International Review of Law and Economics 20 (2000) 187–204

International Review of Law and Economics

Patenting mathematical algorithms: What's the harm? A thought experiment in algebra

Paul B. de Laat*

Faculty of Philosophy, University of Groningen, A-weg 30, 9718 CW Groningen, The Netherlands

Abstract

The patenting of software-related inventions is on the increase, especially in the United States. Mathematical formulas and algorithms, though, are still sacrosanct. Only under special conditions may algorithms qualify as statutory matter: if they are not solely a mathematical exercise, but if they are somehow linked with physical reality. In this article, it is argued that blanket acceptance is to be preferred. Moreover, the best results are obtained if formulas and algorithms are only protected in combination with a proof that supports them. This argument is developed by conducting a thought experiment. After describing the development of algebra from the 16th century up to the 20th (in particular, the solution of the cubic equation), the likely effects on the development of mathematics as a science are analyzed in the context of postulating a patent regime that would actually have been in force protecting mathematical inventions.

Keywords: Algebra; Algorithm; Patent; Software

1. Introduction

Property rights for software traditionally have been rather weak. Only trade secrecy has offered some protection. In the last two decades this state of affairs has changed. Since software piracy has become rampant, the industry has sought more protection. At first firms applied for copyright protection: developers went to the courts and sued supposed infringers of their rights. Until the beginning of the 1990s, this strategy was rather successful: more and

* Corresponding author. Tel.: +31-50-3015902; fax: +31-50-3636160.
E-mail: p.de.laat@philos.rug.nl (P.B. de Laat).

PII: S0144-8188(00)00031-4

more aspects of software were acknowledged as being covered by copyright. Famous lawsuits of the 1990s revolved around the “structure, sequence, and organization” of software, and “look-and-feel” issues. This wave of protection, however, is waning now. More and more, courts limited the actual reach of copyright protection. Nowadays, only verbatim pieces of code may be considered protected by copyright.

The spearhead of software protection is now in patenting, or, to be more precise, software developers seek to patent their software-related *inventions*. In applying for a patent, two hurdles must be negotiated. The invention must be found to be permissible (“statutory”) subject matter, as either a process or a product (i.e., machine, manufacture, or composition of matter), and it should be useful, novel, and nonobvious. If in the end a patent has been obtained, it protects much more than copyright: it covers not only the *expression,* but also the *idea* behind the software. A monopoly has been acquired. Although at first the courts simply refused to grant patents on software, patents are now issued more readily, and it has become a routine matter to apply for them. Since *Diamond v. Diehr* (1981), and especially since *In re Alappat* (1994), the door has been opened wide for software patents, at least in the United States. Thousands of them are issued each year, and the wave is still on the increase. Elsewhere, especially in Europe, things may move in this direction as well.

There is one limit that coverage by a patent may not cross: mathematical formulas, algorithms, and the like may not be patented (for the following account that takes U.S. developments as a point of departure, see Samuelson, 1990; Evans & Johnson, 1994/1995; Stern, 1995; Kuester & Moceyunas, 1995; Patent and Trademark Office, 1996; Bender, 1997; Gibby, 1997). The U.S. Supreme Court holds, that “phenomena of nature [...], mental processes, and abstract intellectual concepts are not patentable, as they are the basic tools of scientific and technological work” (*Gottschalk v. Benson*, 1972). As a consequence, (mathematical) algorithms in software are not deemed patentable.

At first, this threshold was interpreted in an absolute fashion: software-related inventions that included a formula never passed the test. However, since *Diamond v. Diehr* (1981) this limit gradually has eroded. More and more cases involving mathematical formulas have been presented successfully to the courts. As a result, in 1989 the U.S. Patent and Trademark Office (PTO) issued new examination guidelines for such software-related inventions. The change of policy was formulated as follows. If a software-related invention is filed for application as a *process*, and it contains the use of mathematical formulas, it will be considered statutory matter if (and only if) two conditions are satisfied: (a) the algorithm should be applied to physical elements or process steps, in order to avoid the situation that the issuing of the patent would “wholly preempt” use of the algorithm in general; and (b) the end-product should not be “pure number” (*Freeman–Walter–Abele Test*). Later on, this pure-number condition further eroded as a criterion. Even if the output is a pure number, the claim is statutory as long as it represents something other than a mathematical abstraction (cf. *Arrhythmia, Research Technology, Inc. v. Corazonix Corporation,* 1992, see below).

In 1994 these guidelines were severely questioned by rulings from the Federal Circuit (in particular, *In re Alappat* and *In re Warmerdam*). The court argued that if mathematical algorithms are involved, the PTO should look at the invention *as a whole* and judge whether it represents more than only laws of nature, natural phenomena, or abstract ideas. If (and only if) the answer is affirmative, the claimed process is deemed statutory.

After a lengthy public discussion, the PTO in 1996 adopted new, less stringent guidelines for its personnel (PTO, 1996). Although the Freeman–Walter–Abele Test remained in force, the following set of rules was primarily to be relied on in analyzing process claims that involve mathematical algorithms (PTO, 1996, pp. 7483–7485). If a software-related invention consists *solely* of mathematical operations, it will not be considered statutory. This is the case if it is a "mathematical definition of a law of nature" ("for example, a mathematical algorithm representing the formula $E=mc^2$") or an abstract idea that is expressed in a mathematical algorithm. A software process may be acceptable matter, however, if a demonstrable link with physical reality exists. According to the new guidelines, this can be achieved in two ways. First, it can be argued that the process requires physical acts to be performed outside the computer. These acts may take place either *before* the computer starts operating ("pre-computer process activity"; e.g., the case of software for the diagnosis of heart problems, which incorporates an algorithm that measures and manipulates a patient's electrocardiographic signals, *Arrythmia*, 1992), or *after* the software has run ("post-computer process activity"; e.g., a method of curing raw rubber in a mould that relies on computer-generated data, as in *Diamond v. Diehr*, 1981). These categories are referred to by the PTO as "safe harbors." As another possibility for a link with physical reality, the proposed process may be tied to "a practical application within the technological arts" (e.g., a noise-modeling algorithm that is actually used to remove noise from digital signals).

So much for patent claims for process. A second way to claim software-related inventions as patents is not as a process, but as a specific *machine*. Usually, passing this statutory test is much easier. Therefore, many software developers proceeded to formulate their software processes in machine form and to file their applications in the so-called "means-plus-function" format. Although such claims at first were not acknowledged and were simply treated as process claims, since *In re Alappat* (1994) and *In re Warmerdam* (1994) such software applications are increasingly being treated as machine claims. A second avenue for software patent application definitely has opened up. Many applicants, therefore, proceed to file process and machine claims *together*. In particular, this has implications for inventions involving mathematical formulas and algorithms. One may try both approaches; of the two, the machine format option, although more restricted in scope, offers more chances of success. As long as the claimed invention contains specified pieces of software code, it cannot be labeled "disembodied" and will pass the test.

Because of these developments, more and more abstract-looking algorithms acquire the statutory label. The patent that Narendra Karmarkar obtained in 1988 is illustrative (Samuelson, 1990, pp. 1099–1102). It involves a new algorithm in linear programming that addresses the traveling salesman problem and accomplishes the trick of considerably reducing the number of calculations to be made. The algorithm was claimed as a method for "all industrial applications."

Nevertheless, limits still exist. Not everything goes. A recent illustrative example is *In re Schrader* (1994) (Bender, 1997, § G.2.a.i.a). The application claimed a novel auction method for competitive bidding on related items (like contiguous tracts of land). Winning bids are determined by identifying that particular combination of bids that optimizes the return for the seller. The claim was rejected since it was found to involve a mathematical algorithm, without anything physical in the bids per se. Another more complicated example is *In re*

Warmerdam (1994) (Evans & Johnson, 1994/1995, III, § "Divisions within the Federal Circuit"; Gable & Dengler, 1995; and Bender, 1997, § G.2.a.i.c.). The claimed invention controls the motion of objects such as robots to avoid collision with other objects. Its core is an abstract model of bubbles that is manipulated to create a data structure that reflects the so-called bubbles hierarchy. The claim to this as a *method* was found to be unstatutory, as it was judged to be "nothing more than the manipulation of basic mathematical constructs, the paradigmatic 'abstract idea'." Similarly, the claim to this as a *data structure* was found unstatutory, as it was found to be simply another way of describing the manipulation of ideas contained in the bubbles method. But the claim to this as a *machine* programmed according to the bubbles method was found to be statutory.

Mathematical formulas and algorithms in abstract form, one may conclude, are still sacrosanct. In order to let science and technology flourish, these must remain unpatentable. In this article, it will be argued that the opposite applies. This kind of prudence is superfluous, even damaging. Precisely by *granting* patents on abstract formulas and algorithms the development of science and technology may well be furthered. Let me be more precise. Even if practical context is not specified at all (which amounts to a blank check for all commercial applications imaginable), and it is all pure abstraction or pure number, inventive algorithms should be patentable.

The argument will be explored by looking back in time at the Renaissance. Algebra, a prominent branch of mathematics, was still in its infancy. But from 1500 onward, big steps forward were taken. After describing some of these developments until the end of the 19th century, a thought experiment will be carried out: if a patent system would have been in force from 1500 onward, what would have been the effects on the development of algebra proper in that period? Would it have stimulated further advances, or hampered them? It will be argued that by a suitable choice of patent regime intellectual stimuli might, indeed, have been generated.

2. Solution of the cubic equation

Let us go back to the beginning of the 16th century and to northern Italy. Mathematics came into flux, especially the field of algebra. The great breakthrough was the solution of the cubic equation; soon after, by the way, the quartic equation (of degree 4) also was solved. In modern notation, the general cubic is of the form

$$x^3 + ax^2 + bx + c = 0.$$

Before this solution, some specific numerical cases had been solved, but any attempt at a general solution had failed. A mathematician like Pacioli had in 1494 even declared the problem to be unsolvable.

To appreciate the value of this breakthrough, it should be stressed that no mathematician of the time realized that basically there is just *one* cubic equation to be solved. Because negative numbers were not accepted as bona fide, all variations of the cubic, exchanging terms from left to right of the = sign, were seen as separate problems. Similarly, the "reduced" cases with quadratic coefficient a or linear coefficient b zero were treated separately. This can be seen clearly from the following list of "primitive cases" for the cubic (as provided by Cardano, 1968, Chapter II]. First, all "permutations" of $x^3+bx+c=0$ (keeping + signs throughout) were listed as separate problems (Cardano, 1968, p. 25, cases 4–6]:

$$x^3 + bx = c,$$

$$x^3 = bx + c,$$

$$x^3 + c = bx.$$

A similar list was obtained for $x^3+ax^2+c=0$ (Cardano, 1968, p. 25, cases 8–10):

$$x^3 + ax^2 = c,$$

$$x^3 = ax^2 + c,$$

$$x^3 + c = ax^2.$$

Finally, the complete cubic $x^3+ax^2+bx=c$ was "permutated" in a similar fashion, yielding seven more cases (Cardano, 1968, pp. 25–26, cases 12–22). In all, then, 13 cubics were distinguished.

How the cubic came to be solved is a complicated and still partly unclarified story in which several mathematicians were involved (cf. Ball, 1960, Chapter XII; Burton, 1997, Chapter 7; Cardano, 1968, foreword by Oystein Ore, preface by T. Richard Witmer; Matthiessen, 1878, § 127; Restivo, 1992, pp. 62–65; Smith, 1958, vol. I, pp. 295–300 and vol. II, pp. 459–464). Scipione del Ferro of Bologna would seem to have been the first to have solved the cubic (about 1515); at least, that is, of the form $x^3+bx=c$. As we know now, this knowledge is in fact enough to solve *all* kinds of cubics (cf. below). In the tradition of those days, he kept the secret for himself. Only on his deathbed he handed it over to one of his pupils, Antonio Maria Fiore of Venice (1526). To make a name for himself, Fiore in 1535 challenged Niccolo Tartaglia of Brescia to a public contest in problem solving. Rumor had it that Tartaglia, a self-taught mathematician, had solved the cubic; in particular, $x^3+ax^2=c$. Knowing this, Fiore specifically chose Del Ferro-type cubic problems for the contest. Just before the contest took place, however, Tartaglia managed to solve that kind of cubic equation as well. Thereupon, he won the contest by solving Fiore's problems easily and simply submitting only Tartaglia-type cubics to his opponent. To modern standards, therefore, by 1535 both Del Ferro and Tartaglia basically had solved the cubic generally.

Then enters Girolamo Cardano of Milano, a controversial medical doctor, astrologer, and mathematician. As an exception to the rule of secrecy, he was preparing a mathematical textbook. He begged Tartaglia to divulge the cubic secret to him so that he could include it in his book. After initially refusing, Tartaglia finally agreed and divulged the solution to the Del Ferro-type cubic. However, he made Cardano swear a solemn oath never to publish the method before he, Tartaglia, would have done so himself. It was not to be. Upon hearing (in 1542) that Del Ferro had solved $x^3+bx=c$ 20 years before Tartaglia, Cardano felt free to include the solution method in his book *Ars Magna* (1545). This was the first systematic treatment of algebra in Latin, and it had an enormous impact. In the book, Cardano took up the insights of Del Ferro and Tartaglia and developed a systematic treatment of all 13 forms of the cubic. It should be noted, though, that Cardano's understanding of the cubic was limited. He did not know how to handle the roots of negative numbers or, more generally, how to manage the involvement of complex numbers. Although Tartaglia was given full credit in the book (in three places), he was not amused. He attacked Cardano vehemently in his own book *Inventioni*, which appeared in 1546, and an angry exchange of letters ensued between Tartaglia and Ferrari, an assistant of Cardano. In the end, a mathematical duel was arranged between the two (in 1548). Ferrari was declared the winner.

After this historical introduction, I will give a short overview of the systematic development of approaches to solving the cubic equation from 1500 onward. Since the initial hesitant steps by Del Ferro, Tartaglia, and Cardano, many more developments have taken place to clarify the equation. Just the bare minimum of formulas will be mentioned, just enough to illustrate my main point. That point is, in a nutshell, that there is not just one method of solving the cubic, leading to one solution formula, that is implemented in one specific way. On the contrary, *various* solution methods, *various* solution formulas, and *various* ways of implementation exist, which also are interconnected. For the purpose of exposition, I will mainly rely on Ludwig Matthiessen, who wrote a thorough, 1000-page long treatise on historical developments in algebra from ancient times onward: *Grundzüge der antiken und modernen Algebra der litteralen Gleichungen* (Matthiessen, 1878).

The complete cubic equation is of the form

$$x^3 + ax^2 + bx + c = 0.$$

By putting $x=y-a/3$ (called *emendatio*) the squared term can be eliminated to obtain the reduced cubic equation (in canonical form):

$$x^3 + px + q = 0.$$

The classic solution formula for the canonical equation can be expressed as follows. If we define

$$(u,v) = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}},$$

then the three roots are

$$x_1 = u + v,$$

$$x_2 = J_1u + J_2v,$$

$$x_3 = J_2u + J_1v,$$

with J_1 and J_2 the complex cubic roots of unity:

$$J_{1,2} = -\frac{1}{2} \pm \frac{1}{2}\, i\sqrt{3}.$$

If u and v are real ($q^2/4+p^3/27>0$), x_1 is real, while the other two roots are complex numbers. If u and v are complex numbers ($q^2/4+p^3/27\leq 0$), all roots are real.

The complete cubic equation as given above has more complicated roots. If we define

$$u,v = \sqrt[3]{-\frac{1}{2}(2a^3 - 9ab + 27c) \pm \frac{1}{2}\sqrt{(2a^3 - 9ab + 27c)^2 - 4(a^2 - 3b)^3}},$$

then

$$x_1 = -\frac{1}{3}a + u + v,$$

$$x_2 = -\frac{1}{3}a + J_1u + J_2v,$$

$$x_3 = -\frac{1}{3}a + J_2u + J_1v.$$

Under similar conditions, as in the canonical case (see above), the roots are either all real, or one is real and the other two are complex.

The solution formula for the cubic equation — either complete or in canonical form — can be proven in many ways. *Substitution methods* were the first to be developed: x is replaced by some suitable function of other variables (or, more generally, a function of x and appropriate auxiliary variables is introduced into the cubic). Through this substitution the original equation transforms into another equation that can be solved more easily. Many substitution methods exist. Some famous examples for the canonical case are the following.

First, of course, the "inventors" of the canonical formula should be mentioned. Tartaglia and Cardano basically put

$$x = u - v$$

and solved a quadratic in u^3 (Matthiessen, 1878, § 127). Lacroix (in 1799) modified the former method, using the substitution

$$x = y + z$$

and solving a quadratic function of y^3 (Matthiessen, 1878, § 130). Third, Vieta (in 1615) and Hudde (in 1658) both made the clever substitution

$$x = \frac{\frac{p}{3} - z^2}{z}$$

to obtain a quadratic of z^3 (Matthiessen, 1878, § 128–129].

On a finer scale, one may differentiate further and distinguish *different kinds of substitution methods*. In the case of the cubic, Matthiessen distinguishes three classes, which differ as to the actual function of x and auxiliary variables that are being used: a linear function of *x*, a quadratic function of *x*, and a cubic function of *x* (Matthiessen, 1878, § 80, a, b, c).

The second main means to prove the solution formulas is by *combination methods*: combinations of the unknown roots are formed and manipulated ("root types"). These methods were only developed two centuries later by mathematicians like Lagrange, Laplace, and Vandermonde. Laplace (in 1795), for example, could obtain the solution formula for the canonical cubic by putting

$$z = lx_1 + mx_2 + nx_3,$$

with x_1, x_2, and x_3 being the unknown roots of the cubic equation and choosing *l*, *m*, and *n* in such a way that a simple quadratic in z^3 could be obtained (Matthiessen, 1878, § 293).

Also here, one may differentiate further by distinguishing *several kinds of root types*: *z* as a linear combination of x_i, of $x_i x_j$, or of $x_i^2 x_j$; z as a linear combination of $x_i x_j$, divided by a linear combination of x_i; z as a linear combination of x_i^2, divided by a linear combination of x_i; etc., for $\{i,j\}=\{1,2,3\}$ (Matthiessen, 1878, § 284, b).

So far, we have only been talking about proving the solution formulas as given above. It should be pointed out, however, that *other* formulas exist as well. There is not one standard expression but many expressions for the roots of the cubic equation. To be sure, these formulas are all equivalent in the end; but this is not *obviously* so. It would take hard work to prove their mathematical equivalence.

I will mention just one such other solution to the complete cubic. It can be obtained by a particular substitution method: the method of "false substitution" or "false position" (*regula falsorum*). This method, by the way, was first pioneered by Arab scholars for the linear equation

$$ax + b = 0.$$

The idea itself can be used fruitfully for higher order equations as well, as Matthiessen pointed out (in 1870). Consider the complete cubic equation as given above (cf. Matthiessen, 1878, § 161):

$$x^3 + ax^2 + bx + c = 0.$$

Choose two substitutions z_1 and z_2 that are precisely the roots of the quadratic equation in *z*

$$(a^2 - 3b)z^2 + (ab - 9c)z + (b^2 - 3ac) = 0.$$

If we define

$$z_1^3 + az_1^2 + bz_1 + c = \varphi_1,$$

$$z_2^3 + az_2^2 + bz_2 + c = \varphi_2,$$

then it can be derived that

$$\frac{x - z_1}{x - z_2} = \sqrt[3]{1}\,\frac{\sqrt[3]{\varphi_1}}{\sqrt[3]{\varphi_2}},$$

which can also be expressed as the ratio of the "failures of substitution" is equal to the ratio of the "failures of the equation." From this formula it follows that the three roots are

$$x_1 = \frac{z_2\sqrt[3]{\varphi_1} - z_1\sqrt[3]{\varphi_2}}{\sqrt[3]{\varphi_1} - \sqrt[3]{\varphi_2}},$$

$$x_2 = \frac{z_2\sqrt[3]{\varphi_1} - J_1 z_1\sqrt[3]{\varphi_2}}{\sqrt[3]{\varphi_1} - J_1\sqrt[3]{\varphi_2}},$$

$$x_3 = \frac{z_2\sqrt[3]{\varphi_1} - J_2 z_1\sqrt[3]{\varphi_2}}{\sqrt[3]{\varphi_1} - J_2\sqrt[3]{\varphi_2}}.$$

These three roots are "not obviously equal" to the Tartaglia–Cardano-type roots as given above. Note that here is a case of a substitution method that yields a substantially different solution formula. Just as different kinds of method may yield the same formula, sometimes methods of the same kind lead to different formulations of the roots.

As an aside, it should be pointed out that cubic roots also can be constructed *geometrically*. Such geometrical construction methods, for both the canonical and the complete cubic equation, go back to Arab mathematicians of the 9th, 10^{th}, and 11th centuries. Omar ben Ibrahim, also known as Omar Alkhayyami, wrote a textbook about these and other achievements (circa 1080) (cf. Matthiessen, 1878, § 365–366). The geometrical solution, therefore, antedates the algebraic solution by centuries. These methods will further be left aside, as they do not yield numerical solutions.

Finally, besides proving formulas in several ways, they can also be *implemented* in several ways. Of course, the basic implementation that comes to mind is just carrying out the formulas as given: adding, subtracting, multiplying, dividing, and taking roots as the formula goes. However, alternative implementations may calculate the result more efficiently. Consider the case of the complete cubic. If tables with squares of numbers are available, as a minor variation one may readily substitute squares for products:

$$ab = \frac{(a + b)^2 - (a - b)^2}{4}.$$

A more fruitful approach is a *goniometric* one: by introducing suitable goniometric functions, the algebraic formula can be carried out more easily. For the cubic equation, many such transformations exist. Take for example Eytelwein, who in 1824 proposed the following substitution (Matthiessen, 1878, § 339):

$$\frac{4p^3}{27q^2} = tg^2\alpha.$$

By further putting

$$\sqrt[3]{tg\,\frac{\alpha}{2}} = tg\beta$$

one obtains the three roots of the canonical cubic equation in goniometric form

$$x_1 = \mp 2\sqrt{\frac{p}{3}}\,\cot 2\beta,$$

$$x_{2,3} = \pm\sqrt{\frac{p}{3}}\left(\cot 2\beta \pm \frac{i\sqrt{3}}{\sin 2\beta}\right).$$

By looking up goniometric tables, α and β can be calculated.

This systematic treatment of how to solve the cubic equation can be summarized as follows. It can be solved algebraically. To derive the roots, many solution methods can be used: substitution or combination, for example. Sometimes, these yield alternative solution formulas. Although equivalent, these are not trivially and obviously equal to the original solution. Finally, solutions can be implemented otherwise, either exactly or approximately (for example, goniometric implementations). Compare the following figure (in which M_i and M_j are methods of proof, F_i and F_j are solution formulas, and $\mathscr{F}_i$ and $\mathscr{F}_j$ are alternative implementations):

$$\begin{tikzcd}
\boldsymbol{M_i} \arrow[d] \arrow[dr] & \boldsymbol{M_j} \arrow[d] \arrow[dl] \\
\boldsymbol{F_i} \arrow[d] & \boldsymbol{F_j} \arrow[d] \\
\boldsymbol{\mathscr{F}_i} & \boldsymbol{\mathscr{F}_j}
\end{tikzcd}$$

3. Thought experiment: patenting the solution to the cubic equation

At the beginning of the period treated above, around 1500, mathematics was still a unified branch of science: pure and applied mathematics were not yet clearly distinguished. All mathematicians had firm roots in practice, and, however pure their mathematics might become, it all sprang from that. Praxis connected with mathematics was, first and foremost, commercial arithmetics, a field very much in vogue at the time. It dealt with all kinds of applications in the context of trade. Also, the use of mathematics (especially trigonometry) in ballistic science was just starting to develop (cf. Smith, 1958, vol. I, Chapter VIII).

Mathematicians of the time were mostly very secretive about their inventions. They were unlike modern academic mathematicians who routinely publish their findings; they much more resembled the industrial mathematicians of our age, who also as a rule keep their methods a secret. Two factors contributed to this. In the first place, mathematicians were mainly making a living from teaching, either privately or at universities. Compare the episode above: Del Ferro, Fiore, Tartaglia, and Cardano all taught mathematics at one time or another. Often, these mathematicians indulged in public contests with each other in which they posed mathematical problems to each other to solve. The winners received a prize; more importantly, they also secured their fame as outstanding mathematicians. This enabled them to attract even more students or simply to keep their academic appointment. As it was, the results of these contests, that is the solutions to *particular* problems, were made public. But the underlying *general* methods that had been used were kept a secret. Understandably, participants jealously guarded the secrets of their trade; these were regarded as economic assets.

Second, secrecy also was called for in applying their skills to practical problems. This has to do with the fact that mathematical methods spread out easily, more easily in fact than products of any other branch of science. Mathematicians were anxious to become "mathematical advisers" for all kinds of practical problems, and, all the while, they could protect their secrets. However, training pupils in their trade or producing lists of parameters and corresponding solutions (such as for ballistic purposes) would only jeopardize their monopoly. Pupils may defect, and tables can be copied. Mathematical knowledge, once in the open in one form or another, is easily accessed.

So, secrecy was the norm, and publishing (mathematical) texts in print, although it had just become possible, was still the exception. As a result, only a small portion of all mathematical inventions was actually published. This secrecy, of course, greatly held back the spread of important advances in mathematics. Compare the episode described above: Del Ferro, Fiore, and Tartaglia all too readily kept their solutions to the cubic equation a secret. It was only Cardano's exceptional step to publish a book that finally opened up the secret. In the ages to follow, of course, gradually the public conception of science developed in a full-fledged form. In it, publishing one's results has become the norm (cf. Robert Merton's norm of "communism," in the sense of common ownership of goods).

Against this background, let us conduct a thought experiment, starting at the beginning of the 16th century. As explained above, mathematical applications for mercantile and ballistic purposes warrant commercial exploitation, but mathematicians are hesitant to invest in such ventures as their methods are copied too easily. Let us suppose, then, that in order to foster the "useful arts," some kind of patent system is introduced in Europe. The secretive mathematicians of the time, anxious to exploit their discoveries for themselves, would be able to ask for patent protection of their inventions. This is mere fiction, of course. Nevertheless, algebra is a precisely demarcated field, and its development can be studied over a period of centuries. Therefore, I hope some lessons can be drawn from this fictional exercise.

This system would depart from the same basic logic as the currently employed system: the patent-holder gets all rights of exploitation for about 2 decades in exchange for his/her revealing the invention involved. As a whole, this is intended to foster the development of new mathematical inventions and ideas and to get them into the public domain. Actually, some kind of patent system already existed in Europe from the 14th century onward. It originated in Italy, and from there it spread over Europe (cf. David, 1993). However, that system had another objective than that of the current system. A country issuing a patent to foreign artisans hoped that these artisans would immigrate and apply their technologies locally. Inventions could not qualify for a letter patent — let alone mathematical inventions.

In the following analysis of this imaginary patent system for mathematical inventions, three assumptions are made throughout. First, let me say a word about the context of applications. Nowadays, any patent application has to specify the specific commercial uses that are envisaged. In the context of algebra, this would imply, say, that protection is asked for mercantile or artillery applications. In practice, one can formulate specific applications so broadly, that de facto *all* applications are covered (cf. the Karmarkar patent, as discussed above). Therefore, this whole matter of context is rather trivial, and henceforth I will assume, that my imaginary patents routinely cover all commercial uses. Second, it will be assumed that a patent is granted for commercial exploitation *only*. That is, the American *experimental use* exception to patent infringement fully applies (cf. discussion in Chisum, 1986, pp. 1017–1019, especially note 199); similar exceptions exist in European patent law. Finally, I assume that patent disclosure is (almost) perfect: the full text of the application is sufficiently detailed in order to enable informed others to actually reproduce and use the invention, and this text becomes public soon after filing a patent (say, 18 months afterward, which is now customary in most countries outside the United States). This assumption, of course, has not always been fulfilled in practice (cf. Kitch, 1977, pp. 287–288).

The last two assumptions of experimental use exception and perfect patent disclosure, taken together, have important implications. Individuals can and may use patented mathematical algorithms for *private* use. So, a patent will not preclude other individuals from thinking such mathematical thoughts for themselves or will not keep them from use in everyday life (as Allen Newell, cited in Samuelson, 1990, pp. 1109–1110, seems to fear). Moreover, and more importantly, patented algorithms can and may be freely applied inside academia for *scientific* purposes; that is, the further development of mathematics as a science is not hampered. In fact, if anything, it is furthered, because applications for mathematical patents perforce become public.

What effects would such an imaginary patenting system have? From a positive point of view, it is to be expected that all kinds of mathematical inventions, hitherto kept secret, would be taken out of drawers and submitted for patenting in order to be able to apply them with profit. The kind of protection against spillover offered by a patent would induce commercialization (cf. Mazzoleni & Nelson, 1998: Theory II, “induce commercialization theory”). Moreover, it is to be expected, that spillover protection would also stimulate the search for new inventions proper (cf. Mazzoleni & Nelson, 1998: Theory I, “invention motivation theory”). A contrasting view, however, would argue that granting protection on the building blocks of (mathematical) science would hamper scientific progress. One American mathematical society, for example, recently has declared to be opposed to the patenting of algorithms, out of fear that research on algorithms will be slowed down (cf. Samuelson, 1993, p. 303, note 49). These opposite views can be investigated fruitfully, I would argue, by analyzing precisely what elements of (cubic) algebra may be endowed with patent protection. Depending on the kind of regime that obtains in this respect, either more positive or more negative effects will predominate.

What meanings, then, can be given to the term “protecting mathematical inventions?” What is to be understood as a mathematical “process” or “product?” As argued above, the contents of (cubic) algebra can be specified as formulas, proofs, and implementations. These elements may enjoy protection in various combinations. At the lowest level, specific implementations of mathematical solutions ($\mathscr{F}_i$) may become protected. Take, for example, the more efficient calculation of cubic roots using goniometry as proposed by Eytelwein, discussed above. Such a regime, I would argue, corresponds to what is actually allowed in the current U.S. patent system. Compare the Karmarkar patent, which is just a more efficient method of linear programming. As this kind of regime does not seem to go beyond current practice, it is not very useful for our present purposes and will further be left aside.

Next, looking at the highest level, one may consider the protection of solution methods (M_i) in general. Whatever mathematical results a method may achieve, all practical uses of these results are covered. More precisely, such a regime protects the commercialization of results of a method in general, in all possible mathematical fields of application. Say it protects the method of substitution in general, usable in algebraic equations, differential equations, and so on. This must be considered an odd kind of protection. What counts in mathematics is precisely the *specifics*: what specific kind of substitution accomplishes the trick of solving a specified equation? It is not very interesting to hear that substitution methods in general may solve algebraic equations in general, or even that they may be helpful in solving differential equations in general. It *is* interesting to hear, however, that a specific clever substitution (say the one proposed by Vieta and Hudde, cf. above) actually solves the specific canonical cubic equation. Furthermore, granting such a broad protection is also odd, as the original inventor presumably has successfully applied the invention once in a concrete case (say the cubic) and from then on automatically reaps the rewards of all other creative applications of the method in that same field and in others. The effort is in the specifics, but only the inventor's one-time effort, not those of his/her followers, is rewarded. Finally, given the fact that not too many basically different methods exist, it would seem to become possible to cover almost the whole of mathematics with a handful of patents. Clearly, patent scope is too broad in this kind of regime. These drawbacks are not eliminated, only slightly reduced, if the patenting of methods is confined to specific fields of application, say, only to algebraic equations.

The only way to defend such a broad regime at all would be by interpreting the mathematical method involved as a kind of *prospect*, opening up a whole series of follow-up inventions in that same and other mathematical areas (along the lines of the "prospect theory," as originally proposed by Kitch, 1977; cf. Mazzoleni & Nelson, 1998, Theory IV, "exploration control theory"). However, in a field as abstract as mathematics I see no need for the orchestrated coordination of such efforts. For the purpose of applying, say, combination-type methods, as discovered in the case of the cubic, to other unsolved mathematical problems, I doubt that any coordinated master plan would be helpful or necessary. So, I would argue, a blanket protection of methods is too abstract and too broad an option. Only methods with parameters specified, as far as these solve specified mathematical problems, and methods, moreover, that demonstrably work, may count as statutory matter.

This leaves only two useful possibilities for a patenting regime in mathematics. In decreasing degree of protection these are:

1. Protection of formulas tout court (F_i); and
2. Protection of formulas-cum-proof (F_i-cum-M_i).

Apart from protecting alternative implementations of a solution formula, the two regimes protect solution formulas either in a wide sense or in a restricted sense. In the wide sense, protection covers the solution formula tout court, whatever the method by which the result has been achieved (Regime 1). In the case of the cubic, the Tartaglia–Cardano formula for the cubic roots would enjoy blank protection. Note that a patent application will have to supply (at least) one concrete proof; after a patent has been granted, *all* possible methods of proof have in fact been monopolized. When protection in a restricted sense applies, a solution formula is only protected in *combination* with a proof that can stand the test of mathematical scrutiny (Regime 2). Say that Tartaglia–Cardano-type roots become protected on the basis of a substitution type of proof. Again, a patent application will have to provide one sort of proof; this time, however, once the patent has been granted, *only* that one method of proof has been monopolized.

Apart from obviously stimulating the "useful arts" (cf. "induce commercialization theory"), what would have been the effects on the development of algebra as a science (cf. "invention motivation theory") if any of such imaginary regimes had actually come into effect?

In the first kind of regime, a mathematical formula (F_i) tout court becomes the property of its inventor. This might seem like a recipe for stifling further innovation: no room for escape from this monopoly seems available. Such room is an absolute requirement for a sound patent system. Actually, this condition also seems to be met in the current system: according to research and development managers, the most important limitation on the effectiveness of patents is precisely the ability of competitors to invent around the patented invention (Levin, Klevorick, Nelson & Winter, 1987, p. 803). On closer inspection, however, such room also exists in this first kind of hypothetical regime for the patenting of mathematical algorithms. This is so, while in many mathematical problems there are other formulas to express solutions. That is, there are other expressions looking very dissimilar that solve the problem nevertheless. The search for such alternative solutions offers a way out for ambitious mathematicians eager to invent around, as it were, the patented formula. Employing the same method as used in the patent (M_i)—or, of course, another method (M_j)—an alternative formula (F_j) has to be found. Compare the algebra of the cubic equation. Suppose that the formula for the roots of the Tartaglia–Cardano-type solution has become patented tout court. This still leaves room for a way around: develop the age-old method of *regula falsorum* as a solution device for the cubic equation. The formula obtained thus is sufficiently different from the patented one to create room for independent use and patenting.

On the other hand, it has to be taken into account that the disclosure of information in patent documents may not facilitate inventing around the patent too much. In that case, inventors will prefer secrecy over patenting, and the whole system becomes rather useless (cf. Levin et al., 1987, p. 805). Looking at Regime 1, it would seem that the way around, inventing alternative solutions, still requires hard work so that patent disclosure does not amount to a giveaway.

What would have been the effect on the development of algebra as a science? It would have had a beneficial effect, I presume. Del Ferro would have applied for a patent right away, around 1515; if not, Tartaglia would have done so 20 years later. As a result, decades earlier the cubic would have been solved and available for public inspection. Moreover, if the licensing cost had been substantial or even prohibitive, the search for substantially different solution formulas for the cubic would have been speeded up. Application of say the *regula falsorum* might well have happened earlier than 1870. Finally, the early disclosure of the method of substitution might have stimulated earlier applications of this method to higher order algebraic equations (degree 4 and higher). Or, putting it more precisely, actual public disclosure (in patent applications or otherwise) of such efforts would have been furthered.

Second, a restricted regime that protects only formulas-cum-proof (Regime 2) may apply. As in Regime 1, this does not create an absolute monopoly: an independent route to commercial applications of the cubic exists. In comparison, there are now even two routes instead of one. As before, if someone else comes up with a similar procedure that nevertheless yields a *solution* that is substantially dissimilar from the patented formula ($M_i \rightarrow F_j$ instead of $M_i \rightarrow F_i$), she/he will be hailed as an inventor and allowed to use it in any commercial way she/he chooses. But there is also another novel escape route. Someone else may circumvent the patent by coming up with a substantially different *method* to reach the same conclusion ($M_j \rightarrow F_i$ instead of $M_i \rightarrow F_i$). Other mathematicians can effectively invent around such a patent, as it were, by inventing another type of formula, or inventing another type of proof; or, of course, both ($M_j \rightarrow F_j$ instead of $M_i \rightarrow F_i$). Compare, again, the algebra of the cubic. A patent on Tartaglia–Cardano roots, as obtained by substitution, can be circumvented by (re)inventing the *regula falsorum*, which, although being a method of substitution, nevertheless yields something new: another solution formula for the cubic roots that is equivalent to the Tartaglia–Cardano solution, but is not obviously so. But additionally, such a patent can be circumvented by inventing a combination method to arrive at the same type of Tartaglia–Cardano cubic roots. Neither escape route, by the way, is discovered more easily by reading the text of the patent application concerning cracking the cubic by substitution. So, here also patenting is to be preferred over secrecy.

As a whole, this second type of regime would presumably also have been beneficial for the development of algebra. As in the first regime, the original inventors before Cardano would have rushed to obtain a patent on their root expressions as obtained by substitution, thereby speeding up public exposure. Also, as before, patenting would have stimulated the development of alternative solution formulas and the application of substitution methods to higher order equations. But there is one more important effect: the quest for alternative *solution methods* also would have been put on the (commercial) agenda straight away and might have yielded combination-type solutions at an earlier time than only two centuries later, when Vandermonde (in 1772), Laplace (in 1795), and Lagrange (in 1808) pioneered these methods. Also, the application of these combination methods to the solution of higher order equations would have been furthered.

Comparing both regimes, it would seem that the protection of formula-cum-proof would have advanced algebra the most: the second regime has the extra benefit of stimulating the search for new methods of proof. Alternative solutions are of some, mostly practical, importance, but they do not constitute a novel systematic approach to algebraic problems in general. The development of alternative approaches to proof is of a systematic nature and yields much more for algebra in general. The quest for solutions to higher order equations than the cubic has benefited a lot from it. This goes for quartic equations (of degree 4), which have historically been solved along the same lines (i.e., substitution methods and combination methods) as the cubic equation. But broadening the arsenal of proof methods has especially enhanced the understanding of the complexities of equations of degree 5 and higher. It turned out that these can be solved only in specific cases. In general, these do not allow for algebraic solutions in terms of equation parameters; they are "unsolvable by radicals" (theorem of Abel-Ruffini, developed around 1800).

So it would seem important not to lose sight of the distinction between the two proposed regimes. Both protect algorithms and formulas, but there is a fine distinction between the two. Regime 1 protects the formula as a whole (that is, it protects the formula *whatever* is the type of proof that may support it), while Regime 2 only protects a *combination* of formula plus a method to get there. As a result, only Regime 2 stimulates the development of alternative solution strategies. If patent law, then, is about protecting ideas, it would seem that mathematical formulas and algorithms should be allowed protection *with* a proof that supports them. Such a protection strategy effectively conveys the message that a mathematical idea is not to be equated with a formula only, but with a formula-cum-proof. A solution is nothing without corresponding proof.

4. Conclusion

What lessons can be drawn from this thought experiment? If the results are any guide to present-day developments, they would argue for a *blanket acceptance* of mathematical formulas and algorithms as statutory matter for patent application. Let us finally acknowledge that mathematics belongs not only to the "liberal arts" but also to the "useful arts" (cf. *Gottschalk v. Benson*, 1972, discussed in Samuelson, 1990, pp. 1055 and 1112). Let us abolish the "pure-number condition"; even if it is all mathematical abstraction, this is acceptable. There is no need to specify links with "physical elements" or "process steps," with "pre-computer" or "post-computer process activity." Even if nothing of the sort is specified, the claim is acceptable. Let us not bother that the claim "wholly preempts" the use of the algorithm; let all possible commercial applications effectively become protected. It should be stressed that noncommercial, in particular *academic*, use is to remain possible (experimental use exception). Actually, it is the development of both (pure) science and (applied) technology that the system is designed to foster. Such unconditional acceptance, however, should be carefully formulated as to what is protected and what not. While the protection of formulas tout court is acceptable in societal terms, the best results seem to be obtained when algorithms and formulas are only patentable in combination with a supporting proof. That arrangement keeps avenues open for inventive mathematicians to develop alternative methods of proof.

Of course, a word of caution is in order. A thorough examination of present-day mathematics would be required to be able to judge whether these conclusions, based on an examination of early algebra, can simply be copied in the 20th century. In particular, those fields of mathematics should be examined in which the distinction between pure and applied science has almost vanished again (e.g., cryptography, code theory, and the like). These fields nowadays lean heavily on results from abstract branches of mathematics like algebra and number theory. A good object for study would be the famous Rivest-Shamir-Adelman cryptographic patent, which is due to expire soon.

Moreover, problems of implementation and enforcement of such patenting regimes loom large. How to distinguish novel and nonobvious mathematical formulas from non-novel and obvious ones? The same question should be answered as regards the methods of proof. Also, introducing such a regime would take time; how to prevent the patenting process from clogging up in the meantime? And, finally, how to enforce such policies? If Regime 2 obtains, for example, it would be a pressing concern to effectively discriminate between two kinds of users of formulas: those with and those without alternative methods of proof. Only the latter kind infringe on patent law. This requirement possibly could be met by introducing legal clauses that reverse the burden of proof (currently in use for process patents in many European countries): if summoned, a user has to demonstrate that another solution method has been developed.

David Carey (1993) has argued that a fundamental algorithm like the Arabic number system should be “ownable.” Extending this line of reasoning, one could argue the same for the use of matrices in algebra, vector notation, and the like. This position, I would argue, is one bridge too far. I have argued for algebraic solutions to be ownable, on the grounds that at least some room remains to sail past these. With number systems and similar broad-ranging inventions such room seems no longer available. If a monopolist of this kind would refuse licensing, then, given the fact that any compulsory licensing would take time to come into effect, such a monopoly would effectively preempt commercial use by anyone else for a period of 20 years. Competition would come to a halt. Therefore, such a patent regime is not to be recommended.

Acknowledgments

I thank Tomas A. Lipinski, Jaap de Lussanet, and Bill McDaniel for their helpful comments on earlier versions of this article. Also, the comments made by a reviewer of this journal were found to be helpful.

References

Ball, W. W. R. (1960). *A Short Account of the History of Mathematics*. New York: Dover (Original work published 1908).

Bender, D. (1997). Recent developments in software patents. New York: White & Case. Available at: http://cla.org/matsnpubs/memberarticles. Accessed February 18, 1998.

Burton, D. M. (1997). *The History of Mathematics: An Introduction*. New York: McGraw-Hill.

Cardano, G. (1968). *Artis Magnae Sive de Regulis Algebraicis*, trans. *The Great Art or The Rules of Algebra* (T. R. Witmer, Ed. and Trans.). Cambridge, MA.: MIT Press (Original work published 1545).

Carey, D. H. (1993). Should computer programs be ownable? *Metaphilosophy 24*, 76–84.

Chisum, D. S. (1986). The patentability of algorithms. *U Pitt L Rev 47*, 959–1022.

David, P. A. (1993). Intellectual property institutions and the panda’s thumb: patents, copyrights, and trade secrets in economic theory and history. In M. B. Wallerstein, M. E. Mogee, & R. A. Schoen (Eds.), *Global*

Dimensions of Intellectual Property Rights in Science and Technology (pp. 19–61). Washington, D.C.: Office of International Affairs, National Research Council/National Academy Press.

Evans, R. W., & Johnson, M. G. (1994/1995). "The brave new world of software patents", 3 parts. Durham, North Carolina: Moore & Van Allen, PLLC. Available at: http://www.mvalaw.com. Accessed February 10, 1998.

Gable, R. L., & Dengler, A. B. (1995). *In re Warmerdam*: data structures are statutory. Available at: http://www.jurisdiction.com. Accessed February 18, 1998.

Gibby, J. A. (1997). Software patent developments: a programmer's perspective. *Rutgers Comput Technol Law J 23*, 293–355.

Kitch, E. W. (1977). The nature and function of the patent system. *J Law Econ XX*, 265–290.

Kuester, J. R., & Moceyunas, A. K. (1995). Patents for software-related inventions. Atlanta, GA: Thomas, Kayden, Horstemeyer & Risley, LLP. Available at: http://www.kuesterlaw.com. Accessed February 18, 1998.

Levin, R. C., Klevorick, A. K., Nelson, R. R., & Winter, S. G.. (1987). Appropriating the returns from industrial research and development. In M. N. Baily & C. Winston (Eds.), *Brookings Papers on Economic Activity, Vol. 3, Special Issue on Microeconomics* (pp. 783—820). Washington, D.C.: The Brookings Institution.

Matthiessen, L. (1878). *Grundzüge der antiken und modernen Algebra der litteralen Gleichungen*. Leipzig, Germany: Teubner.

Mazzoleni, R., & Nelson, R. R. (1998). The benefits and costs of strong patent protection: a contribution to the current debate. *Research Policy 27*, 273–284.

Patent and Trademark Office. (1996). *Examination Guidelines for Computer-Related Inventions*. Publication No. 950531144-5304-02. Arlington, VA: Patent and Trademark Office.

Restivo, S. (1992). *Mathematics in Society and History: Sociological Inquiries*. Dordrecht, The Netherlands: Kluwer.

Samuelson, P. (1990). *Benson* revisited: the case against patent protection for algorithms and other computer program-related inventions. *Emory L J 39*, 1025–1154.

Samuelson, P. (1993). A case study on computer programs. In M. B. Wallerstein, M. E. Mogee, & R. A. Schoen (Eds.), *Global Dimensions of Intellectual Property Rights in Science and Technology* (pp. 284–318). Washington, D.C.: Office of International Affairs, National Research Council/National Academy Press.

Smith, D. E. (1958). *History of Mathematics*. 2 vols. New York: Dover (Original work published 1923/1925).

Stern, R. H. (1995). Tales from the algorithm war. In D. G. Johnson & H. Nissenbaum (Eds.), *Computers, Ethics & Social Values* (pp. 219–235). Upper Saddle River, NJ: Prentice Hall.