive content. But when it comes to the philosophical implications of his theorem, the story is very different. The irony of Gödel's life, as Goldstein persuasively argues, is that the implications of his theorem were most often construed in precisely the opposite way from what he had intended. "The worst tragedy for a poet," she quotes Jean Cocteau, "is to be admired through being misunderstood" (p. 76). And that, she says, was to be Gödel's fate.

> *The implications of his theorem were most often construed in the opposite way from what he had intended.*

Far from being hailed as the savior of eternal Platonic truths, Gödel's name is most often cited in support of the exact opposite position. Rather than prove the absolute and independent truth of mathematical relations, Gödel's theorems are taken to show the uncertain nature of rationality itself. Since no logical system can be shown to be consistent, or is ever complete, then the whole scientific project of reducing the world to intelligible logical and mathematical formulas is put into doubt. This, Goldstein claims, is broadly the view of Gödel taken by "postmodernists," a group she identifies as possessing extreme relativistic views on the nature of knowledge.

"He [Gödel] is the devil, for math," she cites an unnamed postmodernist as saying. "After Gödel the idea that mathematics was not just a language of God but a language we could decode to understand the universe and understand everything—that just doesn't work any more. It's part of the great postmodern uncertainty we live in" (pp. 24–25). According to the postmodernists, then, Gödel showed that there can be no secure foundations for knowledge and rationality. All knowledge, for them, is ultimately based on the shifting sands of human relations and human society. This is Sophism in spades: for the postmodernists, even more than for the logical positivists, Man is indeed the measure of all things.

Goldstein persuasively shows Gödel's shock and dismay at having his incompleteness theorem associated with such views. Nothing could have been further from his intentions, and nothing more foreign to his own Platonist beliefs. This, for Goldstein, is the true paradox of Gödel's life: having dedicated his life to the incontrovertible mathematical proof of Platonism, he was ultimately hailed as the champion of Platonism's implacable enemy—Sophism. This, she convincingly shows, is also the way Gödel himself viewed his situation in his later years: universally admired, but profoundly misunderstood.

Goldstein's thesis is compellingly argued as well as beautifully written. By placing Gödel's theorem in the context of an ancient philosophical debate, she persuasively demonstrates the centrality of the incompleteness theorem to Western thought; and by pointing out the irony in the manner in which the theorem was received, she artfully depicts the paradox and tragedy of Gödel's life. The boldest part of Goldstein's interpretation, however, is her division of the philosophical tradition into two warring camps, Platonist and Sophist. And while this is undoubtedly arguable in many specific instances, I also find that it does capture a fundamental truth about the Western philosophical tradition.

Nevertheless, like many bold and

clear statements, Goldstein's thesis does show signs of strain when applied to specific historical circumstances. In particular, the lumping together of logical positivism and postmodernism under the common aegis of "Sophism" is difficult to accept. Protagoras's credo, "Man is the measure of all things," is indeed an appropriate dictum for the postmodernist approach, for however vague the "postmodernist" moniker may be, the different strains of thought that come under its umbrella do share a deep suspicion of a single unifying Truth, accessible through universal scientific rationality. In postmodernist thought human history, language, material conditions, and culture may all play a part in the creation of knowledge.

Placing the logical positivists in the same philosophical camp, however, is something of a stretch. For them, knowledge depended on "Man," only in the narrow sense that "observations," the building blocks of knowledge, were conducted by humans and limited by human attributes. Beyond that point, the acquisition of knowledge must follow the strictly impersonal and inhuman rules of logic and mathematics. Any intervention of human factors into this process would render it invalid, and, in fact, "meaningless." All traces of human history, psychology, practices, or culture were ruthlessly weeded out by the logical positivists. Their goal of creating a single unified and impersonal system of knowledge was precisely the postmodernists' nightmare.

It is significant that more recent philosophy of science, influenced by postmodernist thought, emphasizes what human scientists "actually do," over abstract logical systems of knowledge that claim to capture how science "should" be. Invariably, they criticize the logical positivists precisely for being completely devoid of "humanity" and eager to stamp out any human taint to the production of knowledge. From their perspective at least, "Man is the measure of all things" could hardly be further removed from the logical positivist creed.

Furthermore, while Goldstein is undoubtedly correct that many postmodernists would indeed subscribe to Protagoras's creed, that in itself may not imply an implacable opposition to the implications of Gödel's theorem as he understood them. Many postmodernist thinkers (and I'm using the term in the broadest possible sense here), in fact, have drawn back from the extreme position that viewed all knowledge as a facet of human relationships. While certainly no Platonists, they have embraced Realism—the notion that science refers to preexisting external entities that are independent of human will or consciousness. The intractability of a "real" external world allows the postmodernists to retain their emphasis on the socially constructed facets of knowledge, without falling into the trap of extreme and implausible relativism. As realists, they can insist that all knowledge bears the marks of the cultural setting where it was produced, without implying that the content of all knowledge is therefore random—a mere by-product of social power relations. Realism, in short, allows postmodernist philosophers to bring back "Man" into the construction of knowledge, without subordinating all knowledge to human contingencies.

All of which is to say that postmodernists may not have been very far off the mark in their embrace of Gödel's theorems. The fact that mathematical truths can exist independently of mathematical proofs was in itself no more surprising to them than the stubborn materiality of physical objects. Both accorded well with a realist position, which acknowledged the independent reality of external objects. But the blow dealt by Gödel to efforts to reduce all knowledge to a unified logico-mathematical system was of the highest importance for the postmodernists. It provided the conceptual space for the reintroduction of human culture and practices into the production of knowledge. Tragically for Gödel, as Goldstein shows, no conclusion could have been more alien to his own views on the nature of knowledge.

Gödel, reacting to the Vienna Circle's positivism and Hilbert's formalism, set out to establish the independent existence of eternal unchanging truths. The implications of his success, however, went far beyond his intended goals: for along with the positivist position, his theorems also undermined the very possibility of a unified and systematic logical system of knowledge. This was a tragic irony for a man whose faith in rational deduction was so extreme that during his U.S. citizenship hearings he sought to convince the presiding judge that a logical flaw in the Constitution may result in liberty degenerating into tyranny. Ironically, the Platonist Gödel opened a crack in the door separating scientific and mathematical truths from human contingencies. Much to his dismay, his incompleteness theorem ultimately permitted the return of the human world to the process of the production of knowledge.

Department of History
UCLA
Los Angeles, CA 90095-1473
USA
e-mail: amiralex@ucla.edu

The Birth of Model Theory: Löwenheim's Theorem in the Frame of the Theory of Relatives

by Calixto Badesa

PRINCETON, PRINCETON UNIVERSITY PRESS, 2004
XIII + 240 PP. US $52.50. ISBN 0-691-05853-9

REVIEWED BY JEREMY AVIGAD

From ancient times to the beginning of the nineteenth century, mathematics was commonly viewed as the general science of quantity, with two main branches: geometry, which deals with continuous quantities, and arithmetic, which deals with quanities that are discrete. Mathematical logic does not fit neatly into this taxonomy. In 1847, George Boole [1] offered an alternative characterization of the subject in order to make room for this new discipline: mathematics should be understood to include the use of any symbolic calculus "whose laws of combination are known and general, and whose results admit of a consistent interpretation." Depending on the laws chosen, symbols can just as well be used to represent propositions instead of quantities; in that way, one can consider calculations with propositions on par with more familiar arthimetic calculations.

Despite Boole's efforts, logic has always held an uncertain place in the mathematical community. This can partially be attributed to its youth; while number theory and geometry have their origins in antiquity, mathematical logic did not emerge as a recognizable discipline until the latter half of the nineteenth century, and so it lacks the provenance and prestige of its elder companions. Furthermore, the nature of the subject matter serves to set it apart; as objects of mathematical study, formulas and proofs do not have the same character as groups, topological spaces, and measures. Even distinctly mathematical branches of contemporary logic, like model theory and set theory, tend to employ linguistic classifications and measures of complexity that are alien to other mathematical disciplines.

The Birth of Model Theory, by Calixto Badesa, describes a seminal step in the emergence of logic as a mature mathematical discipline. As its title suggests, the focus is on one particular result, now referred to as the Löwenheim-Skolem theorem, as presented in a paper by Löwenheim in 1915. But the book is, more broadly, about the story of the logic community's gradual coming to the modern distinction between syntax and semantics, that is, between systems of symbolic expressions and the meanings that can be assigned to them. A central characteristic of logic is that it deals with strings of symbols, like "2 + 2" and "$\forall x\ \exists y\ (x < y)$," that are supposed to represent idealized mathematical utterances. We take the first string to denote the natural number 4, assuming we interpret "2" and "+" as the natural number, 2, and the operation of addition, respectively; and we take the second string, which represents the assertion that for every x there is a y satisfying $x < y$, to be *true*, assuming, say, the variables x and y range over natural numbers and $<$ denotes the less-than relation. Model theory provides a rigorous account of how syntactic expressions like these denote objects and truth values relative to a given interpretation, or *model*.

The model-theoretic notions of denotation and truth have influenced both the philosophy of language and the philosophy of mathematics. As Harold Hodes colorfully put it, "truth in a model is a model of truth." The tendency to view mathematics as a primarily syntactic activity underlies nominalist accounts of mathematics in scholastic and early modern philosophy, as well as contemporary versions of formalism. These are contrasted with accounts of mathematical practice that focus not on the linguistic practice, but, rather, on what it is that the linguistic terms denote; and contemporary formulations of these positions are typically guided by a model-theoretic understanding.

> *If a sentence is not refutable in a formal system of deduction for first-order logic, then it has a countable model.*

But the simplicity of the model-theoretic viewpoint belies the fact that it arrived relatively late on the scene. Gottlob Frege distinguished between the "sense" of an expression and its "reference" in the late nineteenth century, but he viewed his system of logic as an ideographic representation of the universal laws of correct judgment, with axioms that are simply *true* with respect to *the* domain of logical objects. There was no stepping outside the system; he had no reason to treat either strings of symbols or the structures that interpret them as mathematical objects in their own right. In the latter half of the nineteenth century, algebraic treatments of logic from Boole to Peirce and Schöder *were* in a position to support multiple interpretations of symbolic expressions, but they were not always careful to distinguish the expressions themselves from their interpretations in the algebraic structures. In his landmark *Grundlagen der Geometrie* [*Foundations of Geometry*] of 1899, Hilbert explored various interpretations of geometric axioms, but the axioms were formulated in ordinary mathematical language, and he relied on an intuitive understanding of what it means for an axiom to "hold" under a given interpretation.

In the winter of 1917–1918, Hilbert, with Paul Bernays as his assistant, gave a series of lectures in which the syntactic presentation of a formal system is clearly distinguished from its interpretation in a given domain. (For a discussion of these lectures, see [4].) In his *Habilitationsschrift* of 1918, Bernays showed that the propositional calculus is complete in the modern sense that "[e]very provable formula is a valid formula and *vice-versa*." The question as to whether the usual deductive systems for first-order logic are complete in the same sense is clearly articulated in Hilbert and Ackermann's *Grundzüge der theoretischen Logik* [*Principles of Theoretical Logic*] of 1928, which is based on Hilbert's 1917–1918 lectures.

The distinction between the syntax and semantics is not limited to logic. For example, consider $x^2 + 2x + \sin x + \cos(x + \pi/2)$ and $2x + x^2$. Are these both polynomials? Are they the same polynomial? In contemporary practice, we are apt to dissolve the confusion by distinguishing between polynomial *expressions* and the polynomial *functions* they denote. Although "$x^2 + 2x + \sin x + \cos(x + \pi/2)$" is not a polynomial expression, it denotes the same polynomial function of x as the expression "$2x + x^2$." We tend to forget that this is a relatively modern way of thinking, and that the distinction is often muddied in nineteenth-century work on logic.

This discussion provides some context of Löwenheim's 1915 paper, "Über Möglichkeiten im Relativkalkül" ["On Possibilities in the Calculus of Relatives"]. (A translation, as well as translations of the papers by Skolem and Gödel discussed below, can be found in [5].) The paper's second and most important theorem is stated as follows:

> If the domain is at least denumerably infinite, it is no longer the case that a first-order fleeing equation is satisfied for arbitrary values of the relative coefficients.

In modern terms, a "first-order fleeing equation" is a first-order sentence that is true in every finite model but not true in every model. Löwenheim's theorem

Gödel's and Skolem's proofs. Gödel first proved the completeness theorem in the form "if a sentence φ is not refutable, then it has a model." In doing so, he considered languages without function symbols or the equality symbol. The absence of function symbols is not a serious restriction, since functions can be interpreted in terms of relations. He then extended the result to infinite sets of sentences, by proving what is now known as the "compactness" theorem; and to first-order logic *with* equality, essentially using the modern method of taking a quotient structure.

I know of no contemporary textbook that presents Gödel's proof, which is a shame, since the argument is elegant and direct. The central idea is that any first-order sentence φ is equivalent to a sentence of the form

$$\exists R_1, \ldots, R_l \, \forall x_1, \ldots, x_m \, \exists y_1, \ldots, y_n \, \theta, \quad (1)$$

where θ is a quantifier-free formula, and $R_1, \ldots, R_l$ range over *relations* on the first-order universe. Saying that θ is quantifier-free means that it is a Boolean expression involving the variables $x_1, \ldots x_m, y_1, \ldots, y_n$, the relation symbols $R_1, \ldots, R_l$, and symbols in the language of φ. This normal form was used by Skolem in 1920. Since first-order logic does not allow quantification over relation symbols, (1) is not first-order. But fixing $R_1, \ldots, R_l$ and letting φ' denote

$$\forall x_1, \ldots, x_m \, \exists y_1, \ldots, y_n \, \theta(x_1, \ldots, x_m, y_1, \ldots, y_n), \quad (2)$$

Gödel showed:

- $\varphi' \rightarrow \varphi$ is provable in first-order logic. Thus if φ' has a countable model, so does φ.
- If φ' is refutable, then so is φ.

This reduces the task of proving the completeness theorem to proving it for sentences of the form (2).

To carry out this last step, Gödel first expanded the language by adding a countable sequence of new constant symbols. Let $c_0, c_1, c_2, \ldots$ enumerate both the new constant symbols and the original constant symbols in φ'; in fact, it is these (syntactic!) objects that will constitute the domain of the desired model. In order to satisfy (2), Gödel enumerated all m-tuples of these constants, $\vec{c}_1, \vec{c}_2, \vec{c}_3, \ldots$. He then recursively defined a sequence of formulas

$$\psi_0 = \theta(\vec{c}_1, c_{k_0+1}, \ldots, c_{k_0+n})$$
$$\psi_{i+1} = \psi_i \wedge \theta(\vec{c}_{i+1}, c_{k_{i+1}+1}, \ldots, c_{k_{i+1}+n}),$$

where each k_i is chosen large enough so that the constants $c_{k_i+1}, \ldots, c_{k_i+n}$ are "fresh," i.e., have not appeared in $\psi_0, \ldots, \psi_{i-1}$.

Now, each ψ_i is a Boolean combination of atomic formulas, that is, formulas of the form $R(c_{i_0}, \ldots, c_{i_{l-1}})$, where R is a relation symbol and $i_0, \ldots, i_{l-1}$ is any sequence of indices. Gödel showed that if any of the ψ_i are refutable, then, in fact, so is φ'. On the other hand, if this is not the case, then by the completeness theorem for propositional logic there is a way of assigning truth values to each atomic component in any ψ_i so that the formula comes out true. These satisfying truth assignments can be arranged into a tree, where the i-th level of the tree has all the truth assignments that satisfy ψ_i, and the descendants of a truth assignment are simply those in the tree that extend it. Each level of the tree is finite, and, by hypothesis, there is at least one assignment at each level. By König's lemma, there is an infinite path through this tree: recursively, at each level i, choose any assignment extending the previous with infinitely many descendants. Now consider the model whose domain consists of the constants, where each atomic formula $R(c_{i_0}, \ldots, c_{i_{l-1}})$ is interpreted as true if and only if this atomic formula is interpreted as true at the first level where it appears. This provides a model of $\psi_0, \psi_1, \psi_2, \ldots$, and hence φ', since for each possible instantiation $\vec{c}_i$ of the universal quantifiers in (2) we have chosen constants $c_{k_i+1}, \ldots, c_{k_i+n}$ to witness the existential quantifiers.

This proves the completeness theorem. The weak version of the Löwenheim-Skolem theorem is a consequence: if φ has a model, then it is not refutable, and so φ has a countable model. In fact, if we replace "not refutable" by "has a model" in the second step of Gödel's argument, we end up with, more or less, Skolem's 1922 proof. Skolem's 1920 proof differs in that it uses the axiom of choice but establishes a stronger result: if $\mathcal{M}$ is any model satisfying φ, there is a countable *submodel* of $\mathcal{M}$ satisfying φ. This theorem makes no mention of provability, and it turns out that in this case there is a more convenient choice of normal form. Consider a first-order sentence of the form $\forall x \, \exists y \, \theta(x,y)$. If this is true in a model, then for every element a in the domain, there is an element b such that θ holds of a and b in the interpretation. If f is a function which for every a chooses such a b, we have $\forall x \, \theta(x, f(x))$. In other words, using the axiom of choice, we can see that $\forall x \, \exists y \, \theta(x,y)$ is true in a model if and only if there is a function, denoted by f, such that $\forall x \, \theta(x, f(x))$ is true in the same model. By iterating this move, one can show that every first-order formula ψ is equivalent, in an appropriate sense, to a formula of the form $\exists f_1, \ldots, f_l \, \forall x_1, \ldots, x_m \, \theta$, where θ is quantifier-free. Now suppose φ is true in some model $\mathcal{M}$. Choose interpretations of $f_1, \ldots, f_l$ making $\forall x_1, \ldots, x_m \, \theta$ true. Starting with any element a of the domain of $\mathcal{M}$, consider the submodel of $\mathcal{M}$ generated by a and the functions $f_1, \ldots, f_l$. This is a countable submodel of $\mathcal{M}$ satisfying $\forall x_1, \ldots, x_m \, \theta$, and hence φ, as required.

asserts that such a sentence can be falsified in a model whose elements are drawn from a countably infinite domain. Since a sentence is true in a model if and only if its negation is false, we can restate Löwenheim's theorem in its modern form: if a sentence has a model, it has a countable model (that is, one whose domain is finite or countably infinite). The theorem thus expresses an important relationship between a syntactic object—a sentence—and the class of possible models. In the context of the logic of the time, even stating such a theorem was novel, and Badesa is justified in marking this as the birth of model theory.

The logician Thoralf Skolem presented papers in 1920 and 1922 that clarify and strengthen Löwenheim's theorem. In particular, he showed that the theorem holds not just for single sentences but also for any countably infinite set of sentences. Most importantly, he made clear that there are two ways of stating the theorem. It is, in fact, the case that if a sentence has an infinite model $\mathcal{M}$, then the countable model in question can always be chosen as a *submodel* $\mathcal{M}'$ of $\mathcal{M}$—that is, a restriction of the functions and relations of $\mathcal{M}$ to a countable subset of the domain. In 1920, Skolem proved this stronger version of the theorem, using the axiom of choice. In 1922, he gave an alternative proof of the weaker version, which does not use choice.

In 1929, Gödel proved the completeness theorem for first-order logic in his doctoral dissertation at the University of Vienna (see [3]). To understand Löwenheim's 1915 paper, it will be helpful to work backwards through Gödel's and Skolem's results. Gödel proved the completeness theorem in a form that is analogous to the statement of Löwenheim's theorem: if a sentence is not refutable in a formal system of deduction for first-order logic, then it has a countable model. In outline, the proof proceeds as follows:

1. First, assign to any first-order sentence φ a sentence φ' in a certain normal form.
2. Show that for any sentence φ, the assertion "if φ is not refutable, then it has a countable model" follows from the corresponding assertion for φ'.
3. Prove the assertion for sentences φ' in normal form.

The key idea is that any first-order sentence φ is equivalent, in a precisely specifiable sense, to a sentence φ' of the form

$$\exists R_1, \ldots, R_l \, \forall x_1, \ldots, x_m \, \exists y_1, \ldots, y_n \, \theta,$$

where θ is a formula without quantifiers, and $R_1, \ldots, R_l$ range over *relations* on the first-order universe. This reduces the task to proving the completeness theorem for sentences of a restricted form. Gödel's proof is sketched in the box that accompanies this article.

The weak version of the Löwenheim-Skolem theorem is a consequence of the completeness theorem: if φ has a model, then it is not refutable, and so φ has a countable model. In fact, if we replace "not refutable" by "has a model" in the second step of Gödel's argument, we end up with, more or less, Skolem's 1922 proof. In letters to Jean van Heijenoort and Hao Wang in the 1960s, Gödel indicated that he knew only of the results of Skolem's 1920 paper when he wrote his dissertation, but he acknowledged that the completeness theorem is implicit in Skolem's 1922 paper—it simply did not occur to Skolem to state it.

Recall that Skolem's 1920 paper proves a stronger form of the Löwenheim-Skolem theorem: if $\mathcal{M}$ is any model satisfying φ, there is a countable *submodel* of $\mathcal{M}$ satisfying φ. Here, the key idea is to use the alternative normal form,

$$\exists f_1, \ldots, f_l \, \forall x_1, \ldots, x_m \, \theta,$$

where θ is quantifier-free and $f_1, \ldots, f_l$ are now *function variables*. Today, a sentence of this form is said to be in *Skolem normal form*, and functions witnessing the existential quantifiers are called *Skolem functions for* φ.

We can now describe Löwenheim's proof of his main theorem. Like Gödel's, it has three steps:

1. Replace the sentence in question by one in a specific normal form.
2. Show that if the original sentence is satisfied in some infinite domain, then so is the one in normal form.
3. Show that if a sentence in normal form is satisfied in some infinite domain, then it is satisfied in a countable domain.

The following three questions form the central thrust of Badesa's investigation:

1. Which version of the theorem did Löwenheim intend to prove, the strong version or the weak one?
2. Is Löwenheim's normal form essentially Skolem normal form?
3. Is the proof essentially correct?

As far as the first question is concerned, Löwenheim stated only the weak version of the theorem, and his construction, on the surface, is similar to the ones used later by Skolem in 1922 and Gödel in 1929. The conventional answer is therefore that Löwenheim aimed to prove the weak version.

As far as the second question is concerned, the issue is largely notational. Building on Schröder's notation, Löwenheim uses the formula

$$\Pi i \Sigma k A(i,k) = \Sigma k_i \Pi i A(i,k_i)$$

to express the equivalence between $\forall x \, \exists y \, \theta(x,y)$ and its Skolem normal form, $\exists f \, \forall x \, \theta(x,f(x))$. At times, Löwenheim seems to suggest that one should think of Σk_i as a sequence of quantifiers $\Sigma k_1 \Sigma k_2 \Sigma k_3 \ldots$ where $1, 2, 3, \ldots$ run through the elements of the domain. But this has the net effect of making the function $i \mapsto k_i$ a Skolem function. The conventional wisdom has therefore held that however Löwenheim *thought* of the quantifier Σk_i, he was, in effect, working with Skolem functions.

Just as in Skolem's and Gödel's constructions (see box), at the final stage of the proof, Löwenheim needed to pass from the satisfiability of each of a sequence of formulas $\psi_1, \ldots, \psi_n$ to their joint satisfiability. But where Skolem and Gödel appeal to what is now known as König's lemma, Löwenheim simply makes the inference without comment. Conventional wisdom has therefore held that Löwenheim overlooked the fact that an argument is needed to justify the inference.

Badesa's account is interesting in that it goes against the conventional wisdom, on all three counts. With a careful and thorough analysis not only of Löwenheim's paper but also the historical context in which he worked, Badesa argues forcefully for a novel reading of Löwenheim's proof. According to Badesa, the hypothesis that the original formula is true in a model does more than guarantee consistency at each stage of a syntactic construction, as in Skolem's 1922 proof; the domain of the model constructed does not consist of syntactic objects but, rather, elements of the model that one started with. The resulting proof is thus an amalgam of Skolem's proofs of 1922 and 1920; one starts with a syntactic construction but then uses that construction to pick out a subdomain of the original model. This eliminates the need to appeal to König's lemma. As a result, Badesa argues that Löwenheim offered an essentially complete and correct proof of the strong version of the theorem.

At the heart of the ambiguity is precisely the lack of a clean separation between syntax and semantics. What allows for the two readings is the fact that in Löwenheim's notational and conceptual framework, it is not always easy to tell whether he is referring to symbols or the elements they denote under a particular interpretation. Badesa further notes that this can result in an important difference between the use of Skolem functions and Löwenheim's quantifiers $\sum k_i$. If the indices to variables $k_1, k_2, k_3, \ldots$ are considered to be symbols rather than elements of an underlying domain, it is possible, say, for k_{17} and k_{23} to denote different elements, even though 17 and 23 may stand for the *same* element of that domain.

In the end, Badesa's analysis illuminates not just Löwenheim's proof but also the importance of a conceptual framework that we take for granted today. With his careful analysis of the route by which we arrived at the modern model-theoretic understanding, Badesa has provided us with an insightful account of the emergence of a new field of mathematical inquiry.

REFERENCES

[1] George Boole. *The Mathematical Analysis of Logic*. Barclay and Macmillan, Cambridge, 1847. Reprinted in [2], volume 1, pages 451–509.

[2] William Ewald, editor. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, two volumes. Clarendon Press, Oxford, 1996.

[3] Kurt Gödel. *Collected Works*, volume 1. Solomon Feferman *et al*., eds. Oxford University Press, New York, 1986.

[4] Wilfried Sieg. Hilbert's programs: 1917–1922. *Bulletin of Symbolic Logic* **5**:1–44, 1999.

[5] Jean van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, Cambridge, Massachusetts, 1967.

Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213
USA
e-mail: avigad@cmu.edu

Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s

by Dennis E. Hesseling

BASEL, BIRKHÄUSER, 2003
XXIII + 448 PP. US $139.00. ISBN 3-7643-6536-6

REVIEWED BY JEREMY AVIGAD

The early twentieth century was a lively time for the foundations of mathematics. The ensuing debates were, in large part, a reaction to the set-theoretic and nonconstructive methods that had begun making their way into mathematical practice around the turn of the twentieth century. The controversy was exacerbated by the discovery that overly naïve formulations of the fundamental principles governing the use of sets could result in contradictions. Many of the leading mathematicians of the day, including Hilbert, Henri Poincaré, Émile Borel, and Henri Lebesgue, weighed in with strong views on the role that the new methods should play in mathematics. Dennis E. Hesseling's book, *Gnomes in the Fog*, documents reactions to the "crisis of foundations" that was inaugurated by the attempts of L. E. J. Brouwer to refound modern mathematics on "intuitionistic" principles.

Brouwer's interests in foundational issues extend at least as far back as his 1907 doctoral dissertation at the University of Amsterdam. In that work, he tried to ground mathematics, especially the mathematics of the continuum, on the basis of *a priori* intuition of time together with processes of mental construction. According to Brouwer, such constructions precede language, which provides only a flawed means of communication *ex post facto*. Since logic is typically presented as a collection of linguistic principles, Brouwer took logic, as well, to be secondary to mathematical understanding. In doing so, he countered tendencies, seen in Russell and Hilbert, to view mathematics as based essentially on logic and language. Brouwer's dissertation also includes discussion of the law of the excluded middle and the nature of mathematical existence, two topics that, as Hesseling makes clear, were to become central to the foundational debates later on.

It was, however, in the newly emerging field of topology that Brouwer first made a name for himself. Between the years 1909 and 1913, he not only clarified basic terminology and helped put the subject on a rigorous foundation, but also introduced many of the field's central methods. His results on fixed-points of continuous transformations, degrees of mappings, and invariance of dimension were groundbreaking and seminal. But issues having to do with the grounding of point-set topology in set-theoretic terms were closely related to foundational issues in his dissertation. He returned to these themes in 1918, with a paper, "Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten" ["Foundation of set theory independent of the logical law of the excluded middle"]. Here he put forth a radical alternative to Cantorian set theory, in conformance with his intuitionistic views of mathematics, based on the notion of a "choice sequence." In 1923, he developed an intuitionistic theory of real-valued functions, which showed how notions of continuity, measurability, and differentiability could be developed in intuitionistic terms. A remarkably clear and concise introduction to Brouwer's philosophical and foundational contributions can be found in Mark van Atten's *On Brouwer* [2].

The transition to twentieth-century mathematics involves two important tendencies:

1. the shift of focus from symbolic expressions to abstract objects; and
2. the use of structural, set-theoretic, and infinitary methods to describe such objects, operations on them, and their properties.

Brouwer was by no means rejecting the first tendency, or advocating a return to the algorithmic sensibilities of the nineteenth century. I have already noted his emphasis on mental constructions and skepticism toward the symbolic or linguistic devices used to describe them. Neither did he reject the second tendency; his 1927 proof of the "bar theorem" used transfinite induction, and he treated proofs as infinitary objects on which mathematical operations can be performed. But, in Brouwer's view, the proper methods of *doing* mathematics were to reflect an appropriate "constructive" or "intuitive" understanding. Between 1925 and 1934, despite Brouwer's antipathy toward formalism, various aspects of intuitionistic reasoning received formal treatments in the hands of Kolmogorov, Glivenko, and Brouwer's own student, Heyting. With the appearance of formal definitions of computability in the 1930s, it was not long before mathematical logic had uncovered connections between intuitionistic mathematics and computability, providing ways in which various intuitionistic methods can be understood in computational terms. Among modern practitioners of constructive mathematics, such computational aspects of the practice tend to be emphasized more often than Brouwer's philosophical doctrines. (See, however, [2, Chapter 5] for an aspect of Brouwer's intuitionism that cannot be understood in computational terms.)

In any event, whatever the philosophical presuppositions, Brouwer's methodological prescriptions would have constituted a radical shift in mathematical practice. Coming from a mathematician of Brouwer's stature, this challenge to set-theoretic foundations could not be ignored. But, as Hesseling observes, Brouwer's challenge didn't become a movement until Hermann Weyl's publication of "Über die neue Grundlagenkrise der Mathematik" ["On the New Foundational Crisis in Mathematics"] in 1921. (This and related foundational essays are translated and gathered in [1].) Weyl, a student of Hilbert's, made lasting contributions to diverse branches of mathematics, including function theory, analytic and algebraic number theory, representation theory, and mathematical physics. From an early stage in his career, he was influenced by Edmund Husserl's phenomenology, and traces of this influence can be discerned in his landmark 1918 treatment of general relativity in *Raum, Zeit, Materie* [*Space, Time, and Matter*]. In that same year, he published a work, *Das Kontinuum* [*The Continuum*], in which he developed a foundational approach to replace the contemporary set-theoretic formulations of analysis. Referring to the latter, he wrote that "every cell . . . of this mighty organism is permeated by the poison of contradiction and . . . a thorough revision is necessary to remedy the situation." In 1921, however, he proclaimed, "I now renounce my attempt and join Brouwer's," and joined forces with the new "revolution."

A good deal has been written about Brouwer, including a recent biography [3] by van Dalen. In *Gnomes in the Fog*, Hesseling has made the interesting choice of documenting *reactions* to Brouwer's manifestos rather than focusing on Brouwer himself. He has been extraordinarily thorough. The introduction tells us that he has analyzed more than 1,000 primary sources, including published papers in mathematical and non-mathematical journals, newspaper articles, correspondence, and unpublished manuscripts, drawing on archives from all over Europe. An appendix lists more than 250 published works, almost all of which appeared between 1921 and 1933. Many of these works are discussed explicitly in the text.

> *Classical first-order logic is the result of adding the law of the excluded middle to intuitionistic logic.*

Hesseling draws some interesting conclusions. At a symposium on the foundations of mathematics in Königsberg in 1930, Rudolf Carnap, Heyting, and Johann von Neumann presented papers on logicism, intuitionism, and formalism, respectively. Ever since then, there has been a tendency to characterize the "crisis of foundations" as a struggle among these three. Hesseling notes, however, that in the range of foundational writings he considered, logicism, as characterized by Carnap, plays at best a minor role. Furthermore, Hesseling argues that formalism, as characterized by von Neumann, was not a clearly articulated position from the start but, rather, was gradually shaped in response to intuitionistic challenges.

Hesseling's three opening chapters serve to provide background, to describe Brouwer's work, and to present the opening salvos of the debate. His last, somewhat speculative chapter, explores connections between intuitionism and analogous political and cultural currents. Two chapters in between are titled "Reactions: existence and constructivity" and "Reactions: logic and the excluded middle." These aim primarily to summarize the reactions to intuitionism found in the corpus of documents that Hesseling has analyzed. The chapter titles are somewhat misleading: it would be a mistake to draw a sharp distinction between the intuitionistic views of existence statements (formally, $\exists x\, \psi$) and the law of the excluded middle (formally, $\psi \vee \neg\psi$), since the two are intertwined. The two chapters, rather, aim to separate metaphysical questions from logical ones. From a metaphysical point of view, one is concerned with the task of giving a general account of mathematical knowledge and mathematical objects, whereas from a logical perspective, one is concerned with identifying the proper principles of mathematical reasoning. These are certainly related; one would expect a metaphysical stand to have bearing on the proper principles of reasoning, and, conversely, one would expect logical principles to be justified by reference to a metaphysical view. It is, nonetheless, often useful to keep these two foci distinct.

Despite Brouwer's misgivings about formalism, mathematical logic proved to be a valuable tool in clarifying the differences between intuitionistic and classical practice. Although he is occasionally imprecise with the details, Hesseling does an able job of documenting these developments. For example, formalizations of intuitionistic logic show that the principle $\neg\neg\psi \rightarrow \psi$ of "double-negation elimination," which allows one to prove an assertion by showing that its negation implies a contradiction, is equivalent, as a schema, to the law of the excluded middle, $\psi \vee \neg\psi$. Instances of the latter can also be used, together with intuitionistic logic, to prove $\neg\forall x\, \psi \rightarrow \exists x\, \neg\psi$. This provides a similarly indirect means of proving an existence statement. In fact, classical first-order logic can be characterized as the result of adding the law of the excluded middle to intuitionistic logic. Various "double-negation translations," from Kolmogorov and Glivenko to Gödel and

Gerhard Gentzen, provide ways in which classical forms of reasoning can be interpreted in intuitionistic terms. The differences between Brouwer's proposed methods and those sanctioned by set theory, however, went beyond the axioms and rules governing the logical connectives; set theory sanctions nonconstructive definitions of infinitary sets and sequences that are intuitionistically forbidden. Here, too, logical analysis helped clarify Brouwer's intuitionistic methods, and their contrast to set-theoretic ones.

Hesseling's account of the progress that was made towards clarifying the metaphysical differences between the two conceptions is less satisfying. It is notoriously difficult to find a clear, rational basis from which to address questions having to do with mathematical meaning and existence, since the very project presupposes a general conception of meaning and existence with respect to which one situates the mathematical versions. Since there are no uniformly accepted candidates for the former, one is left addressing weighty issues without a firm ground to stand on. But one can at least try to be clear about one's presuppositions and the philosophical framework on which one's analysis is based, and such clarity and precision is absent from Hesseling's narrative. To be sure, the fault often lies with the original authors; but even the most carefully articulated position sounds flimsy when summarized in a few sentences, and the benefit of hindsight should have afforded a clearer articulation of the philosophical issues at stake.

Hesseling's overview does manage to convey a sense of the issues. Whereas the intuitionistic notion of existence is variously described in terms of definability, describability, algorithmic computability, intuitive construction, or phenomenological experience, the notion of classical existence is typically understood as somehow "independent" of all these. So, the general philosophical problem for classical logic was taken to be finding some sort of justification for the associated knowledge claims. In the debates of the 1920s, one sees traces of the idea, later to become a central part of Quine's views, that theoretical statements are to be justified holistically rather than at the sentential level. In particular, classical existence claims are to be justified in terms of global properties, like consistency, of the theory in which they are a part, rather than in terms of a more "local" analysis of their meaning.

Hesseling's presentation also falls short in its analysis of the mathematical context in which the debates took place. Brouwer's work in topology merits only a brief discussion, with the conclusion that "Brouwer's dissertation and his topological work mostly follow naturally from the same basic principles." It would have been nice to have a better sense of how Brouwer's treatment of the continuum, for example, compared to that of rival foundational approaches, and the substantive mathematical issues the different developments were meant to address. Similarly, it would have been nice to have more insight into the relationship between Weyl's purely mathematical work and his foundational and philosophical views. It is important to keep in mind that Brouwer's critique cut to the core of what it means to do mathematics, and was designed to bear upon the day-to-day practice of every working mathematician. Separating the critique from an understanding of how it affects that practice gives the debate the character of a rhetorical exercise, belying its pragmatic relevance to a subject that plays such a central role in human thought and experience.

> *Following the path that has led to our current understanding might lead us to wonder whether alternative paths may better suit our purposes*

In sum, Hesseling has aimed for breadth rather than depth in his analysis. The result is an interesting sociological study of the cultural processing of new ideas in an important scientific discipline. The work also provides a thorough survey of historical data that should be used to support a better mathematical and philosophical understanding of the foundational issues. It would be unfair to be overly critical of Hesseling for not making further progress towards developing such an understanding, when the mathematical and philosophical communities, themselves, have not done much better. One gets the sense that, over time, mathematicians simply got used to the new set-theoretic methods, and, finding them convenient, grew tired of debate as to whether they are appropriate to mathematics. While this was going on, philosophers of mathematics, understandably reluctant to make declarations as to the proper practice of mathematics, honed the ability to frame issues in such a way that philosophical analyses can't *possibly* have any bearing on what mathematicians actually do. These tendencies towards apathy, on the one hand, and irrelevance, on the other, have been effective in severing substantial communication between the two communities. The sharp disciplinary separation has, in turn, reinforced these tendencies.

Readers of Hesseling's book are likely to be left with conflicting emotions. After seeing the methodological and foundational confusions of the early twentieth century recounted at length, it is hard not to feel a sense of relief that these days are now behind us; coupled, perhaps, with a touch of pity for the mathematicians and philosophers who had to struggle through them. But reflecting on the historical development might also shake us out of our complacency, and following the tortuous path that has led to our current understanding might lead us to wonder whether alternative paths may better suit our purposes. To be sure, the modern set-theoretic methods for reasoning about the mathematical objects have become central to mathematical thought. But we can imagine that one day mathematical sensitivities may evolve so that questions having to do with symbolic representations and algorithms once again form the core of the subject. Or, perhaps, a better synthesis will make it possible to enjoy the heuristic value of modern conceptual methods while preserving the subject's algorithmic content. Mathematicians of the future may well look back at the philosophical and methodological confusions of the early twenty-first century and take pity on *us*, just as we pity our forebears.

Lacking a time machine, we can only speculate as to what the future will bring. But we also play a role in shaping that future, and so it is important that we engage in the subject with an appropriately reflective attitude. The merit of historical works like Hesseling's is that, by bringing our preconceptions to the fore, they help us understand them better.

ACKNOWLEDGMENTS

I am grateful to Harold Edwards, Tobias Nipkow, Gabriel Sabbagh, Mark van Atten, Dirk von Plato, and Richard Zach for comments and corrections on this review and the one preceding it. The attitudes and opinions I have expressed, however, may differ from theirs.

REFERENCES

[1] Paolo Mancosu, editor. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920's.* Oxford University Press, Oxford, 1998.

[2] Mark van Atten. *On Brouwer.* Wadsworth Publishing, Belmont, California, 2003.

[3] Dirk van Dalen. *Mystic, Geometer, and Intuitionist: The Life of L. E. J. Brouwer.* Two volumes. Oxford University Press, New York, 1999 and 2005.

Department of Philosophy
Carnegie Mellon University
Pittsburgh, PA 15213
USA
e-mail: avigad@cmu.edu

Coincidences, Chaos, and All That Math Jazz: Making Light of Weighty Ideas

by Edward B. Burger and Michael Starbird

NEW YORK, W. W. NORTON & COMPANY, 2005, ix + 276 PP. US $24.95, ISBN 0-393-05945-6

REVIEWED BY JOHN J. WATKINS

There is nothing quite like a good thriller. I recently picked one up during a relatively quiet week when I was planning also to read the new Burger and Starbird book that is the subject of this review. (I often like to have several books going at once, to fit whatever mood I am in at the moment.) We all know that the main point of a thriller is page-turning escape, and it was a blurb on the back cover of this one that really caught my attention:

> *Bangkok 8* is one of the most startling and provocative mysteries that I've read in years. . . . Once I started, I didn't want to put it down.—Carl Hiaasen

Surprise. . . . It turned out that *Coincidences, Chaos, and All That Math Jazz* was the page-turning thriller I couldn't put down, for the simple reason that I wanted to see what happened next in the story. And so, I'd like to offer their publishers the following ready-made blurb for the back cover of the next edition of this remarkable book:

> *Coincidences, Chaos, and All That Math Jazz* is one of the most startling and provocative mysteries that I've read in years. . . . Once I started, I didn't want to put it down.—John Watkins

Of course, the publishers might justifiably be concerned about issues of plagiarism, but really, is there any likelihood at all that Carl Hiaasen would ever notice his words being used on the cover of a math book? Well, as it happens, if the publishers (or their lawyers) actually read their own book, they might be able to estimate the probability of such an apparently unlikely coincidence. In fact, they would only need to read as far as the first chapter, where Burger and Starbird begin their story—with coincidences.

And what coincidences! And what a gentle introduction they provide the general reader to the wealth of mathematical ideas in this fine book. The ideas range from the familiar, such as the astonishing number of coincidences that have been noticed in the lives of Abraham Lincoln and John F. Kennedy and the remarkable coincidences that have been reported in the lives of identical twins separated at birth, to fresh examples of "coincidences," such as a stock market scam in which an investment company builds up such an impressive record of stock predictions week after week that it is guaranteed to make buckets of money from an unsuspecting public, and a card trick that would have been completely new to me had I not just seen it performed by a magician only a week earlier. Two people turn up cards one at a time from two decks of cards and the "trick" is to wait and see if two of their cards ever match; that is, will the two people ever simultaneously turn up exactly the same card? What is wonderful about this trick is that knowing the mathematics behind it—and that there will be a match about two-thirds of the time—doesn't in the least detract from the observer's delight when a match actually occurs; it really does seem like magic.

As suggested by the book's title, the authors turn to chaos in the second chapter. The lineup includes all the usual suspects—butterflies from Brazil, dripping faucets, various pendulums both double and magnetic—but they begin this look at non-predictable behavior with a well-chosen hands-on Excel spreadsheet revealing the surprisingly chaotic iterative behavior of the simple polynomial $x^2 - 1$. Burger and Starbird have a real knack for judging exactly how much mathematical detail to include in their discussion of a given topic, and one of the pleasures of reading this book is to watch to see exactly how they manage to communicate mathematical ideas to the reader by finding creative and clever ways to avoid technical matters almost entirely. By the end of this chapter, readers not only have discovered the fundamental nature of chaos but, more important, they appreciate why we shall forever be severely limited in our ability to predict the future of even the simplest of physical systems. (By the way, in my parallel reading universe in Thailand, I was also learning that karma is a weather system far too complex to analyze.)

As the mathematical story continues on, into the "all that math jazz" portion of the book, the topics do begin to feel somewhat disconnected: statistics, cryptography, orders of magnitude, Fibonacci numbers, fractals, topology, the fourth dimension, and infinity. But the reader is having so much fun along the way that there is hardly time to notice. Again, there is much that is familiar; in particular, much of this material has already appeared in their previous book *The Heart of Mathematics: An Invitation to Effective Thinking,* a wonderfully engaging textbook I have greatly enjoyed using in a course introducing students to mathematical thinking. But, in the end, what connects all of these topics is not subject matter, but rather process. The theme that provides the narrative pulse of the book is that very complicated things often can be understood by using simple ideas. The authors begin training the reader in this process early in the book, exploring secrets of nature that are revealed in number patterns, as well as mathematical secrets, such as Goldbach's Conjecture and the Twin Prime Conjecture, that remain tantalizingly beyond reach. Gradually the authors take the reader deeper, into the world of fractals, using nothing more than a strip of paper to be folded again and again, and into the world of topology, using tavern puzzles and doughnuts. They connect each of these wondrous mathematical worlds to the real world with well-timed discussions of Turing machines, DNA replication, and the possible shape of our universe. Momentum builds in the book as readers build their thinking skills: by the final chapters, they will find themselves unlocking the mysteries of 4-dimensional space and going places they never dreamed possible, beyond infinity.

Burger and Starbird write about mathematics with a light touch that may well make them the envy of anyone writing about mathematics today. They take their subtitle, *Making Light of Weighty Ideas*, seriously. They enjoy words and use them with great effectiveness and often with great humor. Frequently, they even achieve the power of poetry in their use of language, as when they write "the stream of natural numbers flows on endlessly," a particularly lovely metaphor that not only conveys the inductive infinitude of the natural numbers but also evokes their timelessness as well. And yet, for my taste, they too often succumb to their fondness for alliteration and puns. I'll confess I was quite amused when they referred to one particularly large never-before-seen-in-all-of-human history integer as a "digital debutante," but then I groaned audibly when, on the very next page, this same number became a "numerical nymphet." When they choose "Turning to Turing" as the heading for a section on computing I am hugely impressed by the clever wordplay, or when they choose the Saint Louis Cardinals as the baseball team that will be checking into the Infinite Hotel, I smile in appreciation at their sly nod to the notion of cardinality, but when they let a pun such as "duck the fowl issue" slip onto a page already full of swans and dragons, I again can only groan. Clearly, they know that they are unable to restrain themselves and even apologize at one point for a particularly bad sentence they use to sum up an otherwise superb discussion of tiling patterns: "In tiling your bathroom, you are now flush with philosophical possibilities that may be far too draining."

This book is written for general readers, even those with little or no math background—the publisher boasts, not very convincingly, I'm afraid: "If you never thought you would read about mathematics, this book is for you."—and I will be recommending it to almost everyone I meet for years to come, but it also contains much to surprise and delight those of us who think we have seen all of this before. You learn how George Gallup (of Gallup Poll fame) got his start during the 1936 U.S. presidential election, what a useful word "quincunx" is, what happens if you balance 100 pennies on edge and then whack the table, how a public policy designed to *reduce* airline safety might actually save lives, why Bill Gates shouldn't even bother to pick up a stray $100 bill lying on the floor, and how we could take a very large stack of playing cards and carefully cantilever them one-by-one off the edge of a table and then walk out and sit ourselves down on the very top card and dangle our feet over the void below just as casually as if we were sitting at the end of a solidly constructed but very high diving board.

This book is beautifully illustrated. In fact, the interplay between text and figures may be its greatest virtue. In some cases this is just absolute first-rate use of a series of figures that vividly bring out a mathematical truth, such as the self-similarity of a continued fraction built from an endlessly repeating sequence of 1s or using an especially nice sketch of a villa designed by Le Corbusier to illustrate the underlying geometry of the Golden Rectangle. The drawing used to illustrate the familiar tale about the height of an ideal person's belly button being a manifestation of the Golden Ratio may not be the divine proportion sketch by da Vinci that one has come to expect, but it is now the one that will forever appear in my mind's eye whenever I think simultaneously of my belly button and the Golden Ratio. The illustrations in the chapter on the fourth dimension are particularly effective: cartoon figures of 2-dimensional creatures with oddly placed eyes and mouths prepare the reader for the far more ambitious mental visualization needed to understand fully the sketches that demonstrate so clearly how one can untie knots in four dimensions or how the Klein bottle can sit quite happily in 4-dimensional space without passing through itself. By taking readers on a journey that sees such marvelous mysteries as these unravel before their eyes, Burger and Starbird have indeed unwittingly placed themselves among the very best thriller writers of our times.

REFERENCES

1. E. B. Burger and M. Starbird, *The Heart of Mathematics: An Invitation to Effective Thinking*, Key College Publishing, 2004.
2. John Burdett, *Bangkok 8*, Alfred A. Knopf, 2005.

Department of Mathematics and Computer Science
Colorado College
Colorado Springs, CO 80903
USA
e-mail: jwatkins@coloradocollege.edu

The index and/or table of contents has been removed and photographed separately within this volume year.

For roll film users, this information for the current volume year is at the beginning of the microfilm. For a prior year volume, this information is at the end of the microfilm.

For microfiche users, the index and/or contents is contained on a separate fiche.

John von Neumann Stamps

Magdolna and István Hargittai

John von Neumann (1903 in Budapest–1957 in Washington, DC) is often called the father of the modern computer, but it is perhaps more accurate to single him out for introducing the stored program into modern computation. He acquired degrees both in chemical engineering and mathematics simultaneously at the Swiss Federal Institute of Technology in Zurich and in Budapest, and started his career with great success in Germany. Anti-Semitism and Nazism forced him out of Europe and he lived the rest of his life in the United States where Princeton, first the University, then the Institute for Advanced Study, provided him the necessary environment for his creative activities. Between the mid-1920s and early 1950s, he published extensively on the axiomatization of mathematics, the mathematical foundations of quantum mechanics, game theory, and other topics, including what later became known as molecular biology. He eagerly participated in the defense of his adopted country in World War II and the Cold War by contributing to a variety of weapon developments, including the atomic and hydrogen bombs. During the last years of his short life he held high appointments in defense and administration. As Commissioner of the U.S. Atomic Energy Commission, he had risen to the highest position among the famous group of Hungarian physicists who have been known as the Martians.[1]

Magdolna Hargittai
Eötvös University
and Hungarian Academy of Sciences

István Hargittai
Budapest University of Technology
and Economics and
Hungarian Academy of Sciences

REFERENCE

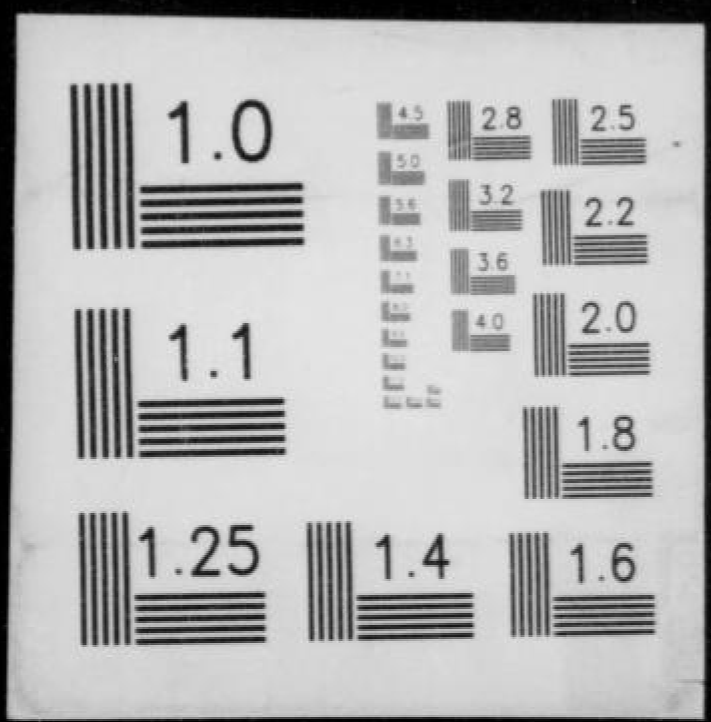
1. I. Hargittai, The Martians of Science: Five Physicists Who Changed the Twentieth Century, Oxford University Press, New York, 2006.

Von Neumann's daughter, Marina von Neumann Whitman in her home with the enlarged image of the U.S. stamp of von Neumann (photograph by the authors).

First-day of issue envelope.

Two Hungarian stamps honoring John von Neumann (Neumann János), issued in 1992 and 2003, and a stamp from Guyana.

Please send all submissions to
the Stamp Corner Editor,
Robin Wilson, Faculty of Mathematics,
The Open University, Milton Keynes,
MK7 6AA, England
e-mail: r.j.wilson@open.ac.uk


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0