Aristotle to Digital

# How do Humans Know How to Know

The last post dealt with the innate and special capacity that humans have that allows them to learn language by distinguishing differences and commonalities in object properties and by correlating what they see with the sounds that they hear.

A skeptical reader may object that this capacity is not particularly special. After all, isn’t a newborn constantly bombarded with messages trying to teach him about the world around him? Doesn’t a newborn have to be repeatedly exposed to various objects and told the various properties before he catches on? Isn’t this prolonged maturation process merely a much more complicated example of training a neural net?

My emphatic answer is no. Maybe the physiological process in our brains is best described by the mathematical and computational models used to describe a neural net. So what? Identifying that to be true is to misses the point. Who

* Objection: Practice countless times
* Isn’t this like training a neural net
* But who wrote the code
* No, despite

A possible object at this point is that Of course, a newborn is surrounded Despite the amount of interaction and on-the-fly teaching that a baby as receives from the people surrounding him as he learns to talk, it would be a mistake to call it formal education. Somehow the baby just knows how to know and knows how to learn. That idea brings us into contact with the philosophical discipline of epistemology which is anotherThis capacity depends very heavily on humans also being able to analyze what constitutes the mode of being of a particular object.

Identification and separation of the essential from the accidental properties is the key faculty that Man possess that currently separates humans from artificial intelligence in much the way that the opposable thumb separates Man from the lower primates, although the gap between the latter is a mere crack compared to the chasm separating the former.

The reason is

# Nuances of Language

In this post, I would like to explore some of the nuances of language and communication, especially as they arise in everyday speaking. Particular focus is on context, tone, and non-verbal cues that add additional complexity to an already complex situation. I close with some thoughts about how we can ever teach a computer to navigate through this highly complex landscape in hopes of someday having software mimic this behavior.

Learning language was the focus of the last post, where I argued that every human shares in a capacity to distinguish and classify object properties and that capacity is what allows humans to learn to speak. In the situations examined in that earlier post, the teacher of the new language (be it the traveler’s foreign friend or the newborn’s friends and family) genuinely wanted the student to learn the new language. That is a special context that is rarely available in common social interactions.

Most communication events take place casually, where both communicants expect roughly the same level of maturity and expertise of the other. The nuances then arise from either an assumption made by one party that is only partially, if at all, shared by the other or from a conscious effort by one party to layer on additional meaning or adapt the plain meaning to something else.

Into this first category fall all the little innocent miscommunications that fill our day-to-day lives. Into the second fall things like half truths, jokes, double entendres, left-handed compliments, sarcasms, hints, lies, and cons. In other words, this category contains all the spicier and more interesting aspects of language. We’ll deal with each category in turn.

First consider the innocent miscommunication. It is best exemplified by an anecdote. Last Friday I was meeting a friend for lunch. He had arranged to pick me up and most of our communication, after the initial phone call, was done by texting. At 11:42 am he issued the following text:

**On my way in 5 min.**

I promptly got up and went to wait for him. At 11:56 am, I called him asking where he was to which he responded “I’ll be there in about 10 minutes”. Needless to say, I was confused. When he arrived, we talked it out. What he meant by ‘on my way’ was that he was leaving in 5 minutes whereas I interpreted it to mean that he would be arriving at my location 5 minutes after the text was sent. Clearly the text meant different things to each of us

**Sender: ‘On my way in 5 min’ = ‘I am leaving my location in 5 minutes’**

**Receiver: ‘On my way in 5 min’ = ‘I will arrive at your location in 5 minutes’**

A simple foul-up really, but one that resulted from context and assumption.

Next consider the much more interesting realm of the intentional addition of meaning to add information or flavor or to deceive. I don’t have the space or the inclination to cover all aspects but I will touch upon 3 of the above-mentioned items: hints; left-handed compliments; and sarcasm.

The hint category is, of course, a staple in murder mysteries and finds frequent expression in the works of Agatha Christie. She actually goes into this point with the words from one her most famous sleuths Miss Marple. In the ‘Thumbmark of Saint Peter’, one of Miss Marple’s nieces calls for help when her entire village begins to ostracize her believing her responsible for the sudden death of her husband. In the process of solving the mystery for her niece, Miss Marple resolves to find meaning to the seemingly feverish last words of the dying man. As a preamble to her explanation, she starts by discussing context. To quote:

“Has it ever occurred to you,” the old lady went on, “how much we go by what is called, I believe, the context? There is a place on Dartmoor called Grey Wethers. If you were talking to a farmer there and mentioned Grey Wethers, he would probably conclude that you were speaking of these stone circles, yet it is possible that you might be speaking of the atmosphere; and in the same way, if you were meaning the stone circles, an outsider, hearing a fragment of the conversation, might think you meant the weather. So when we repeat a conversation, we don’t, as a rule, repeat the actual words; we put in some other words that seem to us to mean exactly the same thing.”

This “as a rule” substitution is part and parcel of how we think and express and describe events in the world around us and it is entirely contextual. It’s behind every hint ever given for a riddle or a problem. The plain meaning of the words in the hint are augmented by the context into which the words belong. It isn’t clear how to describe it, you either get the hint or you don’t.

Next is the idea of a left-handed compliment. When I was in high school, one guy stands out as the king of left-handed compliments. Like most high-schoolers, originality and variety were not high on his list. His standard stock in trade was to walk up to someone in the hall and say, in a fairly high voice “That shirt looks good…on you!” To pull this highly dazzling witticism off, he would exaggerate the pause and then say the last part “on you!” with a clownish emphasis reminiscent of Steve Martin’s “well excuse me!”

<iframe width="420" height="315" src="//www.youtube.com/embed/zANvYB93u2g" frameborder="0" allowfullscreen></iframe>

Of course, he was versatile enough to change “shirt” to “hair cut” or “pants” or whatever part of the victim’s appearance was most deserving of derision.

Finally consider the related category of sarcasm. Often a standard joke when dealing with smart yet social-awkward robots (I’m thinking Red Dwarf here), there is something altogether funny and yet nearly indescribable in the situation where a statement meant sarcastically is interpreted seriously. I really can’t improve on the following clip from the Big Bang theory – so enjoy

<iframe width="420" height="315" src="//www.youtube.com/embed/DF7MroTLDfU" frameborder="0" allowfullscreen></iframe>

Anyway, I hope that after that short tour of the contextual modes of language it is clear that there is far more to meaning than what is found on the written page or in the dry monotone speaking. Artificial intelligence practitioners have yet to capture even the basic capacity that allows a newborn to learns its own mother tongue let alone allow a person (Sheldon not withstanding) to be able to understand hints and humor and sarcasm.

And as far as the dangers of the AI demon escaping the pentagram, well, let me say a couple of things. First, if only AI were that capable, my tech support experiences would be a whole lot better. Second, until that day comes, I’ll keep worrying about those human scam artists, snake-oil salesman, and con artists who are a lot less book-learned and a whole lot smarter than the average PhD in computer science.

## Yogi Berra Logic

Continuing on with the theme of language (natural and otherwise) and expanding on how we take for granted the ability of humans to shift contexts, usually on a moment’s notice, I submit a comical and whimsical topic. ‘Yogi Berra Logic’.

For those unfamiliar with [Yogi Berra](http://en.wikipedia.org/wiki/Yogi_Berra), a brief biographical sketch is in order. Berra is a baseball player and manager. He played for the New York Yankees from 1946-1965 as catcher and outfielder and is generally regarded as one of the best catchers in the history of baseball. He also seems to be both a genuinely good guy as well as an honorable man.

But Yogi Berra is perhaps best remembered these days for a collection of colorful and, at least on the surface, nonsensical sayings called [Yogi-isms](http://en.wikipedia.org/wiki/Yogi_Berra#Quotations). According to the article in Wikipedia, Yogi-isms are ‘…either an apparently obvious tautology or a paradoxical contradiction.’

For the moment, let’s not worry about whether this characterization is correct or whether the Yogi-isms described as [tautologies](http://en.wikipedia.org/wiki/Tautology_(rhetoric)) are rather examples of [circular reasoning](http://en.wikipedia.org/wiki/Circular_reasoning) or of [begging the question](http://afterdeadline.blogs.nytimes.com/2008/09/25/begging-the-question-again/?_php=true&_type=blogs&_r=1) or whether the ones described as paradoxes are in fact not true paradoxes but ironies. It suffices to state that the idea the editor of Wikipedia was trying to convey is that Yogi-isms are pithy and witty but are devoid of much or all meaning. I would like to challenge that assessment.

As an example of a Yogi-ism consider the following [quote](http://www.brainyquote.com/quotes/quotes/y/yogiberra100418.html), attributed to Yogi when he explained why he no longer went to Ruggeri’s restaurant in St Louis:

Nobody goes there anymore. It’s too crowded.

Clearly this quote is funny (if you’re not laughing then stop reading, you don’t have a well-developed sense of humor and will only be harmed by the remainder of this post) but is it nonsensical?

Well that depends on the context and meaning of the words ‘Nobody’ and ‘crowded’. Returning to the ideas of essentials and accidentals, let us view the words ‘nobody’ and ‘crowd’ not as essential terms applying to all humanity (i.e., everybody, nobody, some, few, many) but as accidental names referring to two different subsets or categories of humanity. Next try to match these accidental terms more accurately to the groups that they most likely represent.

Cool = the set of people who Yogi Berra regards as important to him. Maybe they are friends and family. Maybe they are baseball players. Maybe they are authentic Italian people who eat at Ruggeri’s because they like the food.

Jerks = the set of people who Yogi Berra regards as useless or unimportant to him. Maybe they are enemies. Maybe they are pitchers (he is [quoted](http://www.brainyquote.com/quotes/quotes/y/yogiberra139931.html) as saying that ‘all pitchers are either liars or crybabies’). Maybe they are image-conscious Yuppies who began eating at Ruggeri’s when it became known that he ate there and they thought it would be cool to be exposed to ethnic flavor.

Whatever the reasons Yogi had for dividing people up into these two groups, the Yogi-ism can now be translated into

Nobody from the Cool group goes there anymore. It’s too crowded with Jerks.

Of course now the saying makes perfect sense (more precisely it is intelligible) but it isn’t funny at all. In addition, it makes clear a division of people into groups when perhaps the wise and polite thing is to not state this explicitly.

Is this what Yogi meant? Am I attributing to much intelligence to a baseball player? I don’t think so and I am willing to bet that neither do you. As discussed in earlier posts, the common ability of humans to speak and to learn language requires a fundamental capacity to tell essentials from accidentals. Even a baseball player has this ability. Furthermore, being humorous is a clear sign of intelligence and Yogi is clearly funny.

In addition, I am also willing to bet that before you read this analysis, you thought to yourself or muttered under your breath something to the effect ‘Everyone knows what that means. Why are you going on about it?’

But the reason, I am running on about it is that it is difficult to understand how a machine intelligence would be able to parse that Yogi-ism in such a fashion to draw the meaning. How would such an AI determine context? How would it resolve an obvious non sequitur by recognizing that ‘nobody’ refers to one group and ‘crowd’ refers to another? How would it be able to draw from Berra’s background as a baseball player or proud Italian-American to offer reasonable inferences as to who would be in each group?

So at the end of the day, I remain as skeptical of AI taking over has always. I understand that there are theoretical analyses of high sophistication in the science of Artificial Intelligence but I am reminded of another [Yogi-ism](http://www.brainyquote.com/quotes/quotes/y/yogiberra141506.html):

In theory there is no difference between theory and practice. In practice there is.

# A New Turing Test: You Can’t Automate What You Can’t Do Yourself

A common theme that’s been explored in this column for some time is the idea that, at the current state‑of‑the‑art, machine intelligence is clearly inferior to human intelligence to the point that the term machine intelligence should perhaps be regarded as an oxymoron. That isn’t to say that an actual thinking or sentient machine wouldn’t be welcome or that it should be greeted with fear and shunned as an abomination. Rather it is based on a cold eye, unemotional assessment of where we stand and of the huge gap in time and technology that separates us from the apocalyptic stories as featured in the Terminator or in Demon Seed.

As a case in point that illustrates this gap, consider the evolution of machine automation.

The concept of machine automation is nothing new and it dates as far back as man has made machines. But the perceived threat of machine automation as being harmful to mankind seems to have its genesis shortly after the beginning of the [Industrial Revolution](http://en.wikipedia.org/wiki/Industrial_Revolution).

While historians don’t agree on exactly when and how the Industrial Revolution began, it is clear that it got its start in Great Britain sometime around 1780 and eventually completed its spread to most of the Western World by 1840. During that time, machines of all varieties were invented to perform activities that were originally performed by hand. Of course there was a backlash by certain segments of society and perhaps the most notable was by the Luddites in England.

The Luddites were a group of textile artisans who felt threatened by the invention of knitting, spinning and weaving machines that made textile manufacturing accessible by lower skill workers. During the period from about 1811 to 1817, they were known for destroying factory machinery and protesting the encroachment of machines into their economic sphere. And while it is true that these machines probably weakened textile artisan position in society, nowhere do we hear a claim that the machine dislodged the clothing designer. Nor do we hear that the designers of the machine didn’t know how to weave, spin or knit. What we hear is simply that the machines did the same job as the human faster and with fewer errors (and of course, ultimately cheaper) using the techniques developed by human beings.

The idea of machine automation as a threat has waxed and waned over the years. Another example can be found starting in the late 1950s, where workers in office settings felt threatened by the rise of the computer as a business machine. The delightful movie Desk Set, starring Spencer Tracey and Katherine Hepburn, is a comical romp through the fears and realities of machines displacing human beings through automation. Tracey plays the role Richard Sumner, who is a ‘Methods Engineer’ (computer scientist) that has been hired to provide a computer for the research department at a major television network in New York in advance of a corporate merger. Katherine Hepburn plays the role of Bunny Watson (I’m not making that name up), who is the head of the research department whose entire contingent is filled with smart, witty and attractive women. Comic hijinks ensue and Sumner and Watson end up falling in love but there is a very level-headed message that is given in the movie towards the end when Sumner explains that the purpose of his computer is store and retrieve data so as to free the women for tasks best suited for a human being since, as he likes to say, ‘no machine can evaluate.’

Fast forward to today. We don’t just have machines that weave fabric or collate research data. We have machines that act as AI players in video games; that automate complex robotic manufacturing; that print 3-D objects; that control computer updates and traffic lights and billing notice and hundreds of other things. But they only do what we ourselves have taught them to do.

Nowhere is there a record of a single machine inventing a new process, designing a new object, or developing a new idea. True they assist us in all of these tasks but they do so using the well-defined methodologies that we taught them. True that they allow us to comb through vast amounts of data and gain deeper insight than we could have gotten by going through the data by hand. But the patterns they search for and the insights they help bring to light are fundamentally what we put in.

In other words they do what we tell them to do precisely how we told them to do it. That we are sometimes surprised by their results shouldn’t come as a shock. It is well known that any set of rules of reasonable complexity, which are seemingly understandable and sensible on their own, can sometimes produce unforeseen results when they interact. Ask any person harmed by the unintended consequence of a law written by people, administered by people, and acting on people. This doesn’t mean that the law itself somehow became intelligent or achieved sentience but rather we were too busy or too rushed or simply too stupid to think it all the way through.

So I would propose a new type of [Turing test](http://en.wikipedia.org/wiki/Turing_test) to mark the beginning of the era when machines can think on their own. For those who don’t know, the Turing test consists of a remote dialog between a human and a second party that human knows is maybe another human and maybe is a machine. If after some period of time, the human is unable to distinguish that the second party is a machine then the machine has passed the test, successfully mimicking human responses in the dialog.

In my test, I would say that if the human were able to go to the second party with a vague set of requirements for a new thing (a process, a widget, a tool, whatever) and the second party can come back with a design that meets these requirements or explains why they can’t be met then that second party is intelligent, whether or not it is human. When that happens let me know… I would like to hire it.

## Black Swan Science

We are now 55 years removed from the publication of [Karl Popper](http://rationalwiki.org/wiki/Karl_Popper)’s *Logic of Scientific Discovery* (1959), in which Popper argued that falsification is an essential character of science, and it seems that very few of us in society actually embrace this notion to understand how and why they should be skeptical of ‘studies that show…’.

Popper’s argument goes something like this. Before the age of discovery, when explorers from Europe journeyed to the four corners of the globe, Europeans held the belief that all swans were white. This seemed to be a natural conclusion. After all, every swan that had been observed and reported had been white, and it was reasonable to assume that every swan that was, is, or will ever be, is white. On January 10, 1697, the Dutch explorer [Willem de Vlamingh](http://en.wikipedia.org/wiki/Willem_de_Vlamingh) found a habitat sporting a large number of black swans in and around the Swan River on the west coast of Australia. His observation put an end to any validity of the claim that ‘all swans are white’.

In Popper’s logical structure, de Vlamingh’s observations falsified the hypothesis that an essential property of swans is their color. Using this example as a prototype, Popper then generalizes a guiding principle that no scientific theory can be proven, it can only be falsified.

Now, I expect that most educated people in society would reply, if asked, that they are aware that a scientific theory can never be proven, only disproved, and that they accept this as an essential facet of the scientific enterprise. Some of the more knowledgeable may even point out the radical modifications required to Newtonian physics caused by the observations that eventually led to the birth of Quantum Mechanics and General Relativity.

And yet, these very same people seem to blindly believe any ‘fact’ that comes to light, as long as its sales pitch starts with ‘scientist have discovered’. The media continually bombards us with stories of this kind, telling us about how a new study shows that eating this or that raises or lowers the risk of us contracting some horrible malady. That kids who play video games or watch too much TV (however ‘too much’ is defined) are being driven to greater levels of violence or obesity or whatever. Regardless of the subject matter, a vast component of our society is gullible and quite ready to believe as proven cause-and-effect any set of statistical correlations that a scientist may happen to discover in a set of data.

Before I am accused of being unfair and overly critical, I do want to state that I recognize that real life is never as simple as the idealized situation portrayed in the black swan anecdote. For example, imagine ourselves as contemporaries of de Vlamingh who stay at home in Holland. After he arrives home, we happen to be at a meeting where he is presenting his black swan observations. Why should we just give up on the ‘all swans are white’ hypothesis solely on his say so? Perhaps the birds he observed are not actually swans but birds that look similar. Perhaps they were actually white, and a recent fire had covered them with soot. But suppose he came back with the body of a black swan and all our tests and examinations indicate that the bird is indeed black and a swan. Does this mean that we have, with certainty, disproved the white swan hypothesis? I think the answer is a qualified yes. That is to say, we can no longer cling to the notion that all swans are white even if we later narrow the definition of ‘swan’ so that the hypothesis again becomes acceptable.

On the surface, it may seem that the preceding argument invalidates Popper’s approach, and that this whole enterprise is self-contradictory. But some careful thought and identification of what is essential versus what is accidental in the scientific method assures us that we are on firm ground.

The essential aspects of the scientific enterprise is that we can believe that the world is understandable and that logic and the scientific method work as tools to reach this understanding. These beliefs are meta-physical in that they rise above the accidents of any particular scientific hypothesis, theory, or test, and they are not ‘provable’ or ‘disprovable’. We simply identify them as essential aspects of the world and how we interact with it. The accidental aspects are all that remains.

To try to illustrate this, let’s return for the final time to these annoying black swans and to de Vlamingh, who caused so much trouble. The essential aspect in this historical narrative is that Europe held the belief in white swans based on a very large number of observations. To hold a belief about the world is to tacitly assume that the world is understandable and that reason is a tool to understand it. The accidentals of the narrative are that 1) prior to de Vlamingh’s observations all swans were white and 2) after his observations that belief could no longer be held unchallenged or unmodified. The fact that we can argue whether the de Vlamingh’s birds are really swans or really black or whatever is only a discussion about the accidentals of the bird and not the essentials of understanding the world.

So what I am criticizing in society is not that people can be confused about how to draw conclusions from scientific data, nor am I criticizing them for drawing conclusions I would not. What I am criticizing is the acceptance of scientific conclusions without skepticism. I am criticizing misplaced faith, which focuses on the accidental observations of a given study, and loses sight of the essential unprovable nature of the scientific method. I worry about a society that uncritically accepts the term ‘Settled Science’ and turns its back on Black Swan Science.

## Philosophy, Immanuel Kant, and Murder Mysteries – Part 1

I suppose that the genesis of this post comes from one of my current study projects. Over the past several months I’ve been slowly working my way through Harry Gensler’s really fine book ‘An Introduction to Logic’, 2nd edition. As is the case when I learn anything, find that my mind automatically associates many things with many things. It seems to me a good strategy because I remember the information much better and can apply it with greater ease. (This should be contrasted to the way I was taught or learned history – I still don’t know what the Battle of Hastings was, why I should care, and how it effects my life.)

Anyway, Chapter 3 of Gensler’s book deals with definitions and what is essentially epistemology, although I don’t believe that Gensler ever mentions that term explicitly. The most interesting part of that discussion is the presentation of the categories of definition attributed to Immanuel Kant and how they mesh with the two philosophical divisions of knowledge that are traditionally recognized.

Kant divides definitions into two categories:

Analytic statements: Statements whose subject contains its predicate or are self-contradictory to deny

Synthetic statements: Statements that are neither analytic nor are self-contradictory

Traditionally, philosophers recognize two kinds of knowledge which are defined as:

*A posteriori* knowledge: Empirical knowledge based on sense experience

*A priori* knowledge: Rational knowledge based solely on intellect

No doubt a few examples are in order to make these concepts clearer. The examples that Gensler provides (and which I believe an anonymous Wikipedia contributor lifted without attribution) tend to feature the noun ‘bachelor’.

Examples of analytic and synthetic statements are:

All bachelors are unmarried. (analytic)

Daniel is a bachelor. (synthetic)

The first statement is analytic since its subject ‘bachelors’ is synonymous with ‘unmarried’ (that is to say that its subject contains its predicate as an attribute) while the second statement is clearly synthetic since the word ‘Daniel’ is not synonymous with ‘bachelor’ nor is it self-contradictory as it would be if ‘Daniel’ were replaced by ‘Stacey’ (assuming the usual genders denotations of names).

The following statements are examples of *a posteriori* and *a priori* knowledge

Some bachelors are happy. (a posteriori)

All bachelors are unmarried. (*a priori*)

The first piece of knowledge that ‘some bachelor are happy’ can only be obtained by us going out, meeting bachelors and determining (through whatever mechanism we like) that they are happy. The second bit of knowledge is based on our ability to see the essential definition of the word bachelor.

Obviously there is an extremely close tie between a statement being analytic and a piece of knowledge being *a priori*. There is also a very close tie between a synthetic statement and a piece of *a posteriori* knowledge (but I would argue not as close as the association between analytic and *a priori*). Thus there is a tendency in philosophy to equate the two terms in each case and to say that all statements of *a priori* knowledge are analytic and that all statements of *a posteriori* knowledge are synthetic.

This seems to be a natural conclusion and one may dismiss the idea that some statements of *a priori* knowledge can be synthetic or that some statements of *a posteriori* knowledge can be analytic. This dismissal is also supported, at least superficially, by the common notion that all of our mathematics is *a priori* knowledge and all of our science is based on *a posteriori* knowledge.

The problem arises when one starts to examine certain statements that, while not quite self-referential, fall into a category where they at least talk about each other, or more precisely they are statements that explicit talk about the nature of knowledge.

As a possible example of an analytic statement of *a posteriori* knowledge consider the sentence ‘the value of pi is about 3% larger than 3’. That there is a constant of proportionality between the diameter and the circumference of a circle is certainly an analytic statement of *a priori* knowledge but the determination of the actual value (or some decimal approximation to it) is not. Okay, so maybe there is such a thing as an analytic statement of *a posteriori* knowledge, although Gensler leaves the door open for doubt when he says

“But perhaps any analytic statement that is known *a posteriori* also could be known *a priori*”

But apparently the real drama in the philosophical world (I must admit I have fanciful images of Plato and Aristotle, dressed in wrestling tights, as squaring off in a steel-cage match) is over whether there is credible evidence to support the claim of a synthetic statement of *a priori* knowledge. Such a statement Q would be one such that Q is neither self-contradictory to affirm nor to deny, Q is true, and we know Q to be true only using our reason.

Trying to further explain where such a brain-twisting idea can arise, Gensler asks us to consider two types of philosophers: empiricists and rationalists. According to his discussion, the empiricist denies the possibility of synthetic *a priori* knowledge while the rationalist admits such a possibility. The crux seems to come in the examination of the empiricist’s point of view. The first observation is that an empirical point of view seems to equate the experiences of the senses with the actualities of the world. An empiricist is inclined to say something like

“I perceive an object to be red therefore it is a red object.”

Of course the empiricist seems to have no mechanism for embracing the idea that an object is actually red when it is perceived as red except to resort to what seems to be synthetic *a priori* knowledge. It is synthetic because nothing in how the terms are defined requires that an object that is perceived as red to actually be red. It is *a priori* because we use our reason to conclude that it is a tenable assumption that all object perceived as red are, indeed, red.

Perhaps even more interesting is the position that empiricist takes on synthetic *a priori* knowledge in the first place. To say

“There is no such thing as synthetic *a priori* knowledge”

seems to be an example of synthetic a priori knowledge, at least in-so-far as one is willing to agree that the statement, if true, is not true by virtue of the definition of the terms ‘synthetic’ and ‘*a priori*’ and is therefore synthetic and that the statement, if true, cannot be determined to be so by our sense experiences and so must be *a priori*.

Okay, so what does any of this have to do with murder mysteries? Well, as I mentioned above, whenever I am learning something I employ a personal strategy of associating things I understand with things I am trying to grasp. As I was reading Genler’s presentation, I couldn’t help but wonder how mystery writers employ these points to amuse, entertain, and sometimes baffle us.

So next time, I will apply some of these concepts to some of the world’s most famous fictional detectives. We’ll have a chance to see if Sherlock Holmes is synthetic or analytic. We’ll ask how many of Hercule Poirot’s little gray cell depend on *a priori* versus *a posteriori* knowledge. We’ll examine whether Miss Marples understanding of human nature springs from analytic *a posteriori* knowledge. And we’ll explore how logic, reason, and epistemology figure into two to the twentieth centuries most philosophical writers G.K. Chesteron and Umberto Eco through their excellent characters of Father Brown and Brother William of Baskerville.

## Philosophy, Immanuel Kant, and Murder Mysteries – Part 2

In the [last post](http://aristotle2digital.blogwyrm.com/2015/01/09/philosophy-immanuel-kant-and-murder-mysteries-part-1/) we discussed the epistemological divisions in philosophy between *a priori* and *a posteriori* knowledge and the divisions due to Kant between the notions of analytic and synthetic statements. As a brief reminder, *a priori* knowledge stems from first principles and can be understood using the human capacity to grasp the essential nature of things. *A posteriori* knowledge is obtained only after examining a thing and coming to a conclusion about its nature – a conclusion that cannot be grasped by reason alone. An analytic statement is one which is true and in which the subject contains the predicate (that is to loosely say that one defines the other) while a synthetic statement is one that is neither false nor is analytic.

On the surface there seems to be such a strong tie between *a priori* knowledge and analytic statements, on one hand, and between *a posteriori* knowledge and synthetic statements, on the other, that there is a temptation to equate the two concepts in each case. Thus one might want to say that all statements of *a priori* knowledge are analytic and all statements of *a posteriori* knowledge are synthetic.

But as is usually the case with logic when examined very carefully, ideas that seem rock-solid based on a casual examination become a lot more uncertain when looked at more thoroughly. However, these kinds of abstract examinations are often dry. So for this post we’ll try to apply these ideas to the popular medium of the murder mystery.

What should be said about the murder mystery? I think that if Aristotle were alive today one of his favorite past times would be reading and/or writing murder mysteries. This should come as no surprise since Aristotle is credited with formalizing logic and logic and solving mysteries go hand-in-hand. The murder mystery, or detective story as it also called (not all the crimes are murders – only the most enjoyable ones), are individual studies in epistemology. At its heart is the idea of pronouncing a statement of truth; of disclosing ‘whodunnit’.

Consider the analysis of G. K. Chesterton, one of the twentieth century’s most profound thinkers and prolific authors, who penned dozens of works on analysis, philosophy, and social criticism. Chesterton, who was home with logic and critical thinking in its many forms, was particularly fond of the detective story and [wrote often about it](http://www3.dbu.edu/mitchell/chesterton_on_detective_fiction.htm). One of his notable observations was:

The essence of a mystery tale is that we are suddenly confronted with a truth which we have never suspected and yet can see to be true.

Rex Stout, the author of over 70 detective stories, had the following very nice description of the detecting process. Speaking about his gourmand and rotund detective Nero Wolfe, Archie Goodwin, Wolfe’s assistant, has this to say about his boss’s moments of genius:

I knew what was going on, something was happening so fast inside of him and so much ground was being covered, the whole world in a flash, that no one else could ever really understand it even if he had tried his best to explain, which he never did. Sometimes, when he felt patient, he explained to me and it seemed to make sense, but I realized afterward that that was only because the proof had come and so I could accept it. I said to Saul Panzer once that it was like being with him in a dark room which neither of you has ever seen before, and he describes all of its contents to you, and then when the light is turned on his explanation of how he did it seems sensible because you see everything there before you just as he described.

If a detective story is an individual study in epistemology then it should be possible to examine each detective in terms of their where they fall in the division between *a priori* and *a posteriori* knowledge and between analytic and synthetic statements of truth. In this way, maybe we can shed some light on the thornier sides of this debate and also have some fun doing it.

Before examining some of the great literary detectives, let me state that none of them are purely one way or another. There is no author of detective fiction (at least not one I would want to read) who would believe that crime can be solved purely by thinking about the world from first principles nor who would believe that crime can be solved solely by the dry gathering of facts. It is the interplay between the two extremes that is the engine of discovery and truth detection. Nonetheless, each these detectives leans, as does the author who sits behind their adventures, more towards one extreme or another.

We can envision a categorization scheme for detectives where each is placed on a two-dimensional grid. To the left is the extreme of the synthetic and to the right the extreme of the analytic. At the bottom is *a posteriori* knowledge whereas at the top is *a priori* knowledge. An empty grid looks like

and placing a detective in the top right means that he depends more heavily on analytic *a priori* methods to solve crime than by other means.

Our task is then to debate, and argue, and wrestle with where to place each. I won’t pretend to have a well-conceived and impregnable argument for what I present below. Rather I offer it as food for thought and, perhaps, the basis of some really enjoyable discussions with family and friends.

The easiest place of start is with Sherlock Holmes. For this discussion, I will be dealing only with Holmes in his original incarnation as conceived of by Since Sir Arthur Conan-Doyle and not some of the more modern adaptations. The sleuth of 221B Baker Street often solved the mysteries confronting him through observations correlated with dry or obscure facts. Red clay from a particular quarry in northern England combined with an encyclopedic knowledge of the British Rail time tables were a more common route to the solution than ponderings about human nature. Thus we can classify him as predominantly as synthetic and *a posteriori*.

The two famous creations of Agatha Christie, Hercule Poirot and Miss Marple, are cut from a decidedly different cloth. Both of these sleuths depended heavily on their knowledge of human nature and often worked from motive to solution. Clearly they are both analytic, but it seems to me that Poirot starts more from well-articulated first principles and methodical deduction than his female counterpart. Poirot can explain exactly how he arrived at his conclusions (consider his ‘mentoring’ of Doctor Sheppard in the *Murder of Roger Ackroyd*) even if he often won’t and he needs only the bare facts to proceed (*The Disappearance of Mr. Davenheim*). In contrast, Miss Marple relies on a lifetime spent examining human nature ‘under a microscope’ in her village of St. Mary Mead. As she explains in Sir Henry Clithering, her knowledge is akin to an Egyptologist who, due to a lifetime handling Egyptian scarabs, can tell when one is genuine while another is a cheap knockoff even if he can’t explain how. She often jumps to the solution and then gathers or reconciles facts only latter (*Death by Drowning*). Thus I would be inclined to place Poirot in the analytic and *a priori* sector and Miss Marple in just below him somewhat in the *a posteriori* square.

Nero Wolfe, already mentioned above, is more difficult to place. He seems to slide back and forth between the extremes, having the greater fluidity early on in Stout’s writing. In some cases, he is clearly synthetic in his approach. Consider *Fer De Lance*, where he asks a golf club salesman to demo how to swing a club to confirm his suspicions about the delivery method of a poison dart or *The Rubber Band*, where he realizes a connection between two usages of the word ‘rubber’ to impeach the murderer’s alibi. In other cases, including *The Christmas Party* and *Death of a Doxy*, he relies solely on his understanding of human nature and his ability to play upon a murderer’s irresistible compulsion to force a conviction. I place him nearly equally balanced between analytic and synthetic and tipping more towards *a posteriori* than *a priori*.

The final two detectives I’ll discuss both happen to be Roman Catholic priests: Father Brown the creation of G. K. Chesteron and Brother William of Baskerville from Umberto Eco’s brilliant novel *The Name of the Rose*. There is some irony here in that Chesterton was a devout catholic and Eco is a self-declared atheist. Nonetheless, both detectives depend on their training in philosophy (with particular emphasis on Thomas Aquinas) and the intellectual and theological traditions of the Catholic Church to find solutions to their mysteries. Father Brown is deeply logical and staunch defender of reason (*The Blue Cross*) but is prone to inspired deductions where, as Chesteron puts it (*The Queer Feet*):

…in that instant he had lost his head. His head was always most valuable when he had lost it. In such moments he put two and two together and made four million. Often the Catholic Church (which is wedded to common sense) did not approve of it. Often he did not approve of it himself. But it was real inspiration -- important at rare crises -- when whosoever shall lose his head the same shall save it.

In contrast, Brother William seems to take a more measured approach. On one hand he is quite proud and comfortable in his use of logic as in the affair of Brunellus the horse as he and Adso, his novice, approached the unnamed abbey where the bulk of the book is set. At other times, he seems to despair of ever knowing anything or, at least, anything with certainty as in his explanation to Adso of how he got the right answer using from the wrong approach. (An aside: the whole discussion associated with penetrating and navigating the labyrinth is delightful reading and worth studying).

All things considered, I tend to plop Father Brown down into that controversial region where synthetic *a priori* knowledge sits and I place Brother William firmly in the center.

My final diagram looks like:

Obviously, I’ve ignored a host of beloved literary detectives, including C. Auguste Dupin, Perry Mason, Ellery Queen, Lord Peter Wimesy, and Sam Spade. Leave a comment telling where on the diagram you placed your favorites and why.

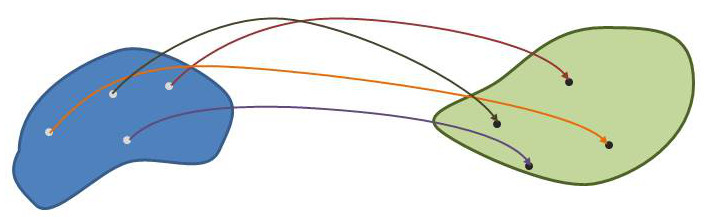
## The Language of Infinity

“How language shapes thought and thought shapes language” is an age old question in linguistics and philosophy. I’m not in any position to give a definitive answer, nor, I suspect, is anyone else. Having taught math and physics at the university level, I am willing to offer some thoughts about how the language of mathematics and the symbols and glyphs used to turn mathematical concepts into written words shape how people think and solve problems.

In this blog I will be focusing in the concept of infinity and the philosophic implications that come from using it. But before I get to infinity directly, I would like to discuss, by way of a warm-up exercise, how the use of the symbol ‘x’ throws off a lot of beginning students.

When describing a function or mapping between two sets of real numbers, without a doubt, the most common notation that teachers use is to allow the symbol ‘*x*’, called the independent variable, to be any member of the initial set and the symbols ‘*y*’ and ‘*f(x)*’ to be the corresponding member of the target set and the function that generates it. The symbolic equation ‘*y = f(x)*’ becomes so rigidly fixed in some students minds, that the idea that the symbols ‘*x*’ or ‘*y*’ could be replaced with any other symbol, say ‘*y*’ and ‘*z*’, never occurs to them. I myself have experiences of students coming and asking if their book has a typo when it asks them to solve ‘*x = f(y)*’ or integrate ‘*f(y) dy*’ or the like (once this happened while I was out to dinner at Olive Garden with my family – but that is a story for another day).

There is no easy way to fix this problem as there is a kind of catch 22 in the teaching of mathematics. One on hand, the mapping between sets exists as a pictorial relation between ‘clouds’ and ‘the points within them’



without the need for written glyphs. One the other hand, an initial set of well-defined symbols keeps initial confusion to a minimum and allows the student to focus on the concepts without all the possible freedom of choice in notion getting in way. (Note: a reader comfortable with classic philosophy may point out that the pictorial that a mapping between sets can be abstracted even further, perhaps to the notion of a Platonic form, but this is a side issue.)

Okay, with the appetizer firmly digesting in our minds, let’s turn to perhaps the most confusing symbol in all of mathematics, the symbol for infinity, ‘∞’. This symbol, which looks like the number ‘8’ passed out after a night of heavy drinking, seduces students and instructors alike into all sorts of bad thoughts.

How does it have this power, you may ask? Well its very shape, small and compact and slightly numberish, encourages our minds to treat it like all other numbers. There are literally countless examples of infinity masquerading as a plain number much like a wolf in sheep’s clothing. One of the most egregious examples is the innocent looking expression

where ‘’ is compared side-by-side with the number ‘0’. There is perhaps more palatable way of writing the integral as

but it still looks like the number ‘a’ can be thought of as becoming or approaching ‘’. A seasoned practitioner actually knows that both expressions are really shorthand for something much more involved. I will summarize what this more involved thing is in one short and sweet sentence. *Infinity is a process that you have the freedom to perform as many times as you like*. Or even shorter: *Infinity is an inexhaustible process*.

Take a moment to think that through. Are you back with me? If you don’t see the wisdom in that maxim consider either form of the integral expression listed above. In both cases, what is really being said is the following. Pick an upper bound on the integral (call it ‘a’ to be consistent with the second form). Evaluate the integral for that value of ‘a’. Record the result. Now increase ‘a’ a bit more, maybe by doubling it or multiplying by 10 or however you like, as long as it is bigger. Now evaluate the integral again and record the result. Keep at it until one of several things has happened: 1) the difference in the recorded values has gotten smaller than some threshold, 2) you run out of time, or 3) you run out of patience and decide to go do something else. The term infinity is simply meant to say that you have the freedom to decide when you stop and you also have the freedom to resume whenever you like and continue onwards.

If you are new to calculus, you will no doubt find this short sentence definition somewhat at odds with what your instructors have told you. Where are all the formal limits and precise nomenclature? Where is all the fancy machinery? If you are an old hand at calculus you may even be offended by the use of the words ‘process’, ‘freedom’, or ‘inexhaustible’. But this sentiment is exactly at the heart of the Cauchy delta-epsilon formalism and the casual nomenclature has the advantage of ruthlessly demolishing the ‘high-brow’ language mathematics to bring what is really a simple idea back to its rightful place an everyday tool in the thinking person’s toolbox.

On the other hand you may be thinking that everyone knows this and that I am making a mountain out of a mole hill. If you fall into that camp, consider this video about the properties of zero by the Numberphiles.

I must admit I like many of the Numberphile’s videos but this one made me shake my head. They allowed language to affect their thinking and they were seduced by the evil camouflage powers of infinity. They go to great trouble to explain why you can’t divide by zero and they note that people say “isn’t dividing by zero just infinity?” and that they point out it isn’t that simple.

The problem is it is that simple! Dividing by zero is infinity as understood by the maxim above. The Numberphiles prove this fact themselves. At about a minute into the video, one of their lot begins to explain how multiplication is ‘glorified addition’ and division is ‘glorified subtraction’. The argument for ‘glorified subtraction’ goes something like this.

If one wishes to divide 20 by 4, then one keeps subtracting 4 until one is lefts with a number smaller than 4 (in this case zero). The number of times one engages in this subtraction process is the answer with whatever piece left over being the remainder. So dividing the number 17 by 5 is a short hand for subtracting 5 from 17 three times and finding that one has 2 leftover. So one then says 17/5 = 3 with a remainder of 2.

The same bloke (I use bloke because of their cool English or Australian or whatever accents), then says that 20 divided by 0 goes on forever because each time you subtract 0 you are left with 20. Here then is the inexhaustible process that lives at the very heart of infinity. Sadly, while he looks like he is about to hit the bulls eye (at 2:20 he even says infinity isn’t a number) his aim goes horrible awry at the last moment when he objects to saying that the expression ‘1/0 = ’ can’t be true because one could then go on to say ‘1/0 = = 2/0’ from which one can say ‘1=2’.

This is, of course, a nonsensical objection since the expression ‘1/0 = ’ is a short hand for saying ‘the glorified subtraction of 0 from 1 (in the sense used above) is an inexhaustible process.’ It is no more meaningful to say that this process is the same as the ‘glorified subtraction of 0 from 2’ as it is to say that ‘1/0’ is the same as any other inexhaustible process like halving a non-zero number until you reach zero.

The fact that the words ‘0’, ‘1’, and ‘ and the sentence ‘1/0 = ’ result in a illogical conclusion is an important warning about the power of language to shape thought. The Numberphile guys had all the right ideas but they came up with a wrong result.

## It Isn’t Elementary

I suppose that this post grew out of a recent opportunity I had to re-watch the [M\*A\*S\*H television series](http://en.wikipedia.org/wiki/M*A*S*H_(TV_series)). One particular episode, entitled [*The Light That Failed*](http://aftermash.blogspot.com/2009/08/episode-126-light-that-failed.html), finds the army surgeons and nurses of the 4077th unit suffering a brutal Korean winter and desperately low on supplies. The supply truck soon arrives bearing a cargo more suited for a military unit in the Guam or the Philippines and not mainland Asia. As the troops are left wondering what they can do with mosquito netting, ice cream churns, and swim fins when the temperature is hovering around freezing, someone notices that one of the doctors, B J Hunnicutt, has received a very rare object – a book.

And this is not just any book, but a murder mystery by the famed writer Abigail Potterfield called the *Rooster Crowed at Midnight*. Either out of the goodness of his heart or out of desire to end the constant nagging (probably both), B J decides to tear portions of the book out so that it can circulate throughout the camp. As the old saying goes, no good deed goes unpunished, and soon he discovers that the last page of the book, in which all is revealed, is missing. And thus begins the great debate as to who committed the murders, how they did it, and why?

The team comes up with many answers, all of which are first widely embraced as the solution and then scuttled when someone gives a counterexample that pokes a hole in the theory. They eventually place a long distance phone call to the author herself, now residing in Australia, to get the answer. But even this authoritative voice doesn’t quell the skepticism. Shortly after they ring off, Col. Potter, the commanding officer, points out that Abigail Porterfield’s own solution can’t be true. The episode closes with the delivery of the much-needed supplies and some comic hijinks but with no satisfactory explanation as to who the culprit was.

I was in middle-school when I first saw that episode and it left a lasting impression on me. For many years I carried misconceptions about mystery stories and I wondered why anyone would ever read them. In particular, I held a very skewed idea about deductive reasoning and what can and cannot be determined from the evidence. With the perspective of years (really decades) I am both happy and disappointed to say that I was not alone in my poor understanding of what logic and reason are capable of achieving.

Let’s talk about deductive logic first. The basic idea behind deductive logic is that the conclusion is infallible if the premises are true. It is a strong approach to logic since it argues from first principles that apply to a broad class of set of object and from these narrows down a conclusion about a specific object. In a pictorial sense, deductive logic can be thought of in terms of Venn diagrams. If we want to conclude something about an object we simply need to know into what categories or classes this object falls and we will be able to exactly conclude something about it by noting where all of the various categories to which it belongs intersect.

Deductive reasoning is unfortunately also limited by the fact that we are not born with nor does anyone have the universe’s user manual that spells out in detail what attributes each object has and into what categories they may be grouped. So the standard objections that are raised in deductive logic fall squarely on disagreements about the truth of one of more premises.

For example, the syllogism

* All men are mammals
* George is a man
* Therefore George is a mammal

is a logically correct deduction, since the conclusion follows from the premises and it is true since the premises are true (or at least we regard them to be true). The syllogism

* + - All men are mammals
    - George is a mammal
    - Therefore George is a man

is invalid, even though its premises are true, since it argues from the specific to the general. In contrast the syllogism

* All white birds are man-eaters
* All swans are white birds
* Therefore all swans are man-eaters

is perfectly valid, since the conclusion follows from the premises, but is not true since neither premise is true (or so I hope!).

All of this should be familiar. But what to make this syllogism (B J’s syllogism) made by B J Hunnicutt in the episode

* Lord Cheevers was murdered in the locked library in Huntley Manner
* Randolf had motive for murdering Lord Cheevers
* Randolf played in Huntley manner as a child
* Randolf would have known if there were secret passages in Huntley Manner
* Therefore Randolf was the murderer.

Is this really a deduction? According to the novel, the first three premises are true. The fourth premise is certainly plausible but is not necessarily true. How then should we feel about the conclusion? What kind of logic is this if it is not deductive? Suppose we knew that Randolf was the murder (e.g. we caught him in the act) what can we infer about the fourth premise?

Before answering these questions, consider what would happen if we modify the argument a bit to simplify the various possibilities. The syllogism (B J’s syllogism revised) now reads

* Lord Cheevers was murdered in the locked library in Huntley Manner
* Randolf had motive for murdering Lord Cheevers
* Randolf played in Huntley manner as a child
* Randolf knew there was a secret passage from the study to the library
* Randolf was seen entering the empty study just before the murder
* Therefore Randolf was the murderer.

This argument is certainly a stronger one than the first one proffered but isn’t really conclusive. But again how should we feel about the conclusion? What kind of logic is this?

In both cases, we know that the conclusion is not iron-clad; that is doesn’t necessarily follow from the premises. But just like those fictional characters in M\*A\*S\*H, we are often faced with the need to draw a conclusion from a set of premises that do not completely ‘nail down’ an unequivocal conclusion.

The type of logic that deals with uncertainty falls under the broad descriptions of inductive and abductive reasoning. Inductive reasoning allows us to draw a plausible conclusion ‘B’ from a set of premises ‘A’ without ‘B’ necessarily following from ‘A’. Abductive reasoning allows us to infer the premise ‘A’ based on our knowledge that outcome ‘B’ has occurred.

In the M\*A\*S\*H examples given above, B J’s revised syllogism is an example of inductive reasoning. All the necessary ingredients are there for Randolf to have committed the crime but there is not enough evidence to inescapably conclude that he did. We can infer that Randolf is the killer but we can’t conclude that with certainty.

B J’s original syllogism is a lot more complicated. It involved elements of both inductive and abductive reasoning. If we believe Randolf is guilty, we might then try to establish that there were secret passages in Huntley Manner that connected the lock library to some other room in the mansion. We would then have to also establish, maybe through eyewitness testimony, that Randolf knew of the passages (e.g. an old servant recalls showing it to a young Randolf). Even still, all we would be doing is establishing the premises with more certainty. The conclusion of his guilt would still not necessarily follow. If on the other hand we knew that he was guilty, perhaps he was seen from someone looking into the library from outside, we might abductive infer that there was a secret passage and that Randolf knew of its existence.

Now if the

## The Dangers of Being Equal

One of the favorite themes of this blog is how language and reasoning affect each other, sometimes to the detriment of both. The overlap between logical reasoning and mathematical language is particularly ripe with possibilities of confusion because the way certain concepts are used contextually. In an earlier post I discussed the seductive properties of the humble symbol . A far more deadly symbol is the highly overloaded glyph described by two parallel horizontal lines – the equal sign ‘=’.

There are so many contextual uses of the equal sign that it is hard to know where to start. And each and every one of them is sinister to the untrained. Like some kind of bizarre initiation ritual, we subject students of all stripes to this ambiguous notation and then we get frustrated when they don’t grasp the subtle distinctions and shaded nuances of meaning that we take for granted. This situation closely parallels the experiences many of us have had learning how to swim, or ride a bike, or ice skate, or drive. Those of us who know how to do something often can’t remember how hard it is to learn when you don’t know.

Of course, this situation is not unprecedented in language. A simple internet search using the search string ‘word with the most definitions’ returns the [following statement from Dictionary.com](http://dictionary.reference.com/help/faq/language/t47.html)

"Set" has 464 definitions in the *Oxford English Dictionary*. "Run" runs a distant second, with 396. Rounding out the top ten are "go" with 368, "take" with 343, "stand" with 334, "get" with 289, "turn" with 288, "put" with 268, "fall" with 264, and "strike" with 250.

So functionally overloading a word with numerous meanings, some of them very closely related and some of them quite distinct, is commonplace.

What makes the equal sign so frustrating is that it is mostly applied in highly technical fields where shades of meaning in thought can have large implication in outcomes. Consider the differences in meaning in the following equations

and

and

Each of them tells us something about the irrational number but in very different ways. In the first equation, we think of as the assigned value for the correlation between the diameter of a circle (and its circumference (. This concept is purely geometric and can be explored with rulers and compasses and pieces of paper. In some sense, it can even be regarded as a causative relation, telling us that the if we make a circle of radius then we are making an object whose perimeter is a distance . The second equation is an identity in the purest sense of that term. It boldly states that one of the many disguises of is an algebraic expression involving the natural logarithm and the imaginary number . The final equation is neither an assignment nor an identity but a set of instructions saying ‘if you want to know how to compute to some accuracy, then set up a computing process that takes the first integers and combines them in this funny way.’

The science of computing has long recognized that the usual ambiguity of human language would be inadequate for machine instructions. All programming languages that I’ve been exposed to clearly distinguish between the concepts of assignment, equivalence, and function definition. Using the pi-equations above, one might express them in the programing languages *Python*, and [*Maxima*](http://maxima.sourceforge.net/) as

|  |  |  |
| --- | --- | --- |
| Pi-equation | Python | Maxima |
|  | pi = C /(2 \*r) | pi : C/(2\*r) |
|  | pi == ln(i\*\*(-2\*i)) | pi = ln(i\*\*(-2\*i)) |
|  | def sum\_sqr(n):  sum = 0  for i in range(1,n+1):  sum = sum + 1.0/(i\*i)  return temp  def approx\_pi(n):  sum = sum\_sqr(n)  return (6\*sum)\*\*(0.5) | calc\_pi(n) := block([sum],  sum : 0,  for i: 1 thru n do  sum : sum + 1/(i\*i),  ans : sqrt( 6 \* sum ) ); |

Note that in each case there is a clear syntactical difference between assignment (‘=’ or ‘:’), the conditional test for identity (‘==’ or ‘=’), and functional definition (‘def…’ or ‘:=’). For anyone who’s been programming for some time, switch back and forth between these ideas of assignment, equivalence, and definition is relatively effortless but for the beginner it is one of the hardest concepts he will have to learn.

The situation is even more complex in the physical sciences for two primary reasons. First, and foremost, because man had been investigating the physical world longer than he has been writing computer programs. As a result, there has been more time for man to layer different meanings and subtle distinctions. Second, computers are intrinsically stupid and require a high degree of precision and clarity to function. A nice discussion of this last point can be found in the prolog of the book [Functional Differential Geometry](http://www.amazon.com/Functional-Differential-Geometry-Gerald-Sussman/dp/0262019345/ref=sr_1_4?s=books&ie=UTF8&qid=1423104141&sr=1-4&keywords=Jack+Wisdom) by Sussman and Wisdom.

As an example, let’s look at perhaps the most famous physical statement – Newton’s second law. Many people, even those lacking formal training in science, know that the expression of the law is ‘force equals mass times acceleration’ or in mathematical terms

But what does the equal sign here mean. The concept of a force tells us that it is a vector quantity that transforms like a position vector. That means that a force relationship in the same in all frames. For example the balancing of the pulls from three ropes tied to an object such that the object doesn’t move is an equilibrium condition that is independent of the frame in which it is expressed. An accelerating observer will make the same conclusion as an inertial observer. So the force on the left-hand side of ‘f=ma’ is geometric in its meaning.

On the other hand, we understand that the acceleration appearing on the right-hand side is kinematic. It describes an objects motion and it the kind of thing measured with rulers and clocks. It is fundamentally frame dependent when described by an accelerating observer. Just imagine the visual perception of someone on a merry-go-round. The mass, which measures the object’s unwillingness to move under influence of a force, simply scales the acceleration and can be regarded as constant.

So how do we reconcile what the equal sign is meaning here? On one side is a geometric quantity as immutable and placid as a mountain. The other side is as ephemeral as rising mist or running water, flowing to and fro. How can they actually be equal?

Well, the answer is that the equal sign should be regarded as relating cause-and-effect. If we regard the force as known (e.g. Newton’s universal law of gravity), then the equal sign allows us to deduce the resulting motion once the force is applied. If we regard the acceleration as known (e.g., we film the motion and do a frame analysis) we can infer the force that caused it.

So

## Aces High

This week’s column has a three-fold inspiration. First off, most of the most exciting and controversial philosophical endeavors always involve arguing from incomplete or imprecise information to a general conclusions. These inferences fall under the heading of either inductive or abductive reasoning and most of the real sticking points in modern society (or any society for that matter) revolve around how well a conclusion is supported by fact and argument. The second source comes from my recent dabbling in artificial intelligence. I am currently taking the edx course CS188.1x and the basic message is that the real steps forward that have been taking shape in the AI landscape came after the realization was made that computers must be equipped to handle incomplete, noisy, and inconsistent data. Statistical analysis and inference deals, by design, with such data and its use allows for an algorithm to make a rational decision in such circumstances. The final inspiration came from a fall meeting of my local AAPT chapter in which Carl Mungan of the United States Naval Academy. His discussion of the [two aces problem](http://usna.edu/Users/physics/mungan/_files/documents/Scholarship/TwoAces.pdf) was nicely done and I decided to see how to produce code in Python that would ‘experimentally’ verify the results.

Before presenting the various pieces of code, let’s talk a bit about the problem. This particular problem is due to Boas in her book [*Mathematical Methods for the Physical Sciences*](http://www.amazon.com/Mathematical-Methods-Physical-Sciences-Mary/dp/0471198269/ref=sr_1_1?ie=UTF8&qid=1424050699&sr=8-1&keywords=boas+physics) and goes something like this.

What is the probability that when being dealt two random cards from a standard deck, that the two cards will be two aces? Suppose that one of the cards is known to be an ace, what is the probability? Suppose that one of the cards is known to be the ace of spades, what is the probability?

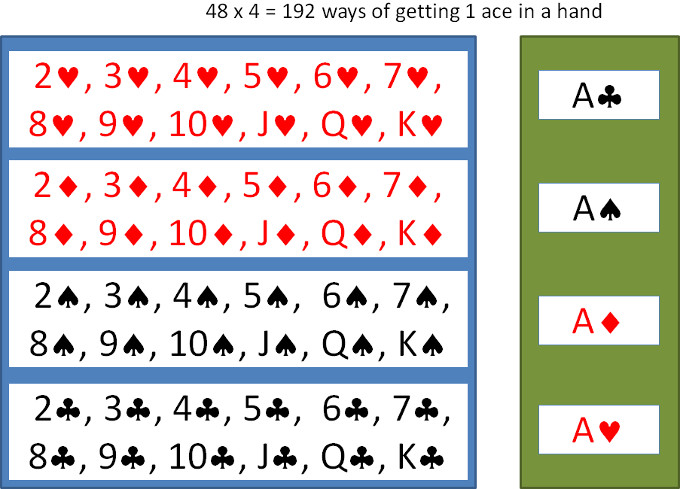
The first part of this three-fold problem is very well-defined and is relatively simple to calculate. The second two parts require some clarification of the phrase ‘is known to be’. The first possible interpretation is that the player separates out the aces from the deck and either choses one of them at random (part 2) or he chooses the ace of spades (part 3). He returns the remaining 3 aces to the deck, which he subsequently shuffles prior to drawing the second card. I will refer to this method as the ***solitaire option***. The second possible interpretation (due to Mungan) is that a dealer draws two cards at random and examines them both while keeping their identity hidden from the player. If one of the cards is an ace (part 2) or is the ace of spades (part 3), the dealer then gives the cards to the player. Parts 2 and 3 of the question then ask for the probability that the hands that pass this inspection step actually have two aces. Since this last process involves assistance from the dealer, I will refer to it as the ***assisted option***. Finally, all probabilities will be understood from the frequentist point-of-view. That is to say that each probability will be computed as a ratio of desired outcomes to all possible outcomes.

With those preliminary definitions out of the way, let’s compute some probabilities by determining various two-card hand outcomes.

First let’s calculate the number of possible two-card hands. Since there are 52 choices for the first card and 51 for the second, there are 52x51/2 = 1326 possible two-card hands. Of these, there are 48x47/2 = 1128 possible two-card hands with no aces since the set of cards without the aces is comprised of 48 cards. Notice that in both of these computations we divide by 2 to account for the fact that order doesn’t matter. For example, a 2-of-clubs as the first card and a 3-of-diamonds is the second is the same hand as a 3-of-diamonds as the first card and a 2-of-clubs as the second.



Likewise, there are 48x4 = 192 ways of getting a hand with only 1 ace. Note that there is no need to divide by 2 since the two cards are drawn from two different sets.



Finally, there are only 6 ways to get a two-ace hand. These are the 6 unique pairs that can be constructed from the green set shown in the above figure.

As a check, we should sum the size of the individual set and confirm that it equals the size of the total number of two-card hands. This sum is 1128 + 192 + 6 for no-ace, one-ace, and two-ace hands and it is totals 1326, which is equal to the size of the two-card hands. So the division into subsets is correct.

With the size of these sets well understood, it is reasonably easy to calculate the probabilities asked for in the Boas problem. In addition, we’ll be in a position to determine something about the behavior of the algorithms developed to model the ***assisted option***.

For part 1, the probability is easy to find as the ratio of the all possible two-ace hands (6 of these) to the number of all possible two-card hand (1326 of these). Calculating this ratio gives 6/1326 = 0.004525 as the probability of pulling a two-ace hand from a random draw from a well shuffle standard deck.

For parts 2 and 3 in the ***solitaire option***, the first card is either given to be an ace or the ace of spades. In both cases, the probability of getting another ace is the fraction of times that one of the three remaining aces is pulled from the deck that now holds 51 cards. The probability is then 3/15 or 0.05882.

The answers for the ***assisted option*** for parts 2 and 3 are a bit more subtle. For part 2, the dealer assists by winnowing the possible set down from 1326 possibilities, associated with all two-card hands, to only 192 possibilities, associated with all one-ace hands, plus 6 possibilities, associated with all two-ace hands. The correct ratio is 6/198 = 0.03030 as the probability for getting a single hand with two aces when it is known that one is an ace. For part 3, the dealer is even more zealous in his diminishment of the set of possible hands with one card the ace-of-spades. After he is done, there are only 51 possibilities of which 3 are winners and so the correct ratio is 3/51 = 0.05882 as the probability of getting a single hand with two aces when it is know that one is the ace-of-spades.

All of these probabilities are easily checked by writing computer code. I’ve chosen python because it is very easy to perform string concatenation and to append and remove items from lists. The basic data structure is the deck, which is a list of 52 cards constructed from 4 suites (‘S’,’H’,’D’,’C’), ranked in traditional bridge order, and 13 cards (‘A’,’2’,’3’,’4’,’5’,’6’,’7’,’8’,’9’,’10’,’J’,’Q’,’K’). Picking a card at random is done importing the random package and using random.choice, which returns a random element of a list passed into it as an argument. Using the random package in the code turns the computer modeling into a Monte Carlo simulation. For all cases modeled, I input the number of Monte Carlo trials to be N = 1,000,000.

Codes to model part 1 and parts 2 and 3 for the ***solitaire option*** are easy to implement and understand so I don’t include them here. The Monte Carlo results (10 cases each with N trials) certainly support the analysis done above but if one didn’t know how to do the combinatorics one would only be able to conclude that the results are approximately 0.0045204 +/- 0.00017.

The code to model parts 2 and 3 for the ***assisted option*** is a bit more involved because the dealer (played by the computer) has to draw a hand then either accept or reject it. Of the N Monte Carlo trials drawn what percentage of them will be rejected? For part 2, this amounts to determining what the ratio of two-card hands that have at least one ace relative to all the possible two-card hands. This ratio is 192/1326 = 0.1448. So, roughly 85.5 % of the time the dealer is wasting my time and his. This lack of economy becomes more pronounce when the dealer rejects anything without an ace-of-spades. In this case, the ratio is 51/1326 = 1/52 = 0.01923 and approximately 98% of the time the dealer is throwing away the dealt cards because they don’t meet the standard. In both cases, the Monte Carlo results support the combinatoric analysis with 0.03031 +/- 0.0013 and 0.05948 +/- 0.0044.

Notice that the uncertainty in the Monte Carlo results grows larger in part 2 and even larger in part 3. This reflects the fact that the dealer only really affords us about 150,000 and 19,000 trials of useful information due to the rejection process.

Finally there are a couple of philosophical points to touch on briefly. First, the Monte Carlo results certainly support the frequentist point-of-view, but they are not actual proofs of the results. Even more troubling is that a set enumeration, such as given above, is not a proof of the probabilities either. It is a compelling argument and an excellent model but it presupposes that the probability should be calculated by the ratios as above. Fundamentally, there is no way to actually prove the assumption that set ratios give us the correct probabilities. This assumption rests on the belief that all possible two-card hands are equally likely. This is a very reasonable assumption but it is an assumption none-the-less. Second, there is often an accompanying statement along the lines that the more that is known the higher the likelihood of the result. For example, knowing that one of the cards was an ace increased the likelihood that both were an ace by a factor of 6.7. While true, this statement is a bit misleading, since, in order to know, the dealer had to face the more realistic odds that 82 percent of the time he would be rejecting the hand. So, as the player, our uncertainty was reduced only at the expense of a great deal of effort done on our behalf by another party. This observation has implications for statistical inference that I will explore in a future column.

## Language and Metaphor

Why do we quote movies? I often think about this fascination we have as a culture to repeat and relive moments from our favorite films. Large number of clips exist on YouTube devoted to capturing that magical moment when a character utters an unforgettable line. But why? What is it that drives us to remember a great line or identify with the person who uttered a quote?

This question came up over the dinner table one night and as I reflected on this question, my thoughts were drawn to an explanation about speed reading that I had once come across. The author went to great trouble to make the point that that the trick to speed reading was to see groups of words in a sentence as a single chunk of information.

To understand that point, consider the words you read in a sentence. To make it concrete, let’s take the word ‘understand’. When you read the word ‘understand’ you are seeing a group of 11 individual letters, but are you really conscious of each letter at a time? Do you really see a ‘u’ followed by an ‘n’ followed by a ‘d’ and so on? No. What each of us with any sophistication in reading accomplishes is to perceive these 11 letters as a unit conveys the word ‘understand’. Of course, this is why we can read a sentence with a misspelling quite comfortably and often we may not even notice.

This idea of chunking and clumping comfortably scales upward to where common phrases can be consumed with a glance without a detailed examination of the individual words. Two common approaches are to either abbreviate the phrase into some short form. Expressions like ‘LOL’ or ‘BTW’ or any of the other text and chat speak concepts are excellent examples. The other approach is pick lyrical or repetitive expressions to encapsulate a phrase. The famous ‘rosy-fingered dawn’ from the Homeric epics or ‘ready, set, go!’ are examples of these kind, respectively.

But there is an even more compelling approach – the concept of the metaphor. The idea here being that we liken an object to another object, one with well-known properties. The properties of the second object are then inherited by the first simply by the equating of the name. Some examples of this include the following sentences.

* ‘That guy’s Benedict Arnold for leaving our team for that other one.’
* ‘How was the date last night Romeo?’
* ‘Stay away from her, she’s Mata Hari through and through!’
* ‘That man is death itself’.

So I think that our desire to quote movies is indicative of this. That by repeating the dialog of the movie, the quote itself becomes a cultural metaphor for the feelings and experiences expressed in the movie. This idea was brilliantly explored in the Star Trek: The Next Generation episode Darmok.

In this episode, the crew of the Enterprise are coaxed into a meeting with a Tamarian ship in orbit around a mostly unexplored planet. Despite their universal translator, no ship of the Federation had ever been able to crack the Tamarian language. The words themselves could be translated but the syntax and context seemed to be utterly devoid of meaning. The Tamarian captain, Dathon, seeing that this meeting was no different from previous encounters and that the ordinary avenues for communication were not working arranged for himself and Captain Picard to be trapped on a the planet.

On that planet, both captains were confronted by a dangerous creature. This shared danger spurred a meeting of the minds and eventual understanding dawned on Picard. The Tamarian race thought and communicated in metaphors. They would say statements like ‘Tember his arms wide’ to mean the concept of giving or generosity. Back on the Enterprise, the crew had also come to a similar epiphany. By analogy, they constructed a Tamarian-like was of expressing romance by saying ‘Juliet on her balcony’, but they lamented that without the proper context, in this case Shakespeare’s tragic play about Romeo and Juliet, one didn’t know who Juliet was and why she was on her balcony.

The episode closes with Dathon dying due the wounds inflicted by the creature and with Picard arriving back aboard the Enterprise just in time to make peace between the Tamarians and the Federation by speaking their language.

<iframe width="420" height="315" src="https://www.youtube.com/embed/ANvlLcOTy6M" frameborder="0" allowfullscreen></iframe>

The episode left some dangling ideas. How do the Tamarians specify an offer that involves a choice between many things or how an abstract idea, like giving someone his freedom, would be expressed. Nonetheless, it was a creative and insightful way of exploring how powerful metaphor can be and how abstracted can be the thought that lies behind it.

So the next time you quote a movie, give a thought to the metaphor that you are tapping into and take a moment to marvel at the miracle of speech and thought.

## Finding the Socratic Method

There is a standard debate in mathematics about the application of the terms ‘invention’ versus ‘discovery’. It resurfaced the other day when a colleague and I were talking about some mathematical graffiti that adorned a door jamb in the conference room in which we were meeting. This graffiti took the form of some mathematical symbols printed on a magnet held in place at the top part of the door. None of us in the room were able to determine what, if any, message was being sent but in the process of discussing the possible meaning, my colleague said, in passing, that Pythagoras had invented the theorem that bears his name. I questioned whether the verb should have been ‘discovered’ rather than ‘invented’. We spent a few minutes discussing that point then we gave up altogether and went our separate ways. On the drive home that evening I began to think about the proper use of those two words and I finished up wondering if Socrates invented or discovered his famous method.

To understand this distinction between ‘invent’ and ‘discover’, let’s return to the Pythagorean Theorem for a moment. Most everyone knows the theorem, namely that the sum of the squares of the lengths of sides of a right triangle is equal to the square of the length of the hypotenuse. The most common proof and, in my opinion, the most elegant, draws a right triangle with squares of the appropriate area on a side. I’ve provided such a diagram for a 3-4-5 Pythagorean triangle below.

<insert triangle diagram>

The proof proceeds by laying the bigger of the two side squares onto the hypotenuse square

<insert overlay>

and then adding up the remaining area and showing that it is equal to the area of the smaller square

<insert summation>

Of course, the steps shown above were quickly done in PowerPoint graphics and there is no reason for a skeptic to actually accept them as proof. But the doubters can go at this ‘proof’ with whatever vigor they desire. The answer will always be the same: a^2 + b^2 = c^2.

And that brings us to the point about invention versus discovery. I would argue that the Pythagorean Theorem is an exercise in discovery. That finding that all right triangles satisfy it is exactly the point – we find or discover that all right triangles obey this relationship.

Contrast this Edison’s invention of the light bulb. I say invention because of a number of factors. First there is the particular form of the object in question. The base, with its two contacts, one at the bottom center and one on the periphery, is choice of form that could have been done many other ways. The shape of the bulb itself is only a suggestion of what could be done given the state of art of glass blowing at the time of its introduction to society. Second there is the particular design and implementation. The placement, current and voltage running through the filament were all carefully chosen to meet specific goals or requirements. The materials that comprise all the parts were chosen to provide the maximum economy based on availability and convenience in pre-existing manufacturing processes. So I would argue that while it is proper to say that laws and properties of electricity were discovered that the light bulb is a true invention.

So far so good, but what about the Socrates? Did he discover or invent the Socratic Method? As quick review, a few brief words are in order about what I mean when I say the [Socratic Method](http://en.wikipedia.org/wiki/Socratic_method) (ironically, if you follow the link to the Wikipedia article you’ll find that both ‘invention’ and ‘discover’ are used in describing Socrates’s contribution).

The Socratic Method is best explained by the Platonic dialog called [Euthyphro](http://classics.mit.edu/Plato/euthyfro.html). In this dialog, we find Socrates and Euthyphro both showing up at the Athenian court but for very different reasons. Socrates is answering a call by the court to make account of his ‘criminal ways’ whereas Euthyphro intends on petitioning the court to bring a charge of murder against his father for the death of a slave in his possession. The two meet on the steps leading inside and exchange with each other their reasons for being there. Socrates expresses surprise that Euthyphro is accusing his father of murder since the slave in question died from being imprisoned for murdering another slave. Euthyphro says that he is compelled to this course of action due to his piety. That’s all the prompting needed by Socrates and soon the two are engaged in a discussion where Socrates asks a question like ‘what is piety’ and Euthyphro attempts to answer with a response like ‘what is pleasing to the gods’. After each answer, Socrates questions some new part of the response as a way a sharpening the reasoning behind the response.

The Socratic Method is a way of examining the logical content of a statement by carefully examining the basic notions that make up that statement. So asking what do ‘piety’, ‘pleasing’ and ‘gods’ mean is a way of finding the truth. Generally, when the method is applied, we are more apt to find out what a particular concept, like ‘piety’, isn’t rather than finding out what it is. Most of the dialogs (and for that matter modern applications) end with both parties departing before the full meaning has been established but at least with a clearer picture of what is not meant.

So with all the preliminaries out of the way, the key question to grapple with is whether Socrates invented this method or discovered it. My vote is for discovery. I say this mostly because of the universal nature of this mode of inquiry but partly because Socrates believed in Truth, in the most absolute sense. If he invented this type of intellectual exploration then the application of it would be necessarily limited to those contexts where its design matched a particular cast of mind or cultural milieu. The fact that it is a successful philosophical pursuit the world over is testament to its ability to transcend the accidentals of human culture. The fact that it was fashioned with the goal of discovering Truth through logic and reason and that Socrates believed in Truth leads me to believe that he would agree that he discovered his method.

I am willing to say that Plato invented the particular encounters presented in the dialogs and that Socrates invented the accidental trappings whereby he applied his method to the Athenian society. Recognizing these inventions is equivalent to recognizing the invention of the particular symbols for denoting an algebraic quantity as ‘a’ or multiplication by ‘x’ or equality as ‘=’. Writing c\*c = a\*a + b\*b versus saying ‘the square of the hypotenuse is equal to the sum of the squares of the two sides’ are two different, invented ways for expressing the same truth.

# States, Abstraction, and Implicit Objects

This post grew out of a conversation in which I recently participated, about the correct way to computationally model a system. The conversation covered the usual strategies of identifying the system’s states and how to abstract the appropriate objects to hold or contain the states based on the question at hand. But what was most interesting were the arguments put forward about the importance of recognizing not only the explicit objects used in the modeling but also the implicit objects that were ignored.

By way of a general definition, an implicit object is an object whose existence is needed in the computer modeling of the system in question but one that is not represented in code as either a real or virtual class. I don’t think this definition is used in computer science or programming books but it should be. This is a point that is almost entirely ignored in the field but is vital for a correct understanding of many systems.

A simple example will go a long way to clarify the definition and to also raise awareness that implicit objects are used more often than people realize. The system to be modeled is the motion of traffic on a highway. For simplicity, we’ll focus on two cars, each traveling along a straight stretch of road. Generalizations to more complicated situations should be obvious but don’t add anything to the discussion. A picture of the situation is

Here we have a blue car and a green car moving to the right as denoted by the matching colored arrows. The typical abstraction at this point is to create a car class with appropriate attributes and member functions for the questions we want to answer. Let’s suppose that we are interested in modeling (with an eye towards prevention) the collision between the two cars. What ingredients are needed to perform this modeling?

A collision between two objects occurs, by definition, when they try to occupy the same space at the same time. Unpacking this sentence leads us to the notion of position for each car as a function of time and a measure of the physical extent of each car. Collision occurs when the relative position between the centers of the two cars is such that their physical extents overlap.

Now, real life can be rather complicated, and this situation is no different. The physical shape of a car is generally hard to model with many faces, facets, and extrusions. And the position of the car is a function of what the driver perceives, how the driver reacts, and how much acceleration or deceleration the driver imposes.

In time-honored fashion, we will idealize the cars as having a simple two-dimensional rectangular shape as shown in the figure, and will also assume that the cars have a simple control law (which we don’t need to specify in detail here except to say that it depends on the position, velocity, and time) that gives the velocity of the car at any instant of time. With these simplifying assumptions, the trajectory of the car is completely specified by giving its position at some initial time and then using the control law to update the velocity at each time step. The position is then updated using the simple formula $${\vec x}(t+dt) = {\vec x}(t) + {\vec v} dt$$.

With all these data specified, we might arrive at a class definition for the car object that looks like:

At this point, we’ve already encountered our first two implicit objects.

The first one is a coordinate frame, comprised of an origin and an orientation of the coordinate axes, which tells us what the position of the object is measured against. This coordinate frame is an object in its own right and, in many circles, worthy of study. Here it is simply important to acknowledge that it exists and that the width and length parameters need to be defined consistently with regards to it. This is an important point, since a change in coordinate frame orientation will change the test (to be described later) for the geometric condition of overlap (i.e., collision).

The second implicit object is much more subtle and it took the discovery of the principles behind special relativity to drive that point home. Both cars share a common clock to describe their motion. That is to say that their velocities are only meaningfully compared if their clocks are synchronized and ticking at the same speed. The role of the implicit clock is usually carried out by the main program of the simulation but, since each object holds its own time, care must be exercised to keep them synchronized. It is possible to make this common-clock assumption stronger by eliminating time from the list of member data, but that approach also has its drawbacks, as will be touched on below.

The final question is, how do we determine if the cars collide somewhere during the simulation? There are two possible approaches. The first is that we give each car a new member function, called detect\_collision, which takes as arguments the position of the other car and its size. This is not an unreasonable approach for simple situations, but it quickly becomes unwieldy should we later want to add fidelity to the simulation. For example, if we wanted to allow for the body of the car changing its orientation as it changes lanes, then the number of parameters that we have to hand into the detect\_collision function has to increase. A more robust way to do this is to perform this computation at the main workspace in the simulation. If this method is chosen, we’ve now introduced a third implicit object. It’s not quite clear from the context what to call this object, but it’is toch notion may seem like a trivial call this object, but it' computation at the main workspace in the simulation. If ts function is absolutely clear – it communicates to the world the message that the two cars have collided.

This last notion may seem like a trivial observation leading to an unnecessary complication but consider the case where each car is equipped with some sensor that warns of an eminent collision. Remote sensing requires some model of the medium between the sensor and thing being sensed. If the collision detector uses sound, then the air between the cars must be modeled as an elastic medium, with a finite propagation speed of the signal, and variations in this medium due to temperature, humidity, and the like may need also need to be modeled. If the collision detector uses electromagnetic waves (radar or laser ranging), then the index of refraction and dispersion properties of the air become the key parameters that again may depend on the thermodynamics of the atmosphere. In either case, this communicating medium now must be promoted from an implicit object to an explicit one.

The lesson, I hope, is clear. It is just as important to pay attention to the implicit objects in a simulation as it is to the explicit ones. There is a parallel here with the spoken word – where it is often as important to pay attention to what was left unsaid as it is to what was.

# Turing, Gödel, and the Universe

Something about the book ‘Five Golden Rules: Great Theories of 20th-Century Mathematics and Why They Matter’ by John L. Casti caught my eye the other day at the library. On a whim, I signed the book out and started reading. Living up to the promise of the title, the book had five chapters, each one devoted to one of the great math theories of 20th century. All in all, I had been exposed, at least superficially, to all the topics covered so I wasn’t sure what I would get out of the book other than some additional insight into material I already knew or confirmation of my wisdom in staying away from topics that I didn’t.

Anyway, I am quite glad that the mechanism of providence pointed me at this book because the connections that Casti draws are worth thinking about. None of these connections was as profound for me as the deep link that he explores in Chapter 4 entitled ‘The Halting Theorem (Theory of Computation)’.

In this chapter, Casti first presents the concept of the universal Turing machine (UTM) as a mechanism for attacking the question of [Decision Problem (or Entscheidungsproblem)](http://en.wikipedia.org/wiki/Entscheidungsproblem) proposed by David Hilbert in 1928. Casti then couples this presentation with a discussion of the famous [Gödel Incompleteness Theorem](http://en.wikipedia.org/wiki/G%C3%B6del%27s_incompleteness_theorems).

I’m simply a beginner in these fields and Casti omits important details and glosses over a variety of things to help the novice. From this point-of-view, I can’t recommend his work. But I am grateful and excited about the perspective he provided by making this linkage.

To understand my excitement, let me first try to fill in some of the background as best as I can.

Hilbert’s Decision Problem asks the following question. Given a set of input data and an algorithm for manipulating said data, is there a way to know if the algorithm will be able to make a yes/no decision about the input. For example, if the input data is a set of axioms in a logical system and some corresponding assertion in the same system and the algorithm is a logical formalism, such as the classical syllogism, will the algorithm be able to prove or disprove the assertion as either true (yes) or false (no)?

While the Decision Problem might seem straightforward enough to talk about it at a seminar, when it comes to actually tackling the question there are some vague conceptions that need further elaboration. Specifically, what does the term ‘algorithm’ really mean and what constitutes a workable algorithm seem to have been the murkiest part when the problem was first posed.

Turing chose to clarify the algorithm concept by the invention of the machine which bears his name. The UTM is a basic model of computation that is often used to understand the theory behind computer programming. It is also the avenue that allowed Turing to a way to tackle Hilbert’s Decision Problem. The basic ingredients for a UTM are a tape or strip of paper, divided into squares, a set of symbols (usually ‘0’ and ‘1’) that can be written into the squares, and a box that has a mechanism for moving the strip left or right, a head that reads from and writes to the strip and that contains an internal state that allows the UTM to select an action from a pre-defined set, given the internal state and current symbol being read. A typical graphical representation of an UTM looks like

The UTM can represent the input to the decision process by a particular pattern of symbols on the strip. It can also implement the step associated with an algorithm by encoding these also to symbols on the tape. So the entire nature of the Decision Problem comes down to the question as to whether once the UTM starts if it will ever halt; hence the name of the name of the chapter.

Kurt Gödel took a different approach to answering the Decision Problem. He developed a way to map any formal system to a set of numbers, called Gödel numbers. Then by formally manipulating these numbers, he was in position to try to answer Hilbert’s challenge.

Now it is reasonable to suppose that all not every question has a yes or no answer. Questions about feeling, opinion, or taste spring to mind. But I think that most of us expect that questions of mathematics and logic and programming have answers and that we should be able to figure them out. The remarkable thing about the work of both Turing and Gödel is that there are cases where the formal logical system simply can’t decide.

In the language of Turing, there are computational problems for which there is no knowing beforehand if the algorithm will terminate with an answer. In the language of Gödel, no matter is done to a logical system in terms of refining existing axioms or adopting new ones, there will always be truths that are unprovable.

I was quite aware of Godel’s theorem and even wrestled with it for a while when I was concerned about its implications to physical systems. Eventually, I decided that while man’s logic may be limited, nature didn’t need to worry about it because she could make decisions unencumbered by our shortcomings.

I was also quite aware that Turing’s machine was used fruitfully as a model computation. What I was unaware of until reading Casti’s book was the parallel between the conclusions of Gödel and of Turing.

And here we arrive at the source of my excitement. As I’ve already said, I remain convinced that Nature can decide – that is to say that Nature is free of the issues discussed above. And yet, in some capacity Nature is an enormous Turing machine. So why does Nature always make a decision and why does the computation of trajectories, and wave, and quantum evolutions always reach a decision? This is the exciting new idea that has occurred to me and one which I will be exploring in the coming months.

# Gödel’s Theorem – General Notions

As I mentioned in [last week’s blog](http://aristotle2digital.blogwyrm.com/2015/05/09/turing-godel-and-the-universe/), I was encouraged by Casti’s chapter on the Halting Theorem and questions of undecidability in logic and computing. In fact, I was inspired enough that I resolved have another go at studying Gödel’s theorem.

To give some background, many years ago I came across ‘Gödel, Escher, and Bach: An Eternal Golden Braid’ by Douglas Hofstadter. While I appreciate that his work is considered a classic, I found it difficult and ponderous. Its over 700 pages of popularized work did little to nothing to really sink home Gödel’s theorem and the connections to Turing and Church. What’s more, Hofstadter seems to say (it’s difficult to tell exactly as he mixes and muddles many concepts) that Gödel’s work supports his ideas that consciousness can emerge from purely mechanistic means at the lowest level. This point always seemed dodgy to me. Especially since Hofstadter left me with the impression that Gödel’s theorem showed a fundamental lack in basic logic not an emergent property.

For this go around, I decided that smaller was better and picked up the slim work entitled ‘Gödel’s Theorem’ by Nagel and Newman. What a difference a serious exposition makes. Nagel and Newman present all the essential flavor and some of the machinery that Gödel used in a mere 102 pages.

Note that formal logic is not one of my strong suits and the intellectual terrain was rocky and difficult. Nonetheless, Nagel and Newman’s basic program was laid out well and consisted of the following points.

To start, the whole thing was initiated by David Hilbert, whose [Second Problem](http://en.wikipedia.org/wiki/Hilbert%27s_second_problem), challenged the mathematical community to provide absolute proof of the consistency of a formal system (specifically arithmetic) based solely on its structure. This idea of absolute proof stands in contrast to a relative proof where the system is question is put into relation to another system, whose validity and consistency is accepted. If the relation between the two is faithful, then the consistency of the second system carries over or is imparted to the first.

Hilbert was unsatisfied by the relative approach as it depended on some system being accepted at face value as being consistent and finding such a system was a tricky proposition.

The preferred way to implement an absolute proof is to start by stripping away all meaning from the system and to deal only with abstract symbols and a mechanistic way of manipulating these symbols using well-defined rules. The example that Nagel and Newman present is the sentential calculus slightly adapted from the ‘Principia Mathematica’ by Whitehead and Russell. The codification of the formal logic system depends on symbols that fall into two classes: variables and constant signs. Variables, denoted by letters, stand for statements. For example, ‘p’ could stand for ‘all punters kick the football far’. There are six constant signs with the mapping

|  |  |
| --- | --- |
| ~ | Not |
| ∨ | Or |
| ⊃ | If… Then… |
| ⋅ | And |
| ( | Left-hand punctuation |
| ) | Right-hand punctuation |

The idea is then to map all content-laden statements, like ‘If either John or Tim are late then we will miss the bus!’, to formal statements like ( (J ∨ T ) ⊃ B) with all meaning and additional fluff removed. Two rules for manipulating the symbols, the Rule of Substitution and the Rule of Detachment, are adopted and four axioms are used as starting points.

In using this system, one has to sharpen one’s thinking to be able to distinguish between statements in the system (mathematical statements like ‘2 > 1’) from statements about the system (meta-mathematical statements like ‘2>1’ is true or that the ‘>’ symbol is an infix operator). One must also be careful to note the subtle differences between the symbol ‘0’, the number 0, and the concept of zero meaning nothing.

The advantage of this approach is that the proofs are cleaner, especially when there are many symbols. The disadvantage is that it takes time and effort to be able to work in this language.

The consistency of the formal system is shown when there is at least one formal statement (or formula) that cannot be derived from the axioms. The reason for this is complicated and I don’t have a good grasp on it but it goes something like this. The following formula ‘p ⊃ (~ p ⊃ q)’ can be derived in the sentential calculus. If the statements S and ~S are both deducible then any statement you like can be derived from the axioms (via the Rules of Substitution and Detachment) and the system is clearly inconsistent. In the words of Nagel and Newman:

‘The task, therefore, is to show that there is at least one formula that cannot be derived from the axioms.’ [p51]

For the sentential calculus, they point out that the formula ‘p ∨ q’ fits the bill since it is a formula, it doesn’t follow from the axiom. Thus this system is consistent. Note that there is no truth statement attached to this formula. The observation simply means that ‘p ∨ q’ can’t be obtained from the axioms by a mechanical manipulation. They present this argument in their chapter titled ‘An Example of a Successful Absolute Proof of Consistency’.

Well despite that lofty achievement, they go on to show how Gödel took a similar approach and mostly ended any hope for a proof of absolute consistency in formal systems with a great deal more complexity. Gödel used as his symbols the numerals and as his rules the basic operations of arithmetic and, in particular, the arithmetic of prime numbers. Using an ingenious mapping from of the variables and constant symbols to numbers, he not only could encode the structure of the formal system itself, he could also encode statements about the formal system as well (meta-mathematics). In the language of Hofstadter, these are self-referential statements, although Nagel and Newman don’t use this term.

Using this approach, Gödel was able to prove that there is no absolute proof of consistency. At best, the system can say about itself that it is either incomplete or inconsistent. If the system is incomplete, there are true statements that are not part of the axioms and that cannot be derived from them. Enlarging the set of axioms to include them doesn’t work since they presence begets new unprovable truths. If the system is inconsistent, then everything is ‘true’ as discussed above.

Nagel and Newman leave the reader with come final thoughts that are worth contemplation. On the hope that Hilbert’s program can be successfully completed they have this to say

‘These conclusions show that the prospect of finding for every deductive system an absolute proof of consistency that satisfies the [finitistic](http://en.wikipedia.org/wiki/Finitism) requirement’s of Hilbert’s proposal, though not logically impossible, is most unlikely.’

They also comment on the possibility of proving consistency from an outside-looking-in approach using meta-mathematical techniques when they say

‘[W]hether an all-inclusive definition of mathematical or logical truth can be devised, and whether, as Gödel himself appears to believe, only a thoroughgoing philosophical ‘realism’ of the ancient Platonic type can supply an adequate definition, are problems still under debate…’

Finally, they have this to say about artificial intelligence (although that term wasn’t in vogue at the time they published

‘[T]he brain appears to embody a structure of rules of operation which is far more powerful than the structure of currently conceived artificial machines. There is no immediate prospect of replacing the human mind by robots.’

And there you have it. A whirlwind tour of Gödel’s theorem with a surprise appearance of the philosophy from antiquity and the ideas about artificial intelligence.

# Self-Reference and Paradoxes

The essence of the Gödel idea is to encode not just the facts but also the ‘facts about the facts’ of the formal system being examined within the framework of the system being examined. This meta-mathematics technique allowed Gödel to prove simple facts like ‘2 + 2 = 4’ and hard facts like ‘not all true statements are axioms or are theorems – some are simply out of reach of the formal system to prove’ within the context of the system itself. The hard facts come from the system talking about or referring to itself with its own language.

As astonishing as Godel’s theorem is, the concept of paradoxes within self-referential systems is actually a very common experience in natural language. All of us have played at one time or another with odd sentences like ‘This sentence is a lie!’. Examined from a strictly mechanical and logical vantage, how should that sentence be parsed? If the sentence is true then it is lying to us. If it is false, then it is sweetly and innocently telling us the truth. This example of the [liar’s paradox](http://en.wikipedia.org/wiki/Liar_paradox) has been known since antiquity and variation of it have appeared throughout the ages in stories of all sorts.

Perhaps the most famous example comes from the original Star Trek television series in an episode entitled ‘I Mudd’. In this installment of the ongoing adventures of the starship Enterprise, an impish Captain Kirk defeats a colony of androids that hold him and his crew hostage by exploiting their inability to be meta.

<iframe width="420" height="315" src="https://www.youtube.com/embed/wlMegqgGORY" frameborder="0" allowfullscreen></iframe>

There are actually host of paradoxes (or antinomies in the technical speak) that some dwerping around on the internet can uncover in just a handful of clicks. They all arise when a formal system talks about itself in its own language and often their paradoxical nature arises when they talk about something of a negative nature. The sentence ‘This sentence is true,’ is fine while ‘This sentence is a lie.’ is not.

Not all of the examples show up as either interesting but useless tricks of the spoken language or as formal encodings in mathematical logic. One of the most interesting cases deals with libraries of either the brick and mortar variety or existing solely on hard drives and in RAM and FTP packets.

Consider for a moment that you’ve been given charge of a library. Properly speaking, a library has two basic components: the books to read and a system to catalog and locate the books so that they can be read. Now thinking about the books is no problem. They are the atoms of the system and so can be examined separately or in groups or classes. It is reasonable and natural to talk about a single book like ‘Moby Dick’ and to catalog this book along with all the other separate works that the library contains. It is also reasonable and natural to talk about all books written by Herman Melville and to catalog them within a new list with a title perhaps with the name ‘Lists of works by H. Melville’. A similar list can be made with grouping criterion selects books about the books by Melville. This list would have a title like ‘List of critiques and reviews of the works by H. Melville’.

An obvious extension would be to construct something like the following list.

*List of Author Critiques and Reviews:*

* List of critiques and reviews of H. Melville
* List of critiques and reviews of J. R. R. Tolkien
* List of critiques and reviews of U. Eco
* List of critiques and reviews of R. Stout
* List of critiques and reviews of G. K. Chesterton
* List of critiques and reviews of A. Christie
* ….

Since the lists are themselves written works what status do they have in the cataloging system? Should there also be lists of lists? If so, how deep should there construction go? At some point won’t we arrive at lists that have to refer to themselves and what do we do when we reach that point? Should the library catalog have a reference to itself as a written work?

Bertrand Russell wrestled with these questions in the context of set theory around the turn of the 20th century. To continue on with the library example, Russell would label the ‘List of Author Critiques and Reviews’ as a *normal set* since it is a collection of things that doesn’t include itself. He would also label as an *abnormal set*, any list that would have itself as a member – in this case a catalog (i.e. list) of all lists pertaining to the library. General feeling suggests that the normal sets are well behaved but the abnormal sets are likely to cause problems. So let’s just focus on the normal sets. Russell asks the following question about the normal sets: Is the set, R, of all normal sets, itself normal or abnormal? If R is normal, then it must appear as a member in its own listing, thus making R abnormal. Alternatively, if R is abnormal, it can’t be listed as a member within itself and, therefore, it must be normal. No matter which way you start you are led to a contradiction.

The natural tendency is, at this point, to cry foul and to suggest that the whole thing is being drawn out to an absurd length. Short and simple answers to each of the questions posed in the earlier paragraph come to mind with the application of a little common sense. Lists should only be themselves cataloged if they are independent works that are distinct parts of the library. The overall library catalog need not list itself because it primary function is to help the patron find all the other books, publications, and related works in the library. If the patron can find the catalog, then there is no need to have it listed within itself. One the other hand, if the patron cannot find the catalog, having it listed within itself serves no purpose – the patron will need something else to point him towards the catalog.

And as far as Russell and perfidious paradox is concerned, who cares? This might be a matter to worry about if one is a stuffy logician who can’t get a date on a Saturday night but normal people (does this mean Russell and his kind are abnormal?) have better things to do with their lives than worry about such ridiculous ideas.

All told, we should care. This application of common sense is actually quite sophisticated even if we are quite unaware of the subtleties involved. In all of these common-sensical responses there is an implicit assumption about something above or outside. If the patron can’t find the library catalog, well then that is what a librarian is for – to point the way to the catalog. The librarian doesn’t need to be referred to or listed in the catalog. He sits outside the system and can act as an entry point into the system. If there is a paradox in set theory, not to worry, there are more important things than complete consistency in formal systems.

This is concept of sitting outside the system, is at the heart of the current differences between human intelligence and machine intelligence. The later, codified by the formal rules of logic, can’t resolve these kinds of paradoxes precisely because they can’t step outside themselves like people can. And maybe they never will.

# Why do We Teach the Earth is Round?

You’re no doubt asking yourself “Why the provocative title? It’s obvious why we should teach that the Earth is round!” In some sense, this was my initial reaction when this exact question was posed in a round table discussion that I participated in recently. The person who posed the question was undaunted by the initial pushback and persisted. Her point was simply a genuinely honest question driven by a certain pragmatism.

Her basic premise is this. For the vast majority of people on the Earth, a flat Earth model best fits their daily experiences. None of us plan our day-to-day trips using the geometry of Gauss. Many of us fly, but far fewer of us fly long enough distances where the pilot or navigator consciously lays in great circle path. And even if all of us were to fly, say from New York to Rome, so what if the path the plane follows is a ‘geodesic on the sphere’, very few of us are either aware or care. After all, that is someone else’s job to do. And certainly gone are the days where we sit at the seashore and watch the masts of ships disappear last over the horizon – cell phones and the internet are far more interesting.

I listened to the argument carefully and mulled it over a few days and realized that there was a lot of truth in it. The points here weren’t that we shouldn’t teach that the Earth is round but rather that we should know with a firm and articulable conviction why we should teach it and that that criteria for inclusion should be open to debate when schools draw up their curriculum.

So what criteria should be used to construct a firm and articulable conviction? It seems that at the core of this question was a dividing line between types of knowledge and why we would care to know one over the other.

The first distinction in our round-Earth epistemological exploration is one between what I will call *tangible* and *intangible* knowledge. Tangible knowledge consists of all those facts that have an immediate impact on a person’s everyday existence. For example, knowing that a particular road bogs down in the afternoon is a slice of tangible knowledge because acting on it can prevent me from arriving home late for dinner (or perhaps having no dinner at all). Knowing that the rainbow is formed by light entering a water droplet in the atmosphere in a particular way so that it is subjected to a single total internal reflection before exiting the drop with the visible light substantially dispersed is an intangible fact, since I am neither a farmer nor a meteorologist. Many are the people who have said “don’t tell me how a rainbow is formed – it ruins all the beauty and poetry!”

An immediate corollary of this distinction is that what is tangible and intangible knowledge is governed by what impacts a person’s life. It differs both from person to person and over time. A person who doesn’t drive the particular stretch of road that I do would find the knowledge that my route home bogs down at certain times and the meteorologist would find the physical mechanism for the rainbow a tangible bit of knowledge, even if it kills the poet in him.

The second distinction is between what I will call *private* and *common* knowledge. The particular PIN I use to access by phone is knowledge that is, and should, remain private to me. In the hands of others it is either useless (for the vast majority who are either honest, or don’t know, or both) or it is dangerous (for those who do know me and are up to no good). Common knowledge describes those facts that can be shared with no harm between all people. Knowing how electromagnetic waves propagate is an example of common knowledge but knowing a particular frequency to intercept enemy communications is private.

With these distinctions in hand, it is now easy to see what was meant by the original, provocative question. As it is taught in schools, knowledge that the Earth is round is, for most people, a common, intangible slice of human knowledge. In this context, it is reasonable to ask why we even teach it to the students.

A far better course of action is to try to transform this discovery into a common but tangible slice of knowledge that effects each student on core level. The particular ways that this can be done are numerous but let me suggest one that I regard as particularly important.

Teaching that the Earth is round should be done within a broader context of how do we know anything about the world around it, how certain are we, and where are the corners of doubt and uncertainty. A common misconception is that the knowledge that the Earth is round was lost during the Dark and early Middle Ages. The ancient Greeks knew with a great deal of certainty that the Earth was round and books from antiquity tell the story of how Eratosthenes determined the radius of the Earth to an astounding accuracy considering the technology of his day. This discovery persisted into the Dark and Middle Ages and was finally put to some practical use only when the collective technology of the world progressed to the point that the voyages of Columbus and Magellan were possible. Framing the lesson of the Earth’s roundness in this way provides a historical context that elevates it from mere geometry into a societally shaping event. Science, technology, sociology, geography, and human affairs are all intertwined and should be taught as so.

Along the way, numerous departure points are afforded to discuss other facets of what society knows and how does it know it. Modern discoveries that the Earth is not a particularly spherical (equatorial bulge) know take on a life outside of geodesy and the concepts of approximations, models, and contexts by which ‘facts’ are known and consumed now become tools for honing critical thinking about a host of policy decision each and every one of us has to make.

By articulating the philosophical underpinnings for choosing a particular curriculum, society can be sure that arbitrary decisions about what topics are taught can be held in check. Different segments can openly debate what material should be included and what can be safely omitted in an above board manner. Emotional and aesthetic points can be addressed side-by-side with practical points without confusion. And all the while we can be sure that development of critical thinking is center stage.

Failure to do this leaves two dangerous scenarios. The first is that student is filled with a lot of unconnected facts that improve neither his civic participation in practical matters nor his general appreciation for the beauty of the world. The second, and more importantly, the student is left with the impression that science delivers to us unassailable facts. This is a dangerous position since it leads to modern interpretations of science as a new type of religion whose dogma has replaced the older dogma of the spiritual simply by virtue that its magic (microwaves, TVs, cell-phones, rockets, nuclear power, and so on) is more powerful and apparent.

# Ideal Forms and Error

A central concept of Socratic and Platonic thought is the idea of an ideal form. It sits at the base of all discussions about knowledge and epistemology. Any rectangle that we draw on paper or in a drawing software package, that we construct using rulers and scissors, or manufacture with computer controlled fabrication is a shadow or reflection of the ideal rectangle. This ideal rectangle exists in the space of forms, which may be entirely within the human capacity to understand the world and distinguish or may actually have an independent existence outside the human mind, reflecting a high power. All of these notions about the ideal forms are familiar from the philosophy from antiquity.

What isn’t so clear is what Plato’s reaction would be if he were suddenly transported forward in time and plunked down in a classroom discussion about the propagation of error. The intriguing question is would he modify his philosophical thought to expand the concept of an ideal form to include and ideal form of error?

Let’s see if I can make this question concrete by the use of an example. Consider a diagram representing an ideal rectangle of length L and height H.

<insert ideal rectangle figure>

Euclidean geometry tells us that the area of such a rectangle is given by the product

<formula for the rectangle area>

Of course, the rectangle represented in the diagram doesn’t really exist since there are always imperfections and physical limitations. The usual strategy is to not take the world as we would like it to be but to take it as it is and cope with these departures from the ideal.

The departures from the ideal can be classified into two broad categories.

The first category, called knowledge error, contains all of the errors in our ability to know. For example, we do not know exactly what numerical value to give the length L. There are fundamental limitations on our ability to measure or represent the numerical value of L and so we know the ‘true’ value of L only to within some fuzzy approximation.

The second category doesn’t seem to have a universally agreed-upon name, reflecting the fact that, as a society, we are still coming to grips with the implications of this idea. This departure from the ideal describes the fact that at some level there may not even be on definable concept of true. Essentially, the idea of the length of an object is context-dependent and may have no absolutely clear idea at the atomic level due to the inherent uncertainty in quantum mechanics. This type of ‘error’ is sometimes called [aleatory error](http://uqtools.larc.nasa.gov/files/2013/02/NASA_LaRC_MUQ.pdf) (in contrast to epistemic error; synonymous with knowledge error).

Taken together, the knowledge and aleatory errors contribute to an uncertainty in length of the rectangle of dL and an uncertainty in its height of dH.

<insert picture of the error rectangle>

Scientists and engineers are commonly exposed to a model in determining the error in the area of such a rectangle as part of their training to deal with uncertainty and error in a formula sometimes called the propagation of error (or uncertainty). For the case of this error-bound rectangle, the uncertainty in the area is determined also in Euclidean fashion yielding

A’ = (L+dL)\*(H+dH) = L\*H + dL\*H + L\*dH + dL\*dH = A + dA

So the error in the area has a more complicated form that the area itself

dA = dL\*H + L\*dH + dL\*dH

Now suppose that Plato were in the classroom when this lesson was taught. What would his reaction be? I bring this up because although the treatment above is meant to handle error it is still an idealization. There is still a notion of an ideal rectangle sitting underneath.

The curious question that follows in its train is this: is there an ideal form for this error idealization? In other words, is there a perfect or ideal error in the space of forms of which our particular error discussion is a shadow or reflection?

It may sound like this question if predicated on a contradiction but my contention is that it only sounds so, on the surface. In understanding the propagation of error in the calculation of the rectangle I’ve had to assume a particular functional relationship.

It is a profound assumption that the object drawn above (not what it represents but that object itself), which is called a rectangle but which is embodied in the real world as made up of atomic parts (be they physical atoms or pixels), can be characterized by two numbers (L and W) even if I don’t know what values L and W take on. In some sense, this idealization should sit in the space of forms.

But if that is true, what stops us there. Suppose we had a more complex functional relationship, something, say, that tries to model the boundaries of the object as a set of curves that deviate much from linearity but enough to capture a shaky hand when the object was drawn or a manufacturing process with deviations when machined. Is this model not also an idealization and therefore a reflection of something within the space of forms?

And why stop there. It seems to me that the boundary line between what is and is not in the space of forms is arbitrary (and perhaps self-referential – is the boundary between what is and is not in the space of forms itself in the space of forms). Like levels of abstraction in a computer model depend on the context, could not the space of forms depend on the questions that are being asked.

Perhaps the space of forms is as infinite or as finite as we need it to be. Perhaps its forms all the way down.

# Bringing Home the Bacon

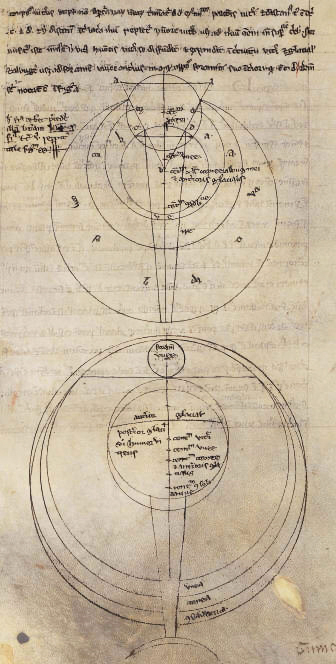
Don’t worry; this week’s entry is not about America’s favorite pork-related product (seriously there exists bacon-flavored candy). It’s about the scientific method. Not the whole thing, of course, as that would take volumes and volumes of text and would be outdated and maybe obsolete by the time it was finished. No, this column is about two men who are considered by science historians to have contributed substantially to the scientific method and the philosophy of science. And it just so happens that both of them bore the last name of Bacon.

[Roger Bacon](https://en.wikipedia.org/?title=Roger_Bacon) was born somewhere around 1214 (give or take – time and record keeping then, as now, was hard to do) in England. Roger became both an English philosopher of note and a Franciscan friar. Most of the best scholastic philosophers of the Middle Ages were monks, and in taking Holy Orders, Bacon falls amongst the ranks of other prominent thinking religious, including [Robert Grosseteste](https://en.wikipedia.org/wiki/Robert_Grosseteste), [Albert Magnus](https://en.wikipedia.org/wiki/Albert_Magnus), [Thomas Aquinas](https://en.wikipedia.org/wiki/Thomas_Aquinas), [John Duns Scotus](https://en.wikipedia.org/wiki/John_Duns_Scotus), and [William of Ockham](https://en.wikipedia.org/wiki/William_of_Ockham).

It seems that the cultural milieu of that time was planting the intellectual seeds for the scientific and artistic renaissance that followed. Roger Bacon cultivated modes of thought that would be needed for the advances to come. Basing his philosophy on Aristote, he advocated for the following ‘modern’ ideas:

* Experimental testing for all inductively derived conclusions
* Rejection of bling following of prior authorities
* Repeating pattern of observation, hypothesis, and testing
* Independent corroboration and verification

In addition, he wrote extensive on science, both its general structure and on specific applications. Among his particular fields of interest was optics, where his diagrams have the look and feel of the modern experimental lab notebook.



He also criticized the Julian day and argued for dropping a day every 125 years. This system would not be adopted until about 300 years after his death with the creation of the Gregorian calendar in 1582. He was almost an outspoken supporter experimental science saying that it had three great prerogatives over other sciences and arts in that:

* It verifies all of its conclusions by direct experiment
* It discovers truths which can’t be reached without observation
* It reveals the secrets of nature

[Francis Bacon](https://en.wikipedia.org/wiki/Francis_Bacon) was born in 1561 in England. He was a government official (Attorney General and Lord Chancellor) and a well-known philosopher. His writings on science and philosophy established a firm footing for inductive methods used for scientific inquiry. The details of the method are collectively known as the [Baconian Method](https://en.wikipedia.org/wiki/Baconian_method) or the scientific method.

In his work [Novum Organum](https://en.wikipedia.org/wiki/Novum_Organum) (literally the new Organon referring to Aristotle’s work on metaphysics and logic), Francis has this to say about induction:

Our only hope, then is in genuine Induction... There is the same degree of licentiousness and error in forming Axioms, as in abstracting Notions: and that in the first principles, which depend in common induction. Still more is this the case in Axioms and inferior propositions derived from Syllogisms.

By induction, he meant the careful gathering of data and then refinement of a theory from those observations.

Curiously, both Bacons talk about four errors that interfere with the acquisition of knowledge: Roger does so in his [Opus Majus](https://en.wikipedia.org/wiki/Opus_Majus); Francis in his Novum Organum. The following table makes an attempt to match up each’s list.

|  |  |
| --- | --- |
| Roger Bacon’s Four Causes of Error | Francis Bacon’s Four Idols of the Mind |
| Authority  (reliance on prior authority) | Idols of the Theater  (following academic dogma) |
| Custom | Idols of the Tribe  (tendency of humans to see order where |
| Opinion of the unskilled many | Idols of the Marketplace  (confusion in the use of language) |
| Concealment of ignorance behind the mask of knowledge | Idols of the Cave  (interference from personal beliefs, likes, and dislikes) |

While not an exact match, the two Baconian lists of errors match up fairly well, which is puzzling if historic assumption that Francis Bacon had no access to the works of Roger Bacon is true. Perhaps the most logical explanation is that both of them saw the same patterns of error; that human kind doesn’t change its fundamental nature in the passage of time or space. Or perhaps Francis is simply the reincarnation of Roger, an explanation that I am sure William of Occam would endorse if he were alive today ☺.

# Bayesian Inference – The Basics

In last week’s article, I discussed some of the interesting contributions to scientific method made by the pair English Bacons, Roger and Francis. A common and central theme to both of their approaches is the emphasis they placed on performing experiments and then inferring from those experiments what the logical underpinning was. Put another way, both of these philosophers advocated inductive reasoning as a powerful tool for understanding nature.

One of the problems with the inductive approach is that in generalizing from a few observations to a proposed universal law one may overreach. It is true that in the physical sciences, great generalizations have been made (e.g. Newton’s universal law of gravity or the conservation of energy) but these have ultimately rested on some well-supported philosophical principles.

For example, the conservation of momentum rests on a fundamental principle that is hard to refute in any reasonable way; that space has no preferred origin. This is a point that we would be loath to give up because it would imply that there was some special place in the universe. But since all places are connected (otherwise they can’t be places) how would nature know to make one of them the preferred spot and how would it keep such a spot inviolate?

But in other matters, where no appeal can be made to an over-arching principle as a guide, the inductive approach can be quite problematic. The classic and often used example of the black swan is a case in point. Usually the best that can be done in these cases is to make a probabilistic generalization. We infer that such and such is the most likely explanation but by no means necessarily the correct one.

The probabilistic approach is time honored. William of Occam’s dictum that the simplest explanation that fits all the available facts is usually the correct one is, at its heart, a statement about probabilities. Furthermore, general laws of nature started out as merely suppositions until enough evidence and corresponding development of theory and concepts led to the principles upon which our confidence rests.

So the only thorny questions are what are meant by ‘fact’ and ‘simplest’. On these points, opinions vary and much argument ensues. In this post, I’ll be exploring one of the more favored approaches for inductive inference known as the Bayesian method.

The entire method is based on the theorem attributed to [Thomas Bayes](https://en.wikipedia.org/wiki/Thomas_Bayes), a Presbyterian minister, and statistician, who first published this law in the latter half of the 1700s. It was later refined by Pierre Simon Laplace, in 1812.

The theorem is very easy to write down and that perhaps is what hides its power and charm. We start by assuming that two random events, $$A$$ and $$B$$, can occur, each according to some probability distribution. The random events can be anything at all and don’t have to be causally connected or correlated. Each event has some possible set of outcomes $$a\_1, a\_2, \ldots$$ and $$b\_1, b\_2, ldot$$. Mathematically, the theorem is written as

where $$a\_i$$ and $$b\_j$$ are some specific outcomes of the events $$A$$ and $$B$$ and $$P(a\_i|b\_j)$$ ($$P(b\_j|a\_i)$$ is called the conditional probability that $$a\_i$$ ($$b\_j$$) results given that we know that $$b\_j$$ ($$a\_i$$) happened. As advertised it is nice and simple to write down and yet amazingly rich and complex in its applications. To understand the theory, let’s consider a practical case where the events $$A$$ and $$B$$ take on some easy-to-understand meaning.

Suppose that we are getting ready for Christmas and want to decorate our tree with the classic strings of different-colored lights. We decide to a purchase a big box of bulbs of assorted colors from the Christmas light manufacturer, Brighty-Lite, who provides bulbs in red, blue, green, and yellow. Allow the set $$A$$ to be color

On its website, Brighty-Lite proudly tells us that they have tweaked their color distribution in the variety pack to best match their customer’s desires. They list their distribution as consisting of 30% percent red and blue, 25% green, and 15% yellow. So the probabilities associated with reaching into the box and pulling out a bulb of a particular color are

The price for bulbs from Brighty-Lite is very attractive, but being cautious people, we are curious how long the bulbs will last before burning out. We find a local university that put its undergraduates to good use testing the lifetimes of these bulbs. For ease of use, they categorized their results into three bins: short, medium, and long. Allowing the set $$B$$ to be the lifetime

the student results are reported as

which confirmed our suspicions that Brighty-Lite is doesn’t make its bulbs to last. However, since we don’t plan on having the lights on all the time, we decide to buy a box.

After receiving the box and buying the tree, we set aside a weekend for decorating. Come Friday night we start by putting up the lights and as we work we start wondering whether all colors have the same lifetime distribution or whether some colors are more prone to be short-lived compared with the others. The probability distribution that describes the color of the bulb and its lifetime is known as the joint probability distribution

If the bulb color doesn’t have any effect on the lifetime of the filament, then the events are independent and the joint probability of say a red bulb with a medium lifetime given by the product of the probability that the bulb is red and the probability that it has a medium lifespan (symbolically $$P(r,m) = P(r) P(m)$$).

The entire full joint probability distribution is thus

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | red | Blue | green | yellow |  |
| short | 0.12 | 0.12 | 0.1 | 0.06 | 0.40 |
| medium | 0.105 | 0.105 | 0.0875 | 0.0525 | 0.35 |
| long | 0.075 | 0.075 | 0.0625 | 0.0375 | 0.25 |
|  | 0.30 | 0.30 | 0.25 | 0.15 |  |

Now we are in a position to see Bayes theorem in action. Suppose that we pull out a green bulb from the box. The conditional probability that the lifetime is short $$P(s|g)$$ is the relative proportion that the green and short entry $$P(g,s)$$ has compared to the sum of the column green $$P(g)$$. Numerically,

Another way to write this is as

which better shows that the conditional probability is the relative proportion within the column headed by the label green.

Likewise the conditional probability that the bulb is green given that its lifetime is short is

Notice that this time the relative proportion is measured against joint probabilities across the colors (i.e. across the row labeled short). Numerically, $$P(g|s) = 0.1/0.4 = 0.25$$

Bayes theorem links these two probabilities through

which is happily the value we got from working directly with the joint probabilities.

The next day, we did some more cyber-digging and found that a group of graduate students at the same university extended the undergraduate results (were they perhaps the same people?) and reported the following joint probability distribution

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | red | blue | green | yellow |  |
| short | 0.15 | 0.10 | 0.05 | 0.10 | 0.40 |
| medium | 0.05 | 0.12 | 0.15 | 0.03 | 0.35 |
| long | 0.10 | 0.08 | 0.05 | 0.02 | 0.25 |
|  | 0.30 | 0.30 | 0.25 | 0.15 |  |

Sadly, we noticed that our assumption of independence between the lifetime and color was not born out by experiment since $$P(color,lifetime) \neq P(color) P(lifetime)$$. However, we were not completely disheartened since Bayes theorem relates relative proportions and, therefore, it might still work.

Trying it out we computed

And

Checking Bayes theorem, we found

guaranteeing a happy and merry Christmas for all.

Next time, I’ll show how this innocent looking computation can be put to subtle use in inferring cause and effect.

## Bayesian Inference – Cause and Effect

In the last column, the basic inner workings of Bayes theorem were demonstrated in the case where two different random variable realizations (the attributes of the Christmas tree bulbs) occurred together in a joint probability function. The theorem holds whether the probability functions for the two events are independent or are correlated. In addition, it can be generalized in an obvious way to cases where there are more than two variables and where one some or all of them are continuous rather than discrete random variables.

If that were all there was to it – a mechanical demonstration between conditional and joint probabilities – Bayes theorem would make a curious footnote in probability and statistics textbooks and would hold little practical interest and no controversy. However, the real power of Bayes theorem comes in ability to link one statistical event with another and to allow inferences to be made about cause and effect.

Before looking at how inferences (sometimes very subtle and non-intuitive) can be drawn, let’s take a moment to step back and consider why Bayes theorem works.

The key insight come from examining the meaning contained in the joint probability that two events, $$A$$ and $$B$$, will both occur. This probability is written as

where the operator $$\land$$ is the logical ‘and’ requiring both $$A$$ and $$B$$ to be true. It is at this point that the philosophically interesting implications can be made.

Suppose that we believe that $$A$$ is a cause of $$B$$. This causal link could take the form of something like: $$A$$ = ‘it was raining’ and $$B$$ = ‘the ground is wet’. Then is it obvious that the joint probability takes the form

which in words says that the probability that ‘it was raining and the ground is wet’ = the probability that ‘the ground is wet given that it was raining’ times the probability that ‘it was raining’.

Sometimes, the link between cause and effect is obvious and no probabilistic reasoning is required. For example, if the event is changed from ‘it was raining’ to ‘it is raining’ it becomes clear that ‘the ground is wet’ due to the rain. (Of course even in this case, another factor may also be contributing to how wet the ground is but that complication is naturally handled with the conditional probability).

Often, however, we don’t observe the direct connection between the cause and the effect. Maybe we woke up after the rain had stopped and the clouds had moved on and all we observe is that the ground is wet. What can we then infer? If we lived somewhere without running water (natural or man-made), then the conditional probability ‘that the ground is wet given that is was raining’ would be 1 and we would infer that ‘it was raining’. There would be no way for the ground to be wet other than to have had rain fall from the sky. In general, such a clear indication between cause and effect doesn’t happen and the conditional probability describes the likelihood that some other cause has led to the same event. In the case of the ‘ground is wet’ event perhaps a water main had burst or a neighbor had watered their lawn.

In order to infer anything about the cause from the observed effect, we want to reverse the roles of $$A$$ and $$B$$ and argue backwards, as it were. The joint probability can be written with the mathematical roles of $$A$$ and $$B$$ reversed to yield

Equating the two expressions for the joint probability gives Bayes theorem and also a way of statistically inferring the likelihood that a particular cause $$A$$ gave the observed effect $$B$$.

Of course any inference obtained in this fashion is open to a great deal of doubt and scrutiny due to the fact that the link backwards from observation to proposed or inferred origin is one built on probabilities. Without some overriding philosophical principle (e.g. a conservation law) it is easy to confuse coincidence or correlation with causation. Inductive reasoning can then lead to probabilistically support but untrue conclusions like all swans are white – so we have to be on our guard.

Next week’s column will showcase one such trap within the context of mandatory drug testing.

Bayes and Drugs

One of the most curious features of Bayesian inference is the non-intuitive conclusions that can result from innocent looking observations. A case in point is the well-known issue with mandatory drug tests being administered in a population that is mostly clean.

For the sake of this post, let’s assume that there is a drug, called Drugg, that is the new addiction on the block and that we know from arrest records and related materials that about 7 percent of the population uses it. We want to develop a test that will detect the residuals in a person’s bloodstream thus indicating that the subject has used Drugg within some period time (e.g. two weeks) prior to the administration of the test. The test will return a binary result with a ‘+’ indicating that the subject has used Drugg and a ‘-‘ indicating that the subject is clean.

Of course, since no test will be infallible, one of our requirements is that the test will provide an acceptably low percentage of cases that are either miss detections or false alarms. A missed detection occurs when the subject uses Drugg but the test fails to return a ‘+’. Likewise, a false alarm occurs when the test returns a ‘-‘ but the subject is clean. Both situations present substantial risk and potentially high costs, so the lower both percentages can be made the better.

In order to develop the test, we gather 200 subjects for clinical trials; 100 of them are known Drugg users (e.g. they were caught in the act or are seeking help with their addiction) and the remaining 100 of them are known to be clean. After some experimentation, we have reached the stage where the 99 percent of the time, the test correctly returns a ‘+’ when administered to a Drugg user and 95 percent of the time, it correctly returns a ‘-‘ when administered to someone who is clean. What are the false alarm and missed detection rates?

This is where Bayes theorem allows us to make a statistically based inference and one that is usually surprising. To apply the theorem, we need to be a bit careful with notation so let’s first define some additional notation. A person who belongs to the population that uses Drugg will be denoted by ‘D’. A person who belongs to the population that is clean will be denoted by ‘C’. Let’s summarize what we know in the following table.

|  |  |  |
| --- | --- | --- |
| Description | Symbol | Value |
| Probability of a ‘+’ given that the person is C | P(+|C) | 0.05 |
| Probability of a ‘-’ given that the person is C | P(-|C) | 0.95 |
| Probability of a ‘+’ given that the person is D | P(+|D) | 0.99 |
| Probability of a ‘-’ given that the person is D | P(-|D) | 0.01 |
| Probability that a person is C | P(C) | 0.93 |
| Probability that a person is D | P(D) | 0.07 |

There are two things to note. First the results of our clinical trials are all expressed as conditional probabilities. Second, the conditional probabilities for disjoint events sum to 1 (e.g. P(+|D) + P(-|D) = 1 since a member of D, when tested, must result in either a ‘+’ or a ‘-‘).

In the population as a whole, we won’t know to which group the subject belongs. Instead, we will administer the test and get back either a ‘+’ or a ‘-‘ and from that observation we need to infer to what group the subject is most likely to belong.

For example, let’s use Bayes theorem to infer what the missed detection probability, P(D|-) (note the role-reversal between ‘D’ and ‘-‘). Applying the theorem we get

Values for P(-|D) and P(D) are already listed above, so all we need is to get P(-) and we are in business. This probability comes is obtained from the formula

Note that this relationship can be derived from , and . The first formula says, in words, that the probability of getting a negative from the test is the probability of either getting a negative and the subject is clean or getting a negative and the subject uses Drugg. The second formula is essentially the [definition of conditional probability](http://aristotle2digital.blogwyrm.com/?p=250).

Since we’ll be needing them P(+) as well, let’s compute them both now and note their values.

|  |  |  |  |
| --- | --- | --- | --- |
| Description | Formula | Symbol | Value |
| Probability of a ‘+’ given that the person either is in C or D |  | P(+) | 0.1158 |
| Probability of a ‘-’ given that the person either is in C or D |  | P(-) | 0.8842 |

The missed detection probability is

So things are looking good and we are happy. But our joy soon turns to perplexity when we compute the false alarm probability

This result says that around 40 percent of the time, our test is going to incorrectly point a finger at a clean person.

Suppose we went back to our clinical trials and came out with the second version of the test where nothing had changed except P(-|C) had now risen from 0.95 to 0.99. As the figure below shows, the false alarm rate does decrease but still remains very high (surprisingly high) when the percentage of the population using Drugg is low.

The reason for this is that when the percentage of users in the population is small in order to get the missed detection rate low we have to do it at the expense of a greater percentage of false alarms. In other words, our diligence in finding Drugg users has made us overly suspicious.

## Images and Representations

It’s an old idea. Someone you know holds up a photograph depicting a beautiful car, say a Corvette, and asks you “what this?”. You answer, “it’s a Corvette,” and are greeted with the cheeky response, “no! silly, it’s a picture.”

This

# Games and Double Effect

This week I decided to take a small break from the heady and deep logical and philosophical issues that the previous several columns have been exploring to talk about a strange and curious double standard in modern discourse. I’ll try to aim the discussion to be a fun and a bit whimsical.

I think that there aren’t many amongst us who don’t enjoy playing games of one sort or another. The

## Bad Causality

Increase in property damage due to weather

Increase skin cancer due to the ozone hole

Increase in the polio once the polio vaccine was discovered.

## Kill Everything

Respect for all life? How? Should we kill all the lions because they eat lambs. Should we let some people die of small pox? How about meningitis?

## Turnabout is Fair Play

Logical argument for turnabout

CS188.1x Artificial Intelligence

* AI = machines act (not think) rationally
* Rational thought
  + Our ability to write down how to do logical deduction is relatively fragile
    - Joe the bear story – one chain in the deduction is missing then the whole thing falls apart
  + Doesn’t help us to deal with uncertainty
* Rational Action
  + It’s not how you think it’s the actions you take that matter
  + Maximally achieve pre-defined goals
    - Goal is the input to the AI
    - Goals are expressed in terms of utilities
  + Maximize your expected utility
    - First third deals with
      * Maximize
      * Utility
    - Middle third with
      * Expectation
      * Probabilistic Inference
  + Intelligence is to thought as wings are to flight
    - Artificial flight is different from biological flight
  + Good decisions rest on good predictions
    - Memory – you did a bad thing in the past don’t do it again
      * Machine learning
    - Thinking through a chain of consequences in the context of a model of the world and a tree of outcomes
  + Focus on uncertainty to emerge from the AI winter
  + Agent – an entity that perceives and acts
    - Sensor with percepts
    - Actuators with actions
    - Agent function maps percepts to actions