The Mathematics Student ISSN: 0025-5742
Vol. 86, Nos. 1-2, January-June (2017), xx-yy

THE ART OF RESEARCH

M. RAM MURTY

As we all know, there is no simple algorithm for research. There is no recipe for making new discoveries. It is a mysterious and inscrutable process. However, we know that this process has some guiding principles. It is the purpose of this article[1] to discuss these principles in a general way, illustrating them with examples from science and mathematics. Naturally, these examples will have some personal bias.

So let us begin. What exactly is research? In one sentence, it can be said to be the art of asking good questions. In our search for understanding, the SOCRATIC method of questioning is the way.

Let us observe that the word 'Question' has as root word 'Quest'. In our quest for understanding, the method of questioning seems to be the only way. Socrates taught Plato that all ideas must be examined critically and fundamental questions must be asked and pursued in order to gain proper understanding. Buddha instructed his disciple Ananda to question, to reflect deeply. As most of you know, Buddha advocated clear thinking. Socrates was adamant about definitions and in mathematics too, definitions are very important.

This method is not infallible. But it is the only way we have available. Some basic questions seem to defy simple answers. But this doesn't stop us from asking them. Often, the inquiry is a good exercise for the mind, and maybe it is the exercise that is the most important thing rather than complete understanding. Nevertheless, one can enquire into the nature of understanding itself. But then, this would take us into the realm of philosophy. This is not our goal here.

Our goal here is to explore how asking proper questions leads to some form of knowledge and understanding. We must keep in mind that each person brings their own past knowledge and experience to deal with the question. Each one brings their own methodology. Let us take an example

In a room, there are five people: an engineer, a physicist, a mathematician, a philosopher and an accountant. They are asked the simple question: what is 2+2?

* Research partially supported by an NSERC Discovery grant.

[1] This article is based on a public lecture given at the Tata Institute of Fundamental Research in Mumbai, India, some years ago. It is presented here in the hope that students may profit from it so that the ideas can gain a wider circulation

The engineer takes out his calculator and says the answer is 3.99. The physicist runs an experiment and finds the answer is between 3.8 and 4.2. The mathematician says he doesn't know the answer but can show that it exists. The philosopher asks for the meaning of the question. The accountant closes all doors and asks, 'What would you like the answer to be?'

Let us begin with some famous questions. What is life? This is the most basic question. It is related to 'What is consciousness?' and we still don't have a satisfactory answer. What is time? This is one of the most difficult questions to deal with and invariably takes us simultaneously into physics and philosophy. What is space? What is light? These questions have baffled the physicists for centuries and much of modern physics is the outcome of this inquiry. What is a number? This question has led to the development of mathematics. There are other questions that seem unrelated to any of these and are seemingly simple, like 'What is a knot?' Yet, on inquiry, we find it is related to the notion of number and the notion of light, as I will briefly indicate at the end of this article.

As I said, existential questions take us into the realm of philosophy. But there are other perhaps simpler questions one can ask and the inquiry into them quickly leads us to some understanding. So how to ask 'good' questions? What is a 'good' question? It is one that leads us to new discoveries. Below, we will present eight methods of generating good questions and we call this the 'Eight-fold way', to borrow a phrase from Buddhist philosophy.

The simplest method of generating good questions is the survey method. This method consists of two steps. After selecting the topic to survey, we gather all the facts about the topic and then organize them. Arrangement of ideas leads to understanding. The amazing thing is that in this process, what is missing is also revealed. The method quickly leads to fundamental questions.

A good example is given by the discovery of the periodic table. Dimitri Mendeleev organized the existing knowledge of the elements and was surprised to find a periodicity in the properties of the elements. Mendeleev was born in 1834 in Siberia and was the last of 17 children. Those who think that writing graduate level textbooks is not research perhaps should think again! In the process of writing a student text in chemistry, Mendeleev decided to gather all the facts then known about the elements and organize them according to atomic weight. In this way, he was able to predict the existence of new elements. In 1875, six years after Mendeleev published his periodic table, the first of his predicted elements was discovered. This was gallium, which is an essential component of the electronics industry. For example, the liquid crystal displays in digital watches and calculators are based on gallium technology. Shortly after the discovery of gallium, scandium was found. Then, germanium was discovered. These names also suggest the nationalities of the discoverers. The race for finding the missing elements soon became a form of national pride! And the race was on! Every time someone discovered a missing element, they got the Nobel Prize in chemistry! Today more than a century after Mendeelev suggested the periodic table, it was finally complete. It now sits as the presiding deity in all chemistry laboratories.

A similar survey method can be found in the writings of the mathematician David Hilbert. Born in 1862, Hilbert studied under Lindemann (who first proved that π is transcendental) and obtained his doctorate in 1885 from the University of Göttingen. Hilbert's list of 69 doctoral students is quite illustrious and includes Courant, Hecke, Takagi, Weyl and Zermelo.

Hilbert's approach to mathematics has always been that of an organizer of knowledge. He would set out to write a definitive textbook on a specific area of mathematics and invariably would find new and fundamental questions the answers to which led to rudimentary discoveries in mathematics. In 1900, at the International Congress of Mathematicians in Paris, Hilbert organized 23 problems which he considered important in mathematics. Six of these problems deal with the notion of number and have acted as a catalyst in the development of number theory.

- The 7th problem led to the development of transcendental number theory.
- The 8th problem is the Riemann hypothesis that plays a major role in analytic number theory.
- The 9th problem led to the development of reciprocity laws in algebraic number theory.
- The 10th problem led to the development of logic and diophantine set theory.
- The 11th problem led to the theory of quadratic forms and the 12th to class field theory, which began as a part of algebraic number theory and now is expanding into the realm of representation theory.

At the dawn of the 21st century, a similar program was launched. In the year 2000, the Clay Mathematical Institute designated 7 problems of mathematics as millenium problems and is offering a prize of one million dollars U.S. for the solution of any of the following problems:

- 1. $P = NP$;
- 2. The Riemann hypothesis;
- 3. The Birch and Swinnerton-Dyer conjecture;
- 4. The Poincaré conjecture;
- 5. The Hodge conjecture;
- 6. The Navier-Stokes equations;

- 7. The Yang-Mills theory.

As far as I know, only one of these problems has been solved and that is the Poincaré conjecture. In 2003, Grigori Perelman in a series of papers posted on the arxiv [4], [5], [6] settled the Poincaré conjecture but turned down the million dollar prize! In 2006, he was awarded the Fields Medal but again turned it down saying that "I am not interested in money or fame. I don't want to be on display like an animal in a zoo!" [1]

Further details about the Clay problems can be found at www.claymath.org. There the reader will find survey lectures in video format. You will undoubtedly find many more questions that need to be answered from these surveys. Thus, we see that the survey method is a powerful way to generate fundamental research questions.

The next method is the method of observations. Careful observations lead to patterns and patterns lead to the question why? In physics, the famous 1887 experiment of Michelson and Morley to determine the speed of light, first with reference to a stationary frame of reference and next with reference to a moving frame of reference was based on careful observations. They found that the velocity of light is constant and no evidence for the postulated ether. This was revolutionary and led to the special theory of relativity by Albert Einstein.

Another example of the power of observation leading to discovery is the apocryphal story of Archimedes. King Hiro commissioned a goldsmith to make a crown and was wondering if the goldsmith had stolen some gold. So he asked Archimedes to find out without destroying the crown. Archimedes started to ponder this and went to take a bath and noticed that the volume of water displaced was proportional to his weight. Immediately, his mind made the connection. The amount of gold given by the king should displace the same amount of water as the king's crown. If not, the goldsmith had taken some of the gold and replaced it with a baser metal. He was so happy with this discovery that he went running through the streets of Syracuse shouting "Eureka" (I have found it!) and forgot that he was taking a bath! Incidentally, the story is that the goldsmith had indeed cheated the king of some gold!

Often, we are unable to determine what impact our discovery will have. The role of the scientist is simply to investigate and report.

Careful observations lead to the discovery of patterns and consequently to conjectures. Certain conjectures gain prominence and act as powerful inducements for the development of a subject. Fermat's last theorem is a good example.

In 1637, Pierre de Fermat, who was a lawyer by profession, was reading Bachet's translation of the work of Diophantus when he came across the discussion

[1]BBC News, March 24, 2010.

of Pythagorean triples. This led him to wonder if the same works for higher powers and he was led to conjecture that for any $n > 2$, one cannot find three positive integers a, b, c such that

$$a^n + b^n = c^n.$$

Then, he wrote in the margin of the book that he had a wonderful proof of this fact, but the margin was too narrow to contain it!

Let us look more closely at Fermat's marginal note. First, the tome that Fermat was reading had very wide margins and so if Fermat had a proof, it must have been very long! One of my students told me that Fermat must have had a proof since he was a lawyer by profession and lawyers always tell the truth!

As most of you know, Fermat's last theorem was finally solved by a galaxy of mathematicians, culminating in the work of Andrew Wiles[11] in 1996. This proof certainly cannot be the one Fermat may have had in mind since it uses many ideas with which Fermat was unfamiliar with and hadn't been discovered yet. To trace the development of these new ideas, we look at another great mathematician who had the uncanny ability of making powerful and incisive conjectures.

This is Srinivasa Ramanujan, who discovered the importance of the τ-function and isolated it for further study. Ramanujan was born in the city of Erode in present-day Tamil Nadu. He is certainly one of the wonders to come out of the dust of India since he more or less educated himself by reading books and doing problems.

Ramanujan was never averse to making extensive calculations on his slate, since he didn't have much paper. Most of his findings, he would store in his brain. He studied the τ-function defined as follows. $\tau(n)$ is the coefficient of q^n in the infinite product expansion

$$q \prod_{r=1}^{\infty} (1 - q^r)^{24}.$$

He computed (see [7] and the table on the next page) by hand the first 30 values of the τ-function. What he observes is that τ is multiplicative and he makes his three famous conjectures concerning its behaviour.

(1) $\tau(mn) = \tau(m)\tau(n)$ for $(m, n) = 1$.
(2) for p prime, and $a \geqslant 1$, we have $\tau(p^{a+1}) = \tau(p)\tau(p^a) - p^{11}\tau(p^{a-1})$.
(3) $|\tau(p)| \leqslant 2p^{11/2}$, p prime.

The first two of his conjectures were proved by Mordell [2] the year after Ramanujan made the conjecture and third one defied attempts by many celebrated mathematicians, until 1974, when Deligne[1] solved it as a consequence of the Weil conjectures. For this work, Deligne was awarded the Fields medal. As we shall see, these conjectures play a vital role in the final solution of Fermat's last theorem. For more details, the reader is referred to the forthcoming monograph [3].

n	$\tau(n)$	n	$\tau(n)$
1	1	16	987136
2	-24	17	-6905934
3	252	18	2727432
4	-1472	19	10661420
5	4830	20	-7109760
6	-6048	21	-4219488
7	-16744	22	-12830688
8	84480	23	18643272
9	-113643	24	21288960
10	-115920	25	-25499225
11	534612	26	13865712
12	-370944	27	-73279080
13	-577738	28	24647168
14	401856	29	128406630
15	1217160	30	-29211840

The fourth method of generating questions (and perhaps the most difficult one) is by the method of re-interpretation. Why I have listed it here will become apparent by the end of the article. This method tries to examine what is known from a new vantage point. An excellent example is given by gravitation.

As most of you know, Isaac Newton first formulated the mathematical theory of universal gravitation. However, much of his theory relied on careful observations that Tycho Brahe and Johannes Kepler made concerning planetary orbits.

For Isaac Newton, gravity is a force and he was able to formulate the inverse square law $F = Gm_1m_2/r^2$ familiar to all of us from high school. On the other hand, for Albert Einstein, gravity is curvature of space.

So let us see how Einstein re-interpreted Newton's theory of gravitation. The surface of the universe is a 3-dimensional manifold. A sun or a planet kind of sits on this surface and consequently distorts the space around it depending on how massive it it. This view has serious implications to the behaviour of light. Thus, one of the consequences is the study of light in such gravitational fields. Light, as it travels on this surface must be therefore influenced by the distortions of space caused by massive gravitational fields and so it was predicted that such a phenomenon can probably be observed during a total eclipse of the sun. So in 1919, scientists were able to verify this phenomenon.

This was one spectacular victory for the theory of relativity. Its mathematical predictions were verified by numerous experiments. Perhaps the most spectacular illustration of this bending of light was the discovery in 1979 of a twin quasar.

It was long predicted that if a massive galaxy was in the line of sight between us and a quasar (quasi-stellar object of the size of our solar system) we would see a "double" and sure enough, this was verified in 1979 and this bending of light phenomenon is now a powerful tool in astrophysics.

Another anomaly resolved by relativity that Newton's theory could not explain is the orbit of Mercury, which was noticed to be not a perfect ellipse. It doesn't quite close upon itself and is called the precession of the perihelion of Mercury. Einstein's theory of relativity could explain this coming from the gravitational field curvature, since Mercury was closest to the sun, and so would feel this effect more than the other planets. This too has now been verified.

Perhaps the most powerful prediction of relativity theory is the existence of black holes, and it was in 1928 that Subramanyan Chandrasekhar worked out this consequence as a graduate student. As we all know, we see objects because light is reflected off of them. When a star dies, it can do one of three things: it can become very cold and become what is called a white dwarf; or it could explode and be a nova, or it can collapse into itself and become a black hole. All of these discoveries were possible only by re-interpreting gravity as curvature.

Let us look at an example of the method of re-interpretation in mathematics. Everyone is familiar with the unique factorization theorem. This says that every natural number can be written as a product of prime numbers uniquely. Euler reformulated this fact in an analytic fashion by introducing the zeta function. He did this by considering the Dirichlet series:

$$\sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Since every natural number can be written as a product of prime numbers uniquely, this series can be written as an infinite product over prime numbers:

$$\prod_p \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Euler gave an analytic proof of the infinitude of primes by noting that both sides converge absolutely for $\Re(s) > 1$ and when we take the limit as $s \to 1^+$, the series diverges and so the product must also diverge, showing the infinitude of primes.

However, it was Riemann who stressed that the zeta function must be studied as a function of a complex variable so that we can gain a better understanding of the distribution of prime numbers. As we shall see, this reformulation of the unique factorization theory re-emerges as a theory of Euler products in the famous Langlands program.

Yet another example of re-interpretation occurs in the work of Dedekind in algebraic number theory. Following some early work of Kummer, it was clear that the unique factorization theorem did not generally extend to the rings of integers of algebraic number fields. Dedekind realized that one needed to replace the notion of a number by the notion of an ideal. He was led to this idea by re-interpreting divisibility. A natural number d divides n if and only if

$$d\mathbb{Z} \supseteq n\mathbb{Z}.$$

"To contain is to divide" became the aphorism for Dedekind's development of algebraic number theory. This re-interpretation transformed number theory and propelled it to major advances in the 19th and 20th centuries.

Another dynamic method for research is the method of analogy. When two theories are analogous, or exhibit some similarities, we try to see if ideas in one theory have analogous counterparts. For instance, the zeta function, and the Ramanujan's zeta function exhibit similarities in that they both have Euler products and functional equations. This analogy was first pushed by Erich Hecke in his study of the theory of modular forms. It also signalled the beginning of a general theory of L-functions and connected representation theory with number theory in a fundamental way. Building on the work of Harish-Chandra, Langlands showed how one can attach L-functions to representations of adele groups. This is the foundation for the Langlands program.

Another profound example of the method of analogy from physics is the Doppler effect. When a train approaches you, the pitch of sound is high and as it moves away, the pitch gets lower. This behaviour with sound waves was extended to light waves by Doppler and used to explain the red shift of stars. When the stars are approaching us, there is a red shift in their spectra and when they are moving away from us, there is a blue shift. This discovery was fundamental in explaining the expansion of the universe.

In a much more down-to-earth application of the Doppler effect, we see that police radar really makes use of the Doppler effect to record the speed of cars.

The strength of analogy is best illustrated in mathematics by the discovery of arithmetic of function fields over finite fields. Hilbert and others already noticed there was an analogy between complex function theory and algebraic number theory. But at the dawn of the 20th century, beginning with the doctoral work of Emil Artin, a new kind of zeta function was discovered which showed structural similarity to the Riemann zeta function but was much simpler to study. Artin conjectured the analog of the Riemann hypothesis for his zeta function and this was proved later by Hasse. But these reflections led Weil to study the zeta functions attached to curves and show the Riemann hypothesis held for these functions as well. Finally, in his epochal paper of 1949 [8], he formulated what became known as the Weil conjectures and these were settled by Deligne [1] in 1974, a part of which led to the proof of the Ramanujan conjecture. In his reflective essay [9], Weil records how he was led to his conjectures. "In 1947, in Chicago, I felt bored and depressed, and, not knowing what to do, I started reading Gauss's two memoirs on biquadratic residues, which I have never read before. The first one deals with the number of solutions of $ax^4 - by^4 = 1$ over finite fields and the second one with $ax^3 - by^3 = 1$. Then I noticed similar principles can be applied to all equations of the form $ax^m + by^n + cz^r + \cdots = 0$ and this implies the truth of the so-called Riemann hypothesis for diagonal equations."

The Rosetta stone was discovered in 1799 and inscribed in the stone were three scripts: hieroglyphics, demotic and ancient Greek. Since scholars knew ancient Greek, they could decipher the other two scripts. It was in this way, the Egyptian hieroglyphs were decoded. Weil makes the analogy to the Rosetta stone when he compares the analogy between the number field, the function field over the finite field case and the complex function theoretical frame with its rich legacy of algebraic topology. The fascinating account is recorded in [10].

The method of transfer is to transfer an idea from one area of study to another. Again, a good example is again of the Doppler effect used in weather prediction. Microwaves are bounced off clouds to see if there are particles there that will cause precipitation and if so, how fast these clouds will be approaching us. As we all know, this is not a fool-proof method but it is approximately true and is a good illustration of the principle of transfer.

A seventh method is induction. This is essentially the method of generalization. Here is a simple example of how one uses the method of induction.

$$1^3 + 2^3 + \cdots n^3 = (1 + 2 + \cdots n)^2 .$$

A more sophisticated example is from the theory of L-functions alluded to earlier. $GL(1)$ and $GL(2)$ are two layers of a larger hierarchy. The Langlands program was largely suggested by induction.

The converse method of generating questions is simple enough. Whenever A implies B, we can ask if B implies A. This is called the converse question. A good example occurs in physics in the discovery of electromagnetism.

Around 1820, Oersted performed a historic experiment to show that an electric current creates a magnetic field. It was only a question of time before some one asked if the converse is true? That someone was Michael Faraday. Shortly after, he showed by experiment that the converse was true. A magnetic field creates an electric current.

The story is that when Faraday gave a public lecture demonstrating electromagnetic induction, the prime minister asked him of what use is it. Faraday responded by saying, "I don't know, but I am sure that someday you will figure out a way to tax it!" And he was right!

The converse method was also fundamental in the resolution of Fermat's last theorem. We have seen that the Riemann zeta function and the Ramanujan zeta functions have similar properties. We also learned that Langlands constructs many more zeta functions from automorphic representations. The question of whether all such objects arise from automorphic representations is called converse theory in the Langlands program. Langlands proved a 2-dimensional special case of a prediction of this theory.

This was the starting point for the proof by Wiles of Fermat's last theorem. To compress three centuries of history is difficult. However many mathematicians played a vital role in the genesis of the solution: Fermat, Euler, Kummer, Riemann, Ramanujan, Hecke, Rankin, Selberg, Taniyama, Shimura, Weil, Iwasawa, Frey, Serre, Mazur, Ribet, Langlands, Taylor and Wiles. Inspired by some earlier work of Hasse, Taniyama predicted that L-functions attached to elliptic curves come from automorphic representations. This was made a bit more precise by Shimura and Weil. Then in 1985, Frey (and independently Hellegouarch), noticed that such a conjecture may imply Fermat's last theorem. This connection was then made more precise in some fundamental conjectures of Serre and Mazur, and then Ribet proved a special case of these conjectures. Ribet then showed that Taniyama's conjecture implies Fermat's last theorem. It was at this point that Wiles was inspired to prove the Taniyama conjecture. Beginning with the fundamental work of Langlands, he showed how one can construct a modular form whose L-series is the same as the L-series of a given elliptic curve over the rationals. At first, his announced proof of 1995 had a gap in it which was subsequently corrected in a joint paper of his with Taylor. With this, the proof of FLT was complete.

What are the future directions of research? In the last two decades, some new connections have been discovered linking Feynman diagrams, knot theory, zeta functions, and more generally, multiple zeta functions. This is a novel theme linking number theory and physics and will undoubtedly inspire many more discoveries. This is the way science progresses: through small steps by innumerable researchers. This gives us hope. We can all join in the adventure of expanding human knowledge.

To summarise, we have seen there are eight methods of generating good questions in the art of research. They are

- Survey,
- Observations,
- Conjectures,
- Re-interpretation,
- Analogy,
- Transfer,
- Induction and
- Converse,

giving us the acronym of SOCRATIC.

Science has come very far in expanding our vision. The Hubble telescope has been able to look very deep into outer space, as far as the coma cluster of galaxies whose movement subtantiates the claim for dark matter. If we are able to see this far, it is not that we have stood on the shoulders of giants, but rather because it is the power of the human mind to question, to inquire, that we have exercised.

References

[1] Deligne, P., La conjecture de Weil. I, *Publications Mathématiques de l'IHÉS,* **43** (1974), 273–307.

[2] Mordell, L. J., On Mr. Ramanujan's empirical expansions of modular functions, *Proceedings of the Cambridge Philosophical Society,* **19** (1917), 117–124.

[3] Ram Murty, M., Panorama Lectures: The Ramanujan conjectures and zeta functions, *Ramanujan Mathematical Society*, (2017), to appear.

[4] Perelman, G., The entropy formula for the Ricci flow and its geometric applications, n arXiv.math.DG/0211159, November 11, 2002.

[5] Perelman, G., Ricci flow with surgery on three-manifolds, n arXiv.math.DG/0303109, March 10, 2003.

[6] Perelman, G., Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, n arXiv.math.DG/0307245, July 17, 2003

[7] Ramanujan, S., On certain arithmetical functions, *Transactions of the Cambridge Philosophical Society,* **22** (1916), 159–184.

[8] Weil, A., Number of solutions of equations in finite fields, *Bulletin of the American Math. Society,* **55** (1949), 497–508.

[9] Weil, A., Two lectures on number theory, past and present, *L'Enseignement Mathématique,* **20** (1974), 87–110.

[10] Weil, A., A 1940 Letter of André Weil on Analogy in Mathematics, Translated by Martin H. Krieger, *Notices of the American Math. Society,* **52** (2005), 334–341.

[11] Weil, A., Modular elliptic curves and Fermat's last theorem, *Annals of Mathematics,* **141** (1995), no. 3, 443–551.

M. Ram Murty
Department of Mathematics, Queen's University
Kingston, Ontario, K7L 3N6, Canada
E-mail: *murty@mast.queensu.ca*