Presented at the workshop *Towards a New Epistemology of Mathematics*, held at the GAP.6 Conference in Berlin, September 14-16 2006.

The useful and reliable illusion of reality in mathematics

Keith Devlin
Stanford University

Abstract

Recent years have seen a growing acknowledgement within the mathematical community that mathematics is cognitively/socially constructed. Yet to anyone doing mathematics, it seems totally objective. The sensation in pursuing mathematical research is of discovering prior (eternal) truths about an external (abstract) world. Although the community can and does decide which topics to pursue and which axioms to adopt, neither an individual mathematician nor the entire community can choose whether a particular mathematical statement is true or false, based on the given axioms. Moreover, all the evidence suggests that all practitioners work with the same ontology. (My number 7 is exactly the same as yours.) How can we reconcile the notion that people construct mathematics, with this apparent choice-free, predetermined objectivity? I believe the answer is to be found by examining what mathematical thinking is (as a mental activity) and the way the human brain acquired the capacity for mathematical thinking.

Introduction

When I do mathematical research, the overwhelming sensation I have is one of discovering facts about a pre-existing world that is "out there," with my research amounting to an exploration of that world. Practically every other mathematician I have spoken to about it says the same, and many have made similar observations in print.

Yet I, along with many (though in this case by no means all) other mathematicians acknowledge that mathematics is an abstract domain of ideas and mental processes that we humans have created — it is a social construct. I (and many of my colleagues) find the idea of mathematical objects having some prior existence simply not tenable, at least if we understand "existence" in a manner even remotely close to its generally accepted meaning.

I need to pause to make an important aside at this point. This paper is not intended as a philosophical analysis of Platonism, a subject that over the years has generated a considerable amount of debate. Rather, my aim is to try to understand, in everyday terms as a practicing mathematician, why doing mathematics creates the sensation of reality that it does. The explanation, if there is one to be had, will surely have to involve both psychological and neurophysiological considerations, and neither of those disciplines is at a sufficiently advanced stage to be able to buttress a solid argument. Thus, my argument will of necessity have the flavor of a "folk theory," and be at best suggestive of a possible more scientific explanation some time in the future.

The puzzles that the social construct view leads to are to explain, first, why the mathematics we develop or discover is the same for all of us (for instance, your number 7 is the same as mine, and we both agree that 3 + 5 = 8), and second, why does doing mathematics create an almost unavoidable sense of discovery?

No other area of human creativity has this feature. True, novelists and playwrights often remark that, once they have created their characters, they take on a life of their own and do things over which the novelist of playwright has little of no control. And Michaelangelo is alleged to have said that he did not create the David, rather it was already present in the block of marble he started with, and he simply removed all the rock that was not the David. It is not hard to understand why they would feel this way. Good novelists and playwrights create characters that follow the norms of human behavior and great sculptors have a keen sense of visual aesthetics. Yet it is easy to imagine a character in a novel or a play do some things differently, or that if Michaelangelo had died before completing the David and someone else had finished the work, then certain features would have come out differently. But with mathematics this is not the case. If certain mathematicians had not lived, it is possible that some of their mathematical results would not have been obtained, or would have been discovered much later than they were, but it is simply not possible that someone else could have come along and proved a contrary result. Mathematics cannot go any other way than the way it does.

I believe that the only way to reconcile the sensation and absolute certainty of mathematics with its nature as a social construct is to look closely at the way the human brain — as a biological organ — does mathematics. Thus, although the question I address is very much one in the philosophy of mathematics, my proposed solution will not be one of philosophical argument, rather will be based on considerations of cognitive science, evolutionary biology, and neurophysiology.

I begin with a brief discussion of the nature of mathematical thought.

Mathematical thought

What is the nature of mathematical thought? Although I have been a mathematician for forty years, I am still not clear exactly what the nature of the mathematical thought process is.

I am sure it is not linguistic. Mathematicians do not think in sentences; at least not most of the time. The precise logical prose you find in mathematical books and papers is an attempt to *communicate* the results of mathematical thought. It rarely resembles the thought process itself.

I am in remarkably good company in having this view of mathematical thinking. For instance, in 1945, the distinguished French mathematician Jacques Hadamard published a book titled *The Psychology of Invention in the Mathematical Field*, in which he cited the views of many mathematicians on what it feels like to do mathematics. Many of them insist that they do not use language to think about mathematics. Albert Einstein, for instance, wrote:

> *Words and language, whether written or spoken, do not seem to play any part in my thought processes. The psychological entities that serve as building blocks for*

> *my thought are certain signs or images, more or less clear, that I can reproduce and recombine at will.*

Hadamard himself makes the same point:

> *I insist that words are totally absent from my mind when I really think . . . even after reading or hearing a question, every word disappears the very moment that I am beginning to think it over.*

Of particular relevance to my thesis are the mathematicians' descriptions of the way they arrived at the solutions to problems they had been working on. Time and again, the solution came at a quite unexpected moment, when the person was engaged in some other activity and was not consciously thinking about the problem. Moreover, in that inspirational moment the whole solution suddenly fell into place, as if the pieces of a huge jigsaw puzzle had been dropped onto the floor and miraculously landed as a complete picture. The mathematician "saw" the solution and instinctively knew it was correct.

No language is involved in this process. Indeed, with a problem for which the solution is fairly complex, it might take the mathematician weeks or even months to spell out (in linguistic form) the step-by-step logical argument that constitutes the official solution to the problem — the *proof* of the result.

So if mathematical thought is not linguistic, if mathematicians do not think in words (or algebraic symbols), how exactly does it *feel* to a mathematician thinking about mathematics? This is, of course, a subjective question, so I will have to give a subjective answer, but I know that my experience is fairly common among mathematicians.

When I am faced with a new piece of mathematics to understand or a new problem to work on, my first task is to bring to life the mathematical concepts involved. It is as if I have been given detailed instructions (including plans and blueprints) to build and furnish a house. By studying the instructions, I can locate and acquire the necessary materials, fittings, and furnishings, and, step by step, I construct the house. When it is finished, I move in. Because I have built the house myself, I know the layout. Although it feels a bit strange at first, within a few days I know the house so well that I can get around quite well in the dark.

Of course, some things in the house were bought as completed units — the boiler, for instance — and I do not fully understand how those items are constructed or how they work. For example, I simply know the manufacturer's specifications of the boiler and how it fits into the heating system. I have neither the time nor expertise to learn boiler design and manufacture. I rely on the claims made by those who have designed and manufactured the boiler that it works as they say it does. If there later turns out to be a problem, I will have to call them in for help.

But I can live in the house without having to think consciously about the building process. The house is there, I am living in it, and I know my way around. My next task is to move the furniture around so that it best suits my lifestyle. Because I am familiar with the house and its contents, I do not have to think about each piece of furniture individually. I can concentrate on the *arrangement* of the furniture, what items go with what other items.

Similarly, when I start to think about a new piece of mathematics or am faced with a new mathematical problem, my first task is to build the “house” — a “house” built of abstract mathematical objects, fastened together by abstract logical and structural relationships. *Understanding* mathematics is like building the house and thereafter knowing my way around it.

Others have used different metaphors, though always the crucial feature is of exploration of a pre-existing domain. For example, David Hilbert, probably the leading mathematician at the end of the nineteenth century, used a garden metaphor when he wrote these words about his mathematical collaboration with Felix Klein, one of the pioneers of group theory:

> *Our science, which we loved above everything, had brought us together. It appeared to us as a flowering garden. In this garden there are beaten paths where one may look around at leisure and enjoy oneself without effort, especially at the side of a congenial companion. But we also liked to seek out hidden trails and discovered many a novel view, beautiful to behold, so we thought, and when we pointed them out to one another our joy was perfect.* [Quoted in Weyl (1944), p.614.]

These days, there is such a great volume of mathematics, much of it so complex, that no one mathematician could possibly master more than a small fraction. Consequently, like every other mathematician, I have to take a lot on trust. For example, to solve a particular problem I may use results obtained by other mathematicians. In order to develop my solution, I must understand those results and be able to apply them to my own problem. But I may well have neither the time nor the necessary expertise to understand just how those other mathematicians obtained those results. I rely on their expertise, together with the fact that, in order to be published, their work had to be examined in detail and certified as correct by other mathematicians — just as when I purchase a sophisticated, computer-controlled boiler for my house, all I need to know are the power ratings and the installation procedures. I do not need to understand how the boiler works. I rely on the expertise of the designers and manufacturers who produced the device, together with the government certification procedures that ensure its safety.

Thus, understanding the mathematics required to solve my problem may well involve learning a number of mathematical facts without knowing exactly how they were established. Just how much I need to know about those supporting facts in order to solve my problem is a matter of mathematical judgment. I may find out later on that I need to learn more about how a particular fact was established by someone else — perhaps because I want to modify that earlier result in some fashion. But I won’t worry about that possibility until it arises.

Once I have understood the mathematics involved in a particular problem, then I can try to solve it. *Trying to solve a problem* is like moving the furniture around in the house to find the best arrangement.

Notice that, once the house has been built and the instructions and plans have been stored away, there is no more need for language. I simply *live* in the house. Language is required only if some problem sends me back to the plans, or if I want to remodel or purchase a new item. And, of course, I need language if I want to describe to someone else how I built the house or why I arranged the furniture the way I did.

To me, then, *learning* new mathematics is like constructing a mental house in my mind; *understanding* that new mathematics is like becoming familiar with the interior of my mental house; and *working on a mathematical problem* is like arranging the furniture. *Thinking* mathematics is like *living* in the house. As a mathematician, I create a symbolic world in my mind and then enter that world.

But why is the mathematical world I create and explore, to all intents and purposes the same as the one created by every other mathematician?

I shall provide three answers, one within the framework of cognitive science, the second based on our knowledge of the evolutionary development of mathematical ability, and the third in terms of neurophysiology. The three answers are all consistent, and can be regarded (I do so regard them) as three different levels of the same argument. My thesis is related to the fact that the above analogy of building, furnishing, and living in a house is much closer to the practice of doing abstract mathematics than might first be supposed.

The cognitive argument

In their book *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*, George Lakoff and Rafael Nunez argue that mathematical entities are constructed by a process of iterated metaphors that are grounded in everyday experience of the physical world. With each new metaphor in the iteration chain, we lift reasoning processes that (therefore) are also grounded in everyday reasoning.

At its heart, the fundamental driving force behind the Lakoff–Nunez approach is the rhetorical question, “How else could a physical brain that evolved to ensure our survival in the physical and later social worlds construct and do mathematics?” In a much earlier work with David Johnson, *Metaphors We Live By*, Lakoff presented a similar argument for our capacity for language.

Since the Lakoff–Johnson and Lakoff–Nunez arguments are very well known, I shall not describe them here. Instead, I will move quickly to establish a consequence of the metaphor theory for the sense of realism that accompanies mathematical thought.

What the metaphor argument comes down to is that we don’t so much acquire new mental skills or abilities as we learn how to use existing skills in new domains. More precisely, we construct (or rather our brain constructs, often without any conscious knowledge on our part) a mapping from an existing domain of expertise to a new domain and we use that mapping to lift the processes we use in the existing domain into the new domain. In other words, we construct a metaphor. One we have mastered a domain in this way, the metaphor construction process may be iterated to extend our abilities to yet another domain. See Figure 1.

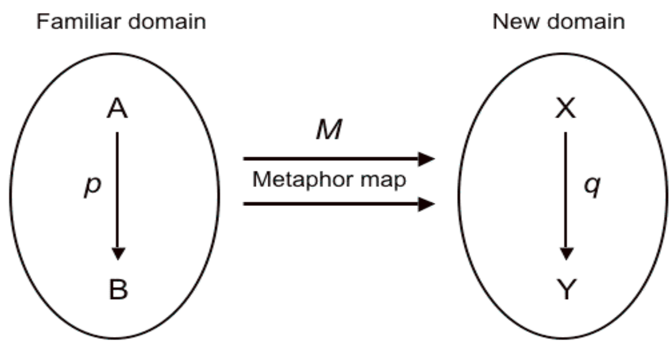

Figure 1. Metaphor. The metaphor maps a familiar domain into a novel domain. It takes familiar concepts such as A and B to new concepts X and Y, and a familiar process *p* that leads from A to B to a new process *q* that takes X to Y. There are two alternate interpretations of the diagram in terms of cognitive activity. Either the brain learns to perform *q* on X to get to Y, $q(X) = Y$, perhaps using *p* as a model, or else the brain pulls back from the new domain to the familiar one, applies *p*, and then passes back to the new domain again, $M\ p\ M^{-1}(X) = Y$. The thesis presented in this paper is that the brain uses the second mechanism. In other words, the way to view the diagram is right-to-left, not left-to-right.[1]

A particularly important feature of this metaphor process from our current perspective is that construction of a metaphor is complete when the new domain appears to us as real and concrete and familiar as the old. Since the entire iteration begins with the real, physical world, a simple induction proof establishes the fact that any domain over which we achieve mastery will seem real and concrete. To put it another way, we never actually learn to reason with abstractions in an abstract domain, rather we continually take a new abstraction and, through the metaphor process, bring it into the cognitively concrete realm that our minds can handle.

It is important not to be deceived by the simplicity of the above argument. The conclusion is deep and profound: The brain can think only in real, concrete terms. We do not learn to think abstractly; we learn to make the abstract appear real. We do not extend into a new realm, we pull back into a familiar one.

I should stress that neither Lakoff and Nunez nor I claim that this metaphor construction process is a conscious one, or even that it is best described at the level of mind. Rather, the issue is what brain circuitry[2] is used when a new concept or domain is being learned. Does the brain create new circuitry or does it map the new domain back to existing circuitry and use that? Reflection on the way the brain works (insofar as neuroscience

[1] Lakoff and Nunez claim that this pullback metaphor process will give all of mathematics. I and others have argued that it does not apply to some of the more advanced parts of the subject, which are linguistically defined. For those parts of mathematics, we suggest, a modified process is used.

[2] I am using the word "circuitry" in a somewhat metaphorical fashion, to refer to established neural activity patterns.

has so far provided us with some knowledge) together with an application of Occam's razor, suggests to me that we should opt for the pullback mechanism.

The experience we have when we learn a new mathematical concept or technique is that it first seems abstract and then, after practice, it becomes concrete. (Think how concrete numbers seem. Yet they are every bit pure abstractions as any other mathematical concept.) It is tempting to view the acquired sense of concreteness as a *consequence* of growing mastery of the new concept or technique, but this is not the case. Rather, the feeling of concreteness is a *precursor* to mastery — it is how the brain achieves the mastery.

This is why mathematics is perceptually Platonistic (to give this phenomenon a name). That is the only way the human brain can do it.

Moreover, since the metaphor-iteration process begins in the physical world, which is the same for all of us, we all end up with the same mathematical domain, or at least mathematical domains that agree in their structure and properties wherever they overlap.

In the remainder of this paper I will present two (inter-related) further arguments that support that last claim, one evolutionary, the other neurophysiological. Preparatory to both arguments, we need to take a brief look at what brains do.

What brains do

Brains are primarily survival devices. They evolved to mediate between input stimuli and output responses, enabling their owners to discriminate between different kinds of environmental influences and to learn from past experiences. Mental activity that supports the creature's survival or helps it achieve goals consistent with its survival is generally referred to as "intelligent behavior."

Though the popular model of the human brain today is a computational device, that is primarily just a reflection of the dominant technology in contemporary society. In the early days of the telephone system, people likened the brain to a telephone exchange. Consideration of the evolutionary development of brains leads to a model that I personally find much more compelling: a brain is a device for discriminating *types* (of objects, people, creatures, situations, emotions, etc.) and associating with discriminated types certain types of action, generally actions appropriate to the creature's survival.

In fact, intelligent behavior can be thought of as (I would go further and say "is") the discrimination of *sufficiently many* types and the association of *sufficiently appropriate* types of response to the types perceived. Let me elaborate this point. In particular, what do we understand by the word "intelligent" and what exactly do I mean by that word "sufficiently" in that previous sentence?

A crucial feature of living entities (although not one unique to living creatures) is that they adapt (some of) their behavior according to changing circumstances. A simple binary thermostat provides a non-living example of such reactive behavior. For an example that is a living organism not having a brain, consider the way that a sunflower orients itself throughout the day so that the flower always faces the sun. We might describe the behavior of the thermostat by saying that it discriminates two types (say, warm and cold). Likewise, the sunflower could be said to discriminate the type “facing the sun.” Though the thermostat acts in the manner for which it was designed and the sunflower’s actions ensure its survival, we would probably not want to refer to their behavior as “intelligent”. Whatever intelligence is, it surely requires a brain. But how much of a brain?

Again, there are simple water-based bacteria, which move toward nutrients in the water and swim away from regions that contain chemicals poisonous to them. Such a creature discriminates the types nutritious and poisonous. From the point of view of the individual bacterium, such discrimination, together with the behaviors it evokes, could be described as “intelligent” behavior, but as a fairly simple, low level, chemical response, most of us would still be reluctant to dignify it with such a loaded term as “intelligence”.

Moving on, *Stomphia coccinea* is a species of sea anemone whose ocean environment typically contains eleven types of starfish, only two of which prey on the anemone. If one of the nine non-threatening species of starfish happens to brush against a *Stomphia*, the anemone does not react. If one of the two predator starfish touches a *Stomphia*, however, the anemone recoils immediately. As with the bacteria, this is purely an automatic response to a particular chemical or collection of chemicals found in the two predator species but not in the other nine. But the *Stomphia’s* ability to distinguish between two types of starfish — dangerous and not — is a cognitive ability that helps to keep it alive. Are we fully justified in refusing to call that behavior “intelligent”?

Proceeding to creatures with brains, when rooks see an animal coming too near their nest, they will pick up stones and drop them on the invader to drive the threat away. The rooks discriminate creatures that are “too near” from those that are not and act accordingly. Is that “intelligent” behavior?

A leaf cutting ant that encounters an opening too narrow for the leaf it is carrying will maneuver the leaf until it can pass through the opening. The ant discriminates openings that are sufficiently wide from those that are not, and adapts its behavior accordingly. Does that qualify as “intelligent” behavior?

How about the now famous octopus that figured out how to unscrew a mason jar in order to get to food inside? Several kinds of type are involved in this sequence. Intelligent behavior or not?

Or consider the chimpanzee which, when presented with a banana suspended beyond its reach from the roof of the cage, thought for a while, then dragged a box from the other side of the cage until it was beneath the banana, climbed onto the box, and grabbed the banana. The chimp clearly had to discriminate many types in order to perform that feat, and it is hard not to classify this action as “intelligent” in the very same problem-solving sense we would apply the term to humans.

Finally, when you or I feel sick we visit a doctor. We do so because the doctor has acquired the ability to discriminate a great number of types of illnesses and symptoms and is able to associate with those types the appropriate kinds (types) of action: particular medicines to take, certain further tests to carry out, perhaps a surgery to perform, and so forth.

Looking back at those examples, we may be unable to draw a clear line between intelligent behavior and mere “stimulus-response” activity, but by the time we get to chimpanzees most of us seem to agree that we are in the realm of intelligence. Though my first few examples might suggest that there is a continuous spectrum of increasing degrees of “intelligent” behavior, with thermostats and sunflowers at one end and humans at the other, this is not the case. There is a spectrum, but it is not continuous. There is a difference in kind between the chimpanzee and all the previous examples, and a second qualitative difference between chimpanzee intelligence and human intelligence.

The McPhail hierarchy

A useful way to approach the behavioral capacities of brains is to make use of a classification of “intelligent activity” introduced by the British psychologist Euan McPhail a few years ago. McPhail provided a three-point scale of intelligent behavior.

Practically all living things exhibit McPhail’s lowest level of intelligence: they can respond (in a manner advantageous to their survival) to certain stimuli. Call this stimulus–response (or S-R) behavior. You don’t require a brain to produce S-R behavior; think of the way a sunflower turns its head to follow the sun each day. McPhail uses the term “S-R intelligence” to refer to any stimulus-response activity advantageous to survival, and I shall follow suit. Apart from the thermostat, all the examples I gave above up to, but not including, the chimpanzee are of S-R intelligence.

The chimpanzee’s solution to the banana problem, in contrast, requires a plan of action. The chimp not only *acts* intelligently, it *is* intelligent. Notice, however, that all the elements of the chimp’s plan were physically present — to use a computing term, the chimp formulated its plan “on line,” while still receiving visual input about the objects it was thinking about.

The chimp exhibits what McPhail calls stimulus–stimulus (or S-S) behavior, and may be said to have S-S intelligence. S-S intelligence requires a brain. Learned S-S behavior involves the forming of mental representations. Most likely brains developed initially as organs to facilitate S-S behavior. It is an effective survival strategy, since it allows an animal to adapt its response pattern based on past experience.

McPhail’s third level of intelligence involves symbolic representation and language, and as far as we know, humans are the only creatures that have attained this level. Call it LI, for linguistic intelligence. (The reason for giving language such prominence in a classification of intelligent *behavior* will become clear presently.)

McPhail’s classification is a rough division of a broad spectrum into three overlapping groupings. At least, the first two groups overlap, and we can understand how it is possible to progress up the spectrum by a long chain of incremental evolutionary developments.

The advantages conferred on an S-R intelligent species in terms of natural selection as it acquires S-S intelligence are fairly clear, since S-R intelligence generates highly predictable behavior patterns that a predator can take advantage of, whereas S-S intelligence means that the creature can mediate its responses, taking account of past experience or novel features of the environmental inputs. Thus, over time, we would expect to see more and more species acquire S-S intelligence. LI intelligence offers still greater advantages in terms of natural selection, but whereas it is easy to envisage mechanisms that would lead from S-R intelligence to S-S intelligence (for example, the introduction of intermediate, or “hidden” layers in a neural net can achieve this effect), it is much harder to postulate a mechanism that will take a species into the LI-intelligence realm. And in fact we have evidence only that this has happened once, namely with the evolution of *Homo sapiens*. But just how did the transition occur?

I will answer this by first presenting an argument having a fictional element, then I shall remove the fiction.[3]

Over a period of three-and-a-half million years, our ancestors brains increased in size and complexity, eventually becoming nine times the size of the brains (per unit weight) of other mammals. As they did so, they developed an increasingly rich capacity to differentiate types of stimuli and forge links between those types. What began as an organ to generate physical responses to physical stimuli evolved into a device that could produce one internal stimulus from another — S-S behavior. Over time, *Homo erectus* brains developed a wide repertoire of such stimulus–stimulus links. One might be willing to call this activity “thinking.” But the subject matter of that “thinking” remained the physical environment. *Erectus’s* world — what he thought about — was the physical world around him.

Imagine now (and this is the fictional part) that a second brain grew, parasitic on the first. This second brain had a similar structure to the first, except that its world was the first brain. Where the first brain received its initial stimuli (its inputs) from the physical world, the second brain received its stimuli from the first brain; and where the eventual responses (the outputs) of the first brain were physical actions in the world, the responses of the second brain were further stimuli to the first brain. We might then be inclined to refer to the activity of the second brain as “symbolic thought.” Whereas the function of the first brain was to manipulate physical objects in the physical world, the function of the second brain was to manipulate symbolic objects that arose in the first brain. The two brains had essentially the same structure, and worked the same way, the only difference between them being the sources of their inputs and the targets of their outputs.

As a first step toward eliminating the fiction now, we can, to a first approximation, think of the prefrontal lobes of the *Homo sapiens* as this second, “parasitic” brain. This part of the brain is a recent addition, and it’s where much of language processing takes place. But given the degree of interconnections between all parts of the brain, this is far too simplistic. Strictly speaking, there is no parasitic second brain. That’s what makes my account so far fictional. Eliminating the fiction completely now, I suggest that the emergence of symbolic thought arose when *the* brain — the first brain, if you like — *itself* developed the ability to function as the second brain. In other words, the brain became able to generate its own stimuli — to create and think about imaginary situations of its own creation, independent of any input from the physical world.

[3] This is essentially the argument presented by Bickerton (1995).

Bickerton (1995) refers to this form of brain activity as "off-line thinking." Another way to think about it is that, instead of directing actions in the world in response to stimuli from the environment, the brain runs simulations of possible actions (and their consequences) — "What if?" scenarios — that would be initiated by imagined (external or internal) stimuli.

Let me explain this change in brain capability in physical terms.

Changing the brain

Our present understanding of the way the brain works is that thinking (conscious or otherwise) involves fluctuating electrical (and electrochemical) activation patterns of networks of neurons. The simplest non trivial example of a brain would have just two sets of neurons, one that accepts inputs from the outside world, another that generates output responses. Each input neuron is connected to several (maybe all) output neurons. The connections are of different capacities. The capacities of some connections may be variable, in that they become stronger the more frequently they are used.

Such a device, which is easily simulated on a digital computer by a neural net, produces stimulus–response behavior. Each pattern of inputs (different strengths of current going in to each input neuron) generates a corresponding pattern of outputs. Where neural connections have variable capacities, the device can "learn" a new stimulus–response pattern if given the appropriate training. (This requires a feedback mechanism to adjust the capacities of the variable connections.) For example, in the early 1980s, cognitive scientists David Rumelhart and James McClelland constructed a neural net with 460 input neurons connected to 460 output neurons that was able to learn how to form the past tense of a number of verbs, including both regular and irregular verbs. (See Rumelhart and McClelland 1986.)

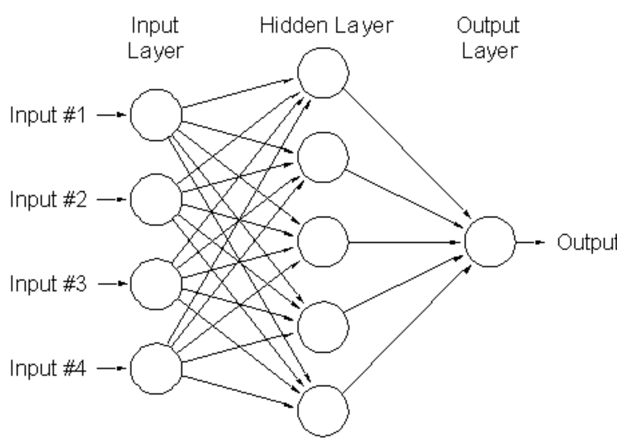

Figure 2. A simple neural network with one hidden layer.

A more complex brain would have one or more layers of neurons (neural "hidden layers") between the input and output neurons, mediating the flow between them. (See Figure 2.) Such a brain, which is also easy to simulate as a neural net, is capable of more sophisticated behavior than a simple two-layer network.

Any mammalian brain is already far more complex than any neural net that can be simulated on a digital computer. The human brain contains roughly 100 billion neurons, each directly connected to between 1,000 and 100,000 other neurons. The number of possible activation patterns is far larger than the number of atoms in the universe.

Suppose for a moment we were able to see inside the human brain, and that a neuron lit up whenever it fired, its brightness depending on the strength of the firing. Given an input pattern from the nervous system, we would first observe a collection of illuminated neurons of different brightnesses spread all over the brain, with concentrations of brightly lit neurons in one or two areas. (See Figure 3.) This pattern would represent the input stimulus. We would then observe a veritable light show as current flowed (in parallel) from neuron to neuron. This is the brain processing (or thinking about) the input stimulus. If this activity resulted in a command to the body to perform a particular action, say to duck to avoid an approaching projectile, we would ultimately see a configuration of neurons light up that would cause signals to travel to the body's muscles, the effect of which would be that the body ducked down.

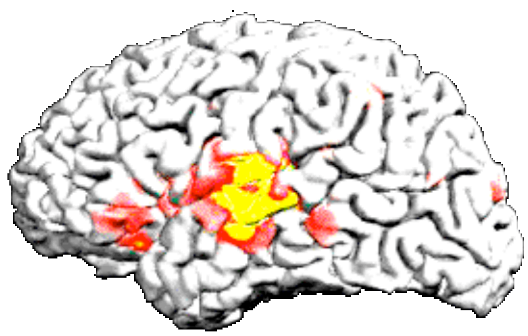

Figure 3. Neuronal activity

We would observe a similar sequence of electrical activity if we were likewise able to observe the internal workings of, say, any mammalian brain or of a pre-human hominid brain.

In terms of brain activity, off-line thinking occurs when the brain simulates an externally produced activation pattern to start the sequence, and then runs through the ensuing sequence without necessarily generating a bodily response.

Not all brain activity requires an external stimulus, of course. The brain of any creature is constantly active, monitoring, controlling, and initiating various body processes. Moreover, animals' brains initiate actions such as obtaining and eating food, and engaging in sex (although bodily signals to the brain are also involved in initiating those actions). Moreover, even when a human is engaged in the supremely off-line thought process of working on a mathematical problem, bodily responses are generated, perhaps the formation of a frown or the repeated tapping of a pencil on a desk. But these are not the kinds of stimuli or responses I am referring to. Rather, I am thinking predominantly of the kind of brain activity that is generally occasioned by an external stimulus.[4]

Of course, an external stimulus can also cause off-line thinking. Indeed, it very often happens that some external event sets us onto a particular train of (off-line) thought. Thus, the description I have given is a simplistic one. The point I am trying to get across is this: An animal having S-S intelligence but not LI cannot initiate a particular pattern of brain activity normally occasioned by an external stimulus in the absence of such a stimulus. The appropriate initial activation pattern requires an input stimulus from the outside world.

In the case of a monkey, for example, the activation pattern we might describe as "thinking about eating a juicy pear" can only be initiated by input from a particular juicy pear — say sight or smell — or perhaps a photograph of a pear. There is, unfortunately, no way of proving this, given present-day methods, since we have no way of knowing what a monkey is thinking about. Likewise, in the case of Pavlov's dogs, salivating at the sound of a bell, we have no way of knowing if they have any *thoughts* about an imminent meal. The most we can know for sure is that, through training, one initiation stimulus (real food) can be replaced by another (a bell).

For an LI animal, however, the initial activation pattern that starts a particular activity sequence can be generated by the brain itself, without any external input. For example, as humans we can *imagine* a juicy pear, complete with sight, texture, smell, taste, and sound, and think about eating it, even when no actual pear is present. (Actually, the activity pattern activated by the sight of a real pear and that initiated by an imagined pear will not be exactly the same. Dieting would be easy indeed if imagined eating were subjectively indistinguishable from real eating.)

Off-line thinking

Off-line thinking is thinking about a world of internally generated symbols. Those symbols may correspond to real objects in the world, such as when we think about our distant relatives whom we have not seen for several years. Or we can think about things that have never existed, such as unicorns.

---

[4] Dreaming is an interesting intermediate case, since dreams do appear to start spontaneously, in many nonhumans as well as in humans. Despite considerable research, however, little is known about dreaming. It appears to involve random combination of parts of activation chains that have already been produced during awake activities. The repeat of parts of "meaningful" activation chains, which were once initiated by physical stimuli, may be what gives dreams their (recollected) content, even though there is a strong element of random firing involved. But since our only conscious knowledge of our dreams is the recollection we have when we wake up, any meaningfulness may be illusory. And we have no way of knowing whether nonhuman dreaming, say by a cat or a dog, involves any mental activity meaningful to that animal. Given the problems involved in trying to understand what goes on in dreaming, it is far more likely that a greater understanding of thinking will cast light on the nature of dreams than that discoveries about dreams will help us understand thinking.

Off-line thought may be initiated spontaneously by the brain or by direct input from a real object, or by some combination of the two, such as when climbing into our car begins a train of thought about whether to trade it in for a newer model. Our existing car is real — we are sitting in it; the new car we think about acquiring in its place might not exist except as a mental amalgam of features we would like it to have.

As far as we know, no animals other than humans are able to think in this off-line fashion. Doing so is clearly a high-risk, high-gain strategy. One obvious gain comes from being able to reflect at length about past events and to plan future actions long in advance, thinking about various alternatives. This is an excellent survival strategy for a creature that is not particularly big, strong, or fast, has an easily penetrated outer skin, and does not have sharp claws or a large jaw and strong teeth. An obvious risk is that a human engrossed in thought is a prime target for a passing saber-toothed tiger or some other mishap.

One way to reduce this risk is to ensure that part of the brain continues to monitor input stimuli from the world. This is exactly what happens. No matter what activity our "symbolic" neocortex is engaged in, the amygdala (the so-called "reptilean brain" that all animals possess, on top of which the human neocortex developed) continues to do what it always did: keep its owner alive. If anything unexpected or threatening arises, it rouses the entire brain into taking evasive action.

For instance, we rely on this alarm system when we drive a car along a familiar route. Most of the time our mind wanders, perhaps reflecting on the day's activities or planning a camping trip. We are only vaguely conscious of our driving, or of what is going on around us. Then suddenly a small child runs out of a doorway toward the road ahead. Whatever we were thinking about is suddenly gone from our mind, and before we realize it we are braking fiercely to avoid an accident. All our attention is now on our driving.

With the amygdala constantly on the lookout for signs of danger, the potential drawbacks of off-line thought are significantly reduced, providing *Homo sapiens* with an opportunity to enjoy some of the benefits. Surely the most significant of those benefits was language. Off-line thinking automatically gives you full language. Or, more precisely, off-line thinking and language are two sides of the same coin. You can't have one without the other. This brilliant observation was first made by Bickerton (1995).

Language for free

To see this, notice first that off-line thinking does not require that the brain generate an exact replica of a real world situation; just an activation pattern that is sufficiently like an externally generated pattern to for the process to have survival value. The question is, what features of the world are absolutely necessary for this to occur?

Surely, agent (subject), action or relationship (verb), and patient (object) are absolutely essential for any useful internal model of the world, and a good case can also be made for at least some temporal (tense) distinctions. If the objects of thought are objects in the world, the off-line thinking amounts to simulation. If, however, the objects of thought are symbolic representations of things in the world ("words", if you like), the off-line thinking amounts to sentence construction.

But for off-line thinking, this distinction between the objects of thought being objects in the world and being symbolic representations of objects in the world is spurious. Everything is symbolic. When the brain develops the ability to generate specific activation patterns that initiate and maintain meaningful chains of activation, (referring) words inevitably arise. What else would constitute the realization in the brain of the word “cat” but the brain activation pattern that stands for “a cat” in off-line thought? To put it another way, it is precisely when the brain acquires the ability to generate such a pattern spontaneously that we are justified in saying that the individual “knows” the word “cat”. This is why McPhail’s third level of intelligence is called “Linguistic Intelligence.”

The story of evolution is one of endless opportunism. Change occurs as a result of random “errors” (mutations) in copying DNA molecules. If a randomly introduced feature offers a certain survival or propagation advantage to the members of a species, that feature will tend to become steadily more prevalent until it is the norm. Sometimes, a feature that has evolved because it offered one survival advantage turns out to be useful in a quite different way. It then evolves further, driven by the new use. Eventually, it may no longer be possible to recognize the feature’s original “function.”

The ability of the human brain to produce and understand language arose in just this opportunistic fashion. On at least two occasions, nature took a feature that had developed for one thing and used it for another, a process evolutionists refer to as exaptation.

We have observed that brains originally functioned as stimulus–response devices; they recognized certain types and produced appropriate responses. Primitive brains, for example those of reptiles or amphibians, have all their stimulus–response mechanism hard-wired into their structure. More complex brains, including human brains, have some stimulus-response mechanisms hard-wired, but they can also acquire additional stimulus–response connections through repetitive learning or training.

That growth was most likely driven primarily by the advantages of having a greater capacity to recognize types and being able to form increasingly rich associations between types of objects or situations encountered and appropriate types of responses. It is, however, likely that elements of this rich repertoire of recognized types were articulated as sounds, yielding an increasingly powerful communication system called protolanguage whose utterances were either single words or else simple object–property combinations. The early hominid line was a highly marginal one that in present-day terms would have easily classified as an Endangered Species. *Erectus’s* brain was its main advantage, and any change in its brain that offered a better chance of survival would have quickly spread through the small population. Thus, once it was possible for hominids to communicate using protolanguage, they surely did so. Once they did, communication became another selection factor driving brain growth.

Indeed, once hominids started to communicate using protolanguage, communication may even have become the *dominant* selection factor driving further brain development. If so, that would give us another instance where a feature initially selected for one function (representation) turns out to provide a second, and eventually more significant selection advantage (communication).

At the end of the long period of brain growth, somewhere between 75,000 and 200,000 years ago, the human brain acquired symbolic thinking. This was caused not by further brain enlargement but by a change in its structure. The brain acquired the ability to simulate the activation patterns normally caused by sensory stimuli, and to isolate from the body's motor centers the outcomes of the resulting thought processes. In short, the brain became able to think "off line."

Off-line thinking could have developed only after the brain had acquired sufficiently many types to support an internal "model" of the world that was adequate for reasoning. But a rich collection of types on its own would not be enough. The brain had to be able to represent, internally, a sufficiently rich world *structure*. The "skeletal structure" of the world that the brain simulated in order to think about the world off line is what we now call syntactic structure, or syntax.

When you add syntax to protolanguage, you get language. Thus, in acquiring the ability to think off line, the brain automatically acquired language. Since protolanguage was almost certainly being used for communication by that stage, language came into being as both a representational medium and a communicative medium.

Mathematics for free as well – almost

Just as language came for free with the acquisition of off-line thinking, so too did the capacity for mathematical thinking.

As explained above, off-line thinking (or LI-thinking) was not a new brain process, rather the brain developed a mechanism for running existing processes in an off-line fashion. The novel feature was a change in the input–output connections. Viewed in terms of the relationship between an individual and the environment, that new brain activity provided simulations of the world. But by shifting the focus from the things in the world being represented in the simulations to the (symbolic) representations themselves, the same brain activity could be viewed as syntax. (Things in the world are replaced by words that denote them: actions are replaced by the verbs that denote them, agents of actions are replaced by subjects of sentences, objects of actions by objects of verbs, etc.) A similar shift in perspective can give mathematical thought. The trick is to approach mathematics in the right way.

Though modern mathematics is often presented (and taught) as a reductive system, based on axioms, that is not how it first developed. Rather, early mathematics was analytic, as our recent ancestors attempted to understand the world in a more precise (and often quantitative) fashion and to be able to build things and carry out negotiations with one another in a precise manner. Reasoning off-line about the world in terms of agents (represented symbolically by subjects and objects) and actions (represented by verbs) gives language. Reasoning off-line about the world in terms of arrangements, relationships, and quantities gives mathematics.

Of course, if mathematical thinking were completely equivalent to language, no one would have difficulty mastering mathematics, something which is very definitely not the case. What off-line thinking gives are the *capacity* for language and the *capacity* for mathematical thinking. In the case of language, young children are exposed to it all the time, and pick it up with ease. (There are a number of well-studied cases of individuals who were not exposed to language during their formative years — roughly up to early teens — and those unfortunate people were never able to acquire language.) With mathematics, not only is there far less exposure to it in everyday life, there is a key feature of mathematical thinking that makes it extremely difficult for the brain to master it: abstraction.

I discussed the role of abstraction in learning mathematics in my book *The Math Gene*, so will just summarize briefly the key points here.

A characteristic feature of the human brain that no other species seems to possess is the ability to think about abstract entities. Many species seem able to reason, if only in a very rudimentary way, about real objects in their immediate environment. Some, including chimpanzees and apes, seem to have an additional ability. A bonobo ape, for example, can carry out very limited reasoning about a single, real object it is familiar with but which is not currently present. The range of human thought, in contrast, is so broad as to constitute a different kind of activity altogether. We can think about practically anything we want: real objects we are familiar with but which are not in our immediate environment, real objects we have never seen but have simply heard or read about, or purely fictitious objects. Thus, whereas a bonobo ape may reason about how to retrieve a banana it just saw its trainer hide, we may think about a six-foot long, gold-plated banana pulled along by two pink unicorns.

How is it possible to think about something that does not exist? To put it another way, just *what* is the object of our thought when we are thinking about, say, a pink unicorn? The standard answer is that the objects of our thought processes are *symbols* (i.e., things that stand for or denote other things). In *The Math Gene* I examine this idea in some detail, but for now let me simply observe this: The symbols that form the object of the thoughts of an ape or a chimpanzee are restricted to symbolic representations of real objects in the world. On the other hand, the symbols that form the objects of our thoughts may also represent imaginary versions of real objects, such as imaginary bananas or imaginary horses, or even wholly imaginary objects put together from symbolic representations of real objects in the world, such as a gold banana or a unicorn.

I find it helpful to view abstract thought in terms of four levels.

Level 1 abstraction is where there is really no abstraction at all. The objects thought about are all real objects that are perceptually accessible in the immediate environment. (However, thinking about objects in the immediate environment might well involve imagining them moved to different locations in the environment, or arranged in different ways in the environment. Thus, I think it is reasonable to view this process as one of abstract thought, even though the objects of that thought are all concrete objects in the immediate environment.) Many species of animals seems capable of level 1 abstraction.

Level 2 abstraction involves real objects the thinker is familiar with but which are not perceptually accessible in the immediate environment. Chimpanzees and apes seem capable of thought at level 2 abstraction.

Level 3 abstraction. As far as we know, only humans are capable of level 3 abstraction. Here, the objects of thought may be real objects that the individual has somehow learned of but has never actually encountered, or imaginary versions of real objects, or imaginary variants of real objects, or imaginary combinations of real objects. Though objects in level 3 abstraction are imaginary, they can be described in terms of real objects — for example, we may describe a unicorn as a horse with a single horn on its forehead. As I shall explain in Chapter 8, the ability to think at level 3 abstraction is, to all intents and purposes, equivalent to having language.

Level 4 abstraction is where mathematical thought takes place. Mathematical objects are entirely abstract; they have no simple or direct link to the real world, other than being abstracted from the world.

Level 3 abstraction comes with off-line thinking. The key to being able to think mathematically is to push this ability to "fake reality" one step further, into a realm that is purely symbolic — level 4 abstraction. This requires effort. Mathematicians learn how to live in and reason about a purely symbolic world. (By "symbolic world" I don't mean the algebraic symbols that mathematicians use to write down mathematical ideas and results. Rather, I mean that the objects and circumstances that are the focus of mathematical thought are purely symbolic objects created in the mind.)

To do mathematics, we need to think off-line about completely abstract objects that bear virtually no relationship to anything in the real world — level 4 abstraction. We need to generate brain activation patterns unlike anything that arises from sensory input. Now, we know that practically everyone can do this, since that is precisely what is required to have a sense of number and to cope with various abstractions that arise in modern life, such as the concepts of marriage, ownership, or indebtedness. The mechanism is the one we discussed earlier. We master a new abstraction by familiarizing ourselves with it until it seems more concrete. The human brain is certainly able to do this, and most people succeed with the most basic mathematical abstractions, namely numbers and basic arithmetic. But far fewer people have the interest or the motivation to continue that familiarization process into the realms of higher mathematics. Given the difficulty the brain has with new abstractions — something it never encountered throughout evolutionary history — it is hardly surprising that so many people never get much beyond arithmetic.[5]

For the purposes of our present discussion, the key point about the account presented above is that the brain handles abstract mathematical concepts and does mathematics

---

[5] A second feature of mathematics that makes it hard is the degree of rigor required in its reasoning processes. Precise reasoning is not something for which our brains evolved. But we need to be careful in drawing conclusions from this observation. Precise, formal reasoning is not required for mathematical *discovery*. Rather, its purposes are verification of things already discovered (or perhaps suspected) and convincing others of the truth of those discoveries.

The need for formal verification is a direct consequence of the nature of mathematical discovery. Trial and error, guesswork, intuition, and conversations with others can go on for days, months, even years, with key steps often being carried out while the mathematician is either asleep or thinking about something else. Although this process need not be mere haphazard stumbling — for a good mathematician it can be highly focused and efficient — it can nevertheless generate errors. Formal proofs are the final (and totally reliable) safeguard against false "discoveries."

by utilizing brain circuits that evolved in order to think about the physical and social environment, first on-line and then off-line. More precisely, and getting away from metaphorical talk of "circuits" or "information processing," the argument just presented amounts to this:

1. Throughout its evolutionary history, the *Homo* brain developed activity patterns in response to various environmental stimuli.
2. Initially, a particular activity pattern would give rise to a physical response.
3. At some stage, somewhere between 75,000 and 200,000 years ago, the brain acquired the ability to simulate the activity patterns normally associated with input stimuli, and to store and then reflect upon the outputs, without those outputs resulting in bodily motor actions — what I am calling off-line thinking.
4. Because the brain activity patterns are essentially the same as those involved in on-line thinking, off-line thinking gives the sensation of thinking about real things in the world. (Sensations are simply manifestations of a familiar activation pattern.)
5. Around 10,000 years ago, with societal structure reaching a sufficiently complex state that there was need for precise means for counting property, tracking the seasons, building houses and temples, and the like, our ancestors started to think off-line about the world not in terms of subject–action–object (giving rise to language) but rather quantity, arrangement, and structure (giving rise to mathematics).
6. Again, because the brain activity patterns are essentially the same as those involved in on-line thinking, this alternative kind of off-line thinking also gives the sensation of thinking about real things in the world.
7. Hence, the reason why doing mathematics yields the sensation of reasoning about concrete entities "in the world" is that such thought processes are comprised of brain activation patterns that are associated with real world stimuli.

Conclusion

In this article I have tried to provide an answer to the questions:

1. Why does doing mathematics carry an overwhelming sensation of reasoning about concrete objects that exist outside of our minds?
2. Why do all mathematicians converge on the same mathematics (to an extremely high degree, and to the extent that this can be checked)?

As I stated at the outset, since these questions are subjective, it is not possible to give a comprehensive, scientific answer. The best one can hope for (at least given the present state of knowledge in the various relevant sciences) is a folk theory or plausibility argument. My goal, however, was simply to provide an answer a question I believe to be both important and interesting. I believe that the discussion I presented above meets that target.

I definitely did not set out to try to explain how people do mathematics. Indeed, I said little about the actual practice of doing mathematics. My focus was on the *capacity* for doing mathematics and how that capacity is realized in the human brain, with a view to explaining why doing mathematics carries the sensations of reality that it does.

Part of my argument was that the capacity for mathematics is inherent in the capacity for language and the capacity for off-line thinking — in fact, all three capacities are different functionalities of the same underlying brain capacity.

Doing mathematics involves far more than simply engaging a latent capacity, however. A characteristic feature of mathematics is that it involves precise reasoning about precisely defined abstract entities. That is something that does not come natural to the human brain. The brain must be trained to reason about such entities in the appropriate way. (If experience in mathematics education is any guide, many people are, for whatever reason, unable to complete the necessary training to the point of competent mathematical performance.) I said nothing at all about those aspects of mathematical thinking.

References

Bickerton, D. *Language and Human Behavior*, Seattle, WA: University of Washington Press (1995).

Devlin, K. *The Math Gene: How Mathematical Thinking Evolved and Why Numbers Are Like Gossip,* Basic Books (2000).

Hadamard, J. *The Psychology of Invention in the Mathematical Field*, Princeton University Press (1949).

Lakoff, G. & Johnson, M. *Metaphors We Live By*, University of Chicago Press (1980).

Lakoff, G & Nunez, R. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being,* Basic Books (2000).

Rumelhart, D. & McClelland, J. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press (1986).

Weyl, H. “David Hilbert and His Mathematical Work,” *Bulletin of the American Mathematical Society* 50. (1944).