;login:

THE MAGAZINE OF USENIX & SAGE

July 2001 • Volume 26 • Number 4

inside:

NEEDLES IN THE CRAYSTACK: WHEN MACHINES GET SICK: PART 5: IN SEARCH OF CLEOPATRAS'S NEEDLES

by Mark Burgess

USENIX & SAGE

The Advanced Computing Systems Association &
The System Administrators Guild

needles in the craystack: when machines get sick

by Mark Burgess

Mark is an associate professor at Oslo College and is the program chair for LISA 2001.

<Mark.Burgess@iu.hio.no>

Part 5: In Search of Cleopatra's Needles

During his fifty-four year reign around 1500 B.C., the Pharaoh Thothmes III saw erected two 70-foot columns before the great temple of Heliopolis, near what is today Cairo. These two exclamatory masts were powerful symbols of Egyptian spiritual and technological supremacy, singular and immutable signals to contrast with North Africa's seething, inconstant desert sands. For almost 3,500 years, they adorned the entrance to the impressive temple as a symbolic gateway, with little competition from their surroundings.

The first of the needles changed hands as a gift to the British people, in 1819 in recognition of Nelson's victory over the French fleet at the Battle of the Nile in 1798. An odd gift perhaps, a lump of stone so heavy that it took years even to muster the effort to move it – but the significance lay more in its shape than its composition. The symbolic gesture was repeated for the United States in 1881, when the Khedive of Egypt, hoping to stimulate economic investment in his country, gave the twin monument to the city of New York. Since then, the obelisks have been emulated in several large cities, including Paris and Washington. They have nothing at all to do with Cleopatra (preceding her by some 1,500 years), nor are they made of any valuable substance, but their singularity of form conveys a powerful symbolism which is recognized by all cultures.

Symbolism is at the very root of our being. It is about the attachment of meaning to patterns (patterns of material or of behavior). It is a seemingly irrational phenomenon, not unique to humans, but quite possibly the very essence of our intelligence. Symbolism is interesting because it underlines an important dichotomy: the difference between information and *meaning*. In Part 4 of this series, we met the three horseman of entropy: the good, the bad, and the ugly. Information (entropy) has three interpretations: the good (that variation is information), the bad (that variation, hence information, brings uncertainty), and the ugly (that information tends to increase with and result in degenerative aging, or disorder).

The essence of these interpretations is that information and meaning are two very different things. Indeed, too much information is simply noise. Imagine a television from which the antenna has been removed. The screen is a speckled swarm of fuzzy dots, devoid of any obvious meaning. However, it is not devoid of information. On the contrary, the length of a message you would have to send to a friend, so that he or she could reproduce the precise picture, would have to be very great indeed. The point is not that there is little information, but that there is so much information that it is impossible to find any meaning in it. If one could trace the effect back to its many causes, one would find the history of colliding particles in the early universe which led to today's cosmic background radiation. All that information is represented on our television screens, but the essential connection from cause to effect is a little more than we are willing to grasp, or care about.

Failing the Half-Closed-Eye Test

It is necessary to limit information in order to attach meaning. Symbolism is about attaching meaning to limited lumps of information. Lighthouses, totem poles, monuments, company logos, and even icons of the natural world, such as mountains (e.g., Mount Fuji), are all simple statements with very low entropy: concentrated powerful signals which are easily seen against their surrounding environments. The more complex a message is, the more analysis it requires, and the harder it is to discern its meaning.

The importance of such strong signals is not just symbolic, rather the opposite: the reason they have symbolic significance is because they are *effective competition* with their surroundings. Competition is not a word normally used in computer science, certainly not in information theory: we are used to the relative certainty of propositions and logic, not to the bullying and submission of the boxing ring; but competition is a crucial concept whenever systems interact.

This has broad ramifications for information. Strong signals have a powerful effect on the environment. Conversely, too much information or meaning becomes garbage. It is indistinguishable from no information. Indeed, the whole concept of "renormalization" in science is about trimming away the noise to see what is left. If every signal is equally loud, the result is just loud noise.

Look at the Web pages of many ISPs, for instance, and one finds a seething jumble of signals, from advertisements to links, to imagery, to text and color – not unlike the pornographic neon jumble of urban market districts. If one half closes one's eyes and looks at a page of information, the immediate impact of its layout and any strong symbols is all that can be seen. Today, multimedia messages assail us from every angle, detracting from overall impact of any one part. Commercial television is perhaps the worst example in some countries, where invasion by commercials, rapid cuts, overlaid text information, unnatural emphasis of voice, inappropriate music and never a second of silence, pollute any message being conveyed with toxic irrelevancy.

To make a garbage heap, one only needs to collect a sufficient number of individually meaningful things and put them all together. There is a reason why the Washington Monument (another obelisk) is used as a symbol of strength and unity, rather than a haphazard slum or shanty town, where there is much more information. There is a reason why Big Ben in London is well known; the Vigeland Monument in Oslo; the Eiffel Tower in Paris; the Statue of Liberty in New York. These are strong signals, dominating their environments.

To make a garbage heap, one only needs to collect a sufficient number of individually meaningful things and put them all together.

Today, we constantly erode the meaning of symbols by abusing them out of context. Think of mobile phones which now spam us with well-known music, rather than the low-info bell, or imagine the Eiffel Tower in the Amazon rain forest. Such loss of context demeans the significance of symbols, creating garbage out of art. Understanding the significance of mixed signals is a subtle business; in fact, it is one of the perpetually unsolved problems which demands our attention. Part 4 of this series was about how to use information maximally, by limiting and structuring it; it was also about the fundamental limitations incurred when information is limited for the purpose of ordering. That discussion provided a mapping from cause to effect and showed the limits incurred on going back the other way.

But what about this opposite direction? Separating signal from noise is much harder. Going from observed effect to the possibly many causes is a much harder problem, because summarial effect is often a mixture of many causes (a many-to-one mapping), and the combination is not necessarily a linear superposition. Yet this problem is clearly at the heart of all diagnostics, fault analysis, intrusion detection, and performance tuning. What can we hope to find out from observations?

For most things, random means "too complicated to really analyze."

"Imagination is more important than knowledge"

We take many complex things around us for granted. For instance, according to statistics textbooks, coin tosses and rolled dice are considered random events. Is this true? A coin has only two sides, it is circular: the simplest shape possible in two dimensions. A die is scarcely more complicated. Isn't it possible to predict the outcome? How hard could it be?

What we know is that, if we do toss coins or dice, the distribution of results, over many throws, seems pretty random. The reason is, of course, only that we can't be bothered to figure it out. It is not all that difficult, *in principle*, to predict the outcome. We know Newton's laws, we know about gravity, and we know about geometry. So what's so hard? The answer is: everything else – the environment. The environment (fingers, the table, floor, air, height, etc.) comprises a bunch of variables which are quite complicated. Since we are lazy, and the problem of coin tossing is not the least bit interesting, we pretend we don't know how to do it and call the result "random." So random means "other variables" which we don't account for. It turns out that, when we get down to quantum mechanics, and the subatomic, the world turns out to be unpredictable for completely unknown reasons, but for most things, random means "too complicated to really analyze." In a coin toss, we classify all of the complex variables into just two outcomes – a many-to-two mapping.

If we want to know what causes the result, it is not possible to extract information about all of those causal variables just from the two outcomes. One cannot reverse a many-to-few mapping. Lost detail is gone forever.

Statistics was invented in order to get something out of situations like this. Statistics says: "Even though we can't reasonably analyze some problem in complete detail, we can still try to say something about them." The usual way that statistics is presented is rather obtuse, and even a little dishonest. The idea is that we go out and measure a bunch of stuff in some problem and plot the outcomes in some way. Then we try to fit some off-the-shelf probability model (Gaussian, Poisson, Parot, etc.) and work out a bunch of standard things. That approach is a little bit like trying to say something about human behavior by matching hairstyles.

In order to use statistics meaningfully, we need a model of what causes the results: a hypothesis which can be tested. That means we essentially have to guess how the mapping from many causes to few results works. From such a model, one can then work out the consequences and see whether they match observation. This is how science is done. Einstein, always good for a quote, said that "imagination is more important than knowledge." What he meant was that, to understand nature, we need to dream up models by imaginative means. They can then be confirmed or denied. He did not mean, as some have suggested, that imagination overrides knowledge, only that knowledge itself is not enough for finding answers: creativity is needed.

Once information has been discarded or lost, it is not possible to go back and recreate it, unless it is somehow coded into a local potential, as was discussed in Part 4, or journaled in an ever growing archive. The only way to refine one's understanding of events, to amplify on "it was just random," is to postulate a reason and then collect evidence which supports or denies it. Imagination is itself just a random entropic walk of possibilities, loosely constrained by a hypothesis.

A Comedy of Errors

Our brains are extremely good at fitting models to data – almost too good, in fact. It is as inconceivable as it is inevitable. When I was younger, I would sit and watch the television noise, after everyone had gone to bed, with a friend. While listening to Led Zeppelin, we would wait for the rabbit on the bicycle to race across the bottom on the screen, as it often did, late at night. The rabbit had apparently been observed by many of our friends, while watching the fuzzy dots. Of course, some prefer to look for faces in clouds, or emotional expressions on car radiator grills. The essence is the same.

At the end of the 19th century, the Italian astronomer Sciaparelli was having much the same problem. He trained his telescope on Mars and was amazed to find a criss-cross pattern of lines which he called *canali*. Percival Lowell misinterpreted his notes and thus began the legend of the Canals of Mars. The canali were later discovered to be a trick of the low-resolution distortion, or a loss of information, about random dark spots on the surface. A century later, another blurred picture of Mars apparently revealed a gigantic human face carved into the surface. On closer inspection, it was another trick of the light, fueled by lack of resolution, much to the disappointment of UFO enthusiasts and indeed Hollywood, who went ahead and made the movie anyway.

Model fitting is intrinsic to the way our cognition works, and we use that to good effect. When we look for meaning, we do so by attaching interpretive models to data we perceive. The process of problem solving is the forward process of mapping cause to effect. The reverse process is that of diagnostics, mapping effect back to cause. Our capacity for reasoning might even have developed as a by-product of this ability for causal analysis.

Model fitting is intrinsic to the way our cognition works, and we use that to good effect.

Models are essential to the way we think, so to understand any data, to diagnose any situation, we need some kind of causal model. We build mental-model imagery by generating high-entropy associations, when we parse language; this gives us the ability to infer things from previous experience and leave things unsaid. When we say “cake,” we don’t just think of a description of physical attributes, we think of the cake Granny made on Saturdays when we were small. We think that it was like the one in the cafe the other day. We think, not of one cake but of all cakes, of different colors, shapes and sizes, and of all our experiences eating cakes, and of parties. Perhaps we think of a cookery program on the television or a famous chef. From one starting place, our thinking diffuses into every niche of possible meaning. Those who are good at diagnostics are often those who are good at this random walk of association.

In other words, concepts are not isolated things in our minds. The robustness of meaning has only to do with the competitive strengths of the different interpretations. Our memories are dense networks, tied into many models. We have no unique search-key: there is no unique way of attaching meaning to memory; rather, we tie memories into many different models of meaning, which compete for attention. This feature of human cognition is what we exploit in blending systematic, rational thought with the apparently random walk of imagination. It is how we solve problems and diagnose causal change. This is not at all like databases. Computers work more like warehouses. We find a piece of information and store it on a shelf at a particular place. None of the other information is aware of it. No free relationships are forged. We have learned to make simple associations with relational databases, but these are very primitive.

Time’s Causal Arrow

One of the apparent paradoxes of change, which was first pondered in the world of physics, is how the apparently reversible laws of physics can lead to irreversible conse-

Complex changes are not easily undone without a memory of exactly what transpired.

quences. Reversibility is a property of physical law which means that time has no arrow. The fundamental equations of physical systems work just as well forwards as backwards. There is no concept of past or future. This resembles the problem of correlation in statistical analysis. If two things are correlated, A being correlated with B means that B is correlated with A. There is no arrow which tells us whether A caused B, whether B caused A, or whether they are both by-products of some deeper cause. This is precisely the problem that models try to unravel.

The "paradox" of reversibility is not really such a mystery to the three horsemen of entropy, because it all has to do with how things spread out into an environment. Many apparent paradoxes of physics arise because physicists forget the environment surrounding their object of attention, as a matter of routine. Physics is about isolating the signals of nature into single threads which are independent of one another. The aim is to find the arrow from cause to effect as clearly as possible, by stripping away the jungle of irrelevancy. Amongst the approaches used is to try to physically isolate effects from other signals by putting them in a box, or by shielding systems from their environments, or by performing control experiments and subtracting one from the other to obtain the variance.

The mystery of reversibility is this: the laws of physics describe infinitesimal changes; they are formulated in terms of "differentials," or changes so small that the environment is irrelevant to them. However, the laws also include recipes for combining small changes into large ones. It is at this larger scale, where we take a step back from the small details and do the half-closed-eyes test, that the environment becomes important. How we combine the infinitesimal changes is important, and that combination puts out feelers into the environment, where other signals and effects are lurking. Every time we combine tiny changes into larger ones, the effects of the environment play an infinitesimal role. As we get farther and farther from the starting point, and after many such combinations, we begin to see a real change, reflected in the landscape of influences from the surroundings. Systems spread out into their environments, mixing with other signals as they go. Suddenly, the way the infinitesimal changes were combined (their history) becomes very important to the final result.

The following analogy might be helpful. If you drive your car one meter in any direction, the world around does not have any great effect on you, and the change is easily undone. However, if you drive half way across the continent, then the route you take begins to play an important role: the features of the landscape make one route unequivalent to another. That places an implicit arrow on the journey. Complex changes are not easily undone without a memory of exactly what transpired. The way that physics is formulated, the memory of how changes transpire is usually lost (dissipated) to the environment; one chooses to ignore the information and treat it as ambient noise, and thus it seems like a mystery how the rest of the changes happened.

When we look for changes in the machinery of computer communities, this principle is of central importance. It is not enough to collect data about changes, do little statistical analyses, plot graphs, etc. if one cannot separate the important signals from the environment. Physicists throw away parts of a signal which they are not interested in (and sometimes get confused, but not as often as one might expect); this is how they find order in chaos.

The meaning of signals is a mapping from cause to effect. Signals which change things are more than just correlations (which are bi-directional), they are directed arrows, like conditional probabilities. Auto-correlations have direction in a time-series only by virtue of time's arrow.

Pins and Needles

When machines get sick, they exhibit certain observable symptoms. These are clues as to the cause. Humans, for instance, get sore throats, perhaps skin coloration, headaches, pain, and so on. The same symptoms characterize most illnesses because there are only so many things which can hurt or change visibly. Computers run slowly, perhaps even stop working or jumble data. There is only a finite number of symptoms which we observe, but the number of possible causes is far greater. In a sense, finding out the cause of machine illness from a few symptoms is like the problem of determining why a flipped coin shows heads or tails. It is a many-to-one mapping, which one is trying to reverse.

Of course, we would like to correct an illness at its source, if possible. Merely addressing the symptoms ignores the causal chain of events which led to the problem, and the fact that the chain is often unidirectional. Turning over a tossed coin does not change the conditions of the environment which selected one of its faces in the first place. The level of detail in such a response would be incommensurate with the level of detail in the environment which caused the result. That cannot be a cure. Similarly, patching symptoms does not cure the illness which causes them.

In complex machines, the cause of sickness has to be a strong signal. There is a lot of complexity, or entropy in modern computers, in biological systems which compete for resources. A weak signal would just be a whisper in the wind. In order to be noticed amongst the other things going on, the problem has to be sufficiently strong. That often means that by the time the signal has grown to noticeable levels, it is already well established, and hard to counteract.

The significance of such a signal would be its strength. Humans get sick when bacteria or viruses replicate themselves to such a degree that they present a signal in our bodies which is so strong as to be toxic. They start drawing resources from other tasks which suffer as a result. The point is not whether they are foreign or not. Cancers are not foreign, but they are also strong signals which are pathological. The effect of a strong dose of almost any substance or signal (even water!) is toxic to a system, because it drives that system in a direction which is not its usual one.

The Search for Extra-Network InteLligencE?

In recent years, it has become popular to build anomaly and intrusion detection systems that listen for rabbits on bicycles amidst the storm of traffic on networks. What is the likelihood of finding anything interesting?

We would like to be on the lookout for signals which could be dangerous or helpful to us. There are many signals out there, waiting to be understood. Searching for messages in complex signals is like trying to find a needle in a haystack. There are many examples of this problem: all kinds of diagnostics, fault detection (including medicine) may be viewed as such a search. Looking for genes in DNA by examining the coding and looking for patterns is another example. Even more difficult is the problem of figuring out the causal relationships between gene action and manifestations of phenotype (species and characteristics) and cell function. In the Search For Extra-Terrestrial Intelligence (SETI), scientists look for what might be a meaningful signal in the bath of fuzzy dots on your TV screen. Cryptanalysis has many of the same difficulties: how to tell a bitstream encrypted message from noise?

Searching for messages in complex signals is like trying to find a needle in a haystack.

In system administration we are often trying to walk the fine line between necessary complexity and the ability to control.

In each case, the problem is to identify meaning in a mass of data by looking for causal threads. If we were lucky, every important signal would be a strong, low-entropy obelisk contrasting with a sandy desert. Unfortunately, nothing is so simple. Where you find one obelisk, others appear. Often, for the sake of efficiency, one compresses signals into as small a space as possible. This is certainly true of network traffic. In a desert full of obelisks, none of them seem to distinguish themselves anymore. This, of course, is one thorn in the side of intrusion detection systems. How do we distinguish a good signal from a bad signal? How do we attach meaning to patterns of traffic? Can we, for example, see the difference between a denial of service attack and a legitimate transfer of data? Stealth attacks are deliberately low-entropy signals, designed to be hidden.

In system administration we are often trying to walk the fine line between necessary complexity and the ability to control, but as we have seen in the previous issues, complexity cannot really be controlled, it can only be regulated by competitive means. Sometimes ecologists introduce a species of animal or plant into an ecology, perhaps a predator which will control a parasite or infestation (like the snails or hedgehogs of Hawaii). They introduce one species because that is a simple signal, something which seems to be controllable. Sometimes this happens by accident (as with the grey squirrel in England). In order to make a difference, such a species needs a selective advantage, which is a very low entropy signal. The problem with such a signal is that it will tend to dominate. Dominant signals are often toxic. On the other hand, if one were to just throw a bunch of random animals into a system, their effect would be unpredictable, but perhaps more balanced. This is the dilemma discussed in Part 3.

Another example is drugs. Drugs (medications) are low-entropy signals designed to target specific “problems,” where problem is defined as something possibly dangerous or undesirable to us. Drugs are failing today because they make themselves very obvious. If you keep hitting something in the same place, it will either wear out or move out of the way. A good analogy would be that, if the Germans had invaded Poland and kept on invading Poland during the Second World War, then everyone would simply have moved somewhere else and left them to it. That would not have had the desired effect. Had they attacked everyone at the same time, the result would probably have been a failure, because each signal would have been too weak to be noticed. (“You, soldier! Surround the enemy!”)

When the environment or an external influence acts with a strong signal, it leaves a marked effect and tends to order the system. This is the way that certain genes become important. This is the way that computer programs are able to learn by finding signals which can be digitized into a particular path through a data structure or a digital value. This is how the brain learns, we presume, although no one actually knows how the brain learns. Bacteria become resistant to drugs by genetic selection in the face of a low-entropy signal. A single threat is easy to resist. Too much of a good thing, and it loses its significance.

Rotate Shields; Long Live the Collective!

Machines are subject to a variety of influences. Every influence is a signal, a message to the system which changes it somehow. Some changes are good, some bad, and some neutral. This bath of information is called the environment. It acts like a potential, or fitness landscape in which a machine is meant to survive and fulfill its function. Those machines which have evolved in complex environments have evolved defenses to many of the signals which are contrary to their stability. Human-made machinery, however, is put together in the way that humans see the world: by the half-closed-eye method. We extract the simplest machine which solves the main part of a complex task. We build to solve the problem as we see it.

Simplification helps understanding, but to quote Einstein again: “Everything should be made as simple as possible but no simpler!” If animal evolution had simplified to the extent that human machine-building does, they would all be dead. Human machines have traditionally been built for protected environments: assuming that the user will do no harm. Today, the network has unleashed a force which has taken the propositional world of computing by surprise: *competition*, fuelled by its ally *diversity*. Entropy arrives center stage, and our machines are ill-equipped to deal with it. Our diagnostic abilities have not been built into our technology in the same way that evolution pits competition against strong signals.

“Everything should be made as simple as possible but no simpler!“

Once machines, driven by humans, were connected in a network, it was inevitable that there would be a variety of players with a variety of motivations, some good, some bad, and some ugly. The environment of users and other machines, which our computers bathe in, by console and by network, contains many signals which help and oppose the system. We are just learning how to shield ourselves from such signals, as DNA did with cell walls, as organisms did with skin: we can build firewalls and filters for the network and access controls for resources, so that unwanted signals are not passed.

These controls are not perfect, however. The environment is huge, the machine is small; the resources available to a complex environment, to prod and to poke the system for weakness, are huge. Unlike, genetically determined machines, computers are reliant on humans to make changes. Adaptive systems do not really exist today.

Strong signals are needles in the side of the machine. There are a few ways that machines can resist such toxic signals. The first is to have a defense of any kind. The next level is by redundancy: if one strategy for protection fails, another can take over. The next is the bane of the Borg: rotating shield modulations. We vary the defensive strategy constantly, to prevent an attacker from adapting to the defensive signal. This is the strategy used by screen-savers, to avoid permanent damage to monitors from persistent, strong signals. In a competitive environment, constancy is both your friend and your enemy. It is an arms race, i.e., a race to stand still. It is the principle of the Red Queen (a quote from *Through the Looking Glass*):

> *“Well in our country,” said Alice, still panting a little, “you’d generally get to somewhere else – if you ran very fast for a long time as we’ve been doing.”*
>
> *“A slow sort of country!” said the Queen. “Now here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run twice as fast as that!”*

In computer systems, one could rotate system policy at random (cfengine can do this, for instance) to prevent users from learning system policy and adapting to it. By making the signal less obvious, one can actually win stability. This is the lesson of non-cooperative Game Theory. In the future, we shall have to learn to adapt and embrace complexity, if we are to create machines which have the ability to communicate in a collective. If we haven’t learned this already from society (love thy neighbor and lock thy door), then we’ll have to learn it all over again, the hard way.

The Babel Acupuncture

Our world is teeming with information: its jungles, its species, its cities and the ever changing dynamics of it all are an information system of formidable complexity. From this complex world, humans have imagined meaning in the simplicity of low-entropy symbols. It is a Tower far greater than Babel. Then, having built the Tower, we push it down again, polluting precious meaning through repetitive strain, and the meaning becomes the garbage which we fail to tidy up after ourselves, like a multimedia virus which sweeps to every quiet corner of the world. The result might be something indistinguishable from the fuzzy snow on a TV screen. The “Snow Crash” is no myth.

All the needles we have made, we turn on ourselves, in our community bartering grounds, as we compete for supremacy in whatever value system we hold dear. Whether economist, hacker, warmonger, or merely egoist, we slowly shoot ourselves in the foot with tiny needles.