LECTURE 19

SIMULATION -II

We now take up the question of the reliability of a simulation. I can do no better than quote from the Summer Computer Simulation Conference of 1975,

"Computer based simulation is now in wide spread use to analyse system models and evaluate theoretical solutions to observed problems. Since important decisions must rely on simulation, it is essential that its validity be tested, and that its advocates be able to describe the level of authentic representation which they achieved."

It is an unfortunate fact that when you raise the question of the reliability of many simulations you are often told about how much man power went into it, how large and fast the computer is, how important the problem is, and such things, which are completely irrelevant to the question that was asked.

I would put the problem slightly differently:

Why should anyone believe the simulation is relevant?

Do not begin any simulation until you have given this question a great deal of thought and found appropriate answers. Often there are all kinds of reasons given as to why you should postpone trying to answer the question, but unless it is answered satisfactorily then all that you do will be a waste of effort, or even worse, either misleading, or even plain erroneous. <u>The question covers both the accuracy of the modeling and the accuracy of the computations.</u>

Let me inject another true story. It happened that one evening after a technical meeting in Pasadena, California we all went to dinner together and I happened to sit next to a man who had talked about, and was responsible for, the early space flight simulation reliability. This was at the time when there had been about eight space shots. He said that they never launched a flight until they had a more than 99 point something percent reliability, say 99.44% reliability. Being me I observed that there had been something like eight space shots; one simulation had killed the astronauts on the ground, and we had had one clear failure, so how could the reliability be that high? He claimed all sort of things, but fortunately for me the man on his other side joined in the chase and we forced a reluctant admission from him that what he calculated was not the reliability of the flight, but only the reliability of the simulation. He further claimed that everyone understood that. Me, "Including the Director who finally approves of the flight?" His refusal to reply, under repeated requests, was a clear admission that my point went home, that he himself knew that the Director did not understand this difference but thought the report was the reliability of the actual shot.

He later tried to excuse what he had done with things like, what else could he do, but I promptly pointed out a lot of things that he could do to connect his simulation with reality much closer than he had. That was a Saturday night, and I am sure that by Monday morning he was back to his old habits of identifying the simulation with reality and making little or no independent checks that were well within his grasp. That is what you can expect from simulation experts - they are concerned with the simulation and have little or no regard for reality, or even "observed reality".

Consider the extensive business simulations and war gaming that goes on these days. Are all the essentials incorporated correctly into the model, or are we training the people to do the wrong things? How relevant to reality are these gaming models? And many other models?

We have long had airplane pilot trainers that in many senses give much more useful training than can be given in real life. In the trainer we can subject the pilot to emergency situations that we would not dare to do in reality, nor could we ever hope to produce the rich variety that the trainer can. Clearly these trainers are very valuable assets. They are comparatively cheap, efficient in the use of the pilot's time, and are very flexible. In the current jargon, they are examples of "virtual reality".

But as time goes on, and planes of other types are developed, will the people then be as careful as they should be to get <u>all</u> the new interactions into the model, or will some small, but vital, interactions of the new plane be omitted by oversight, thus preparing the pilot to fail in these situations?

Here you can see the problem clearly. It is not that simulations are not essential these days, and in the near future, rather that it is necessary for the current crop of people, who have had very little experience with reality to realize that they need to know enough so that the simulations include all essential details. How will you convince yourself that you have not made a mistake somewhere in the vast amount of detail? Remember how many computer programs, even after some years of field use, still have serious errors in them! In many situations such errors can mean the difference between life or death for one or more people, let alone the loss of valuable equipment, money, and time.

The relevant accuracy and reliability of simulations are a serious problem. There is, unfortunately, no silver bullet, no magic incantation you can perform, no panacea for this problem. All you have is yourself.

Let me now describe my sloppiest simulation. In the summer of 1955 Bell Telephone Laboratories decided to hold an open house so that the people living nearby, as well as relatives and friends of employees, could learn a little about what the people who worked there did. I was then in charge of, for that time, a large analog differential analyzer, and I was expected to give demonstrations all day Saturday. Much of what we were doing at that time was trajectories of guided missiles, and I was not about to get into security trouble showing some sanitized versions. So I decided that a tennis game, which clearly involves aerodynamics, trajectories, etc. would be an honest demonstration of what we did, and anyway I thought that it would be a lot more appealing and interesting to the visiting people.

Using classical mechanics I set up the equations, incorporated the elastic bounce, set up the machine to play one base line with the human player on the other, along with both the angle of the racket and the hardness with which you hit the ball which were set by two dials conveniently placed. Remember, in those days (1955) there were not the game playing machines in many public places, hence the exhibit was a bit novel to the visitors. I then invited a smart physicist friend, who was also an avid tennis player, to inspect and tune up the constants for bounce (asphalt court) and air drag. When he was satisfied, then behind his back I asked another physicist to give me a similar opinion without letting him alter the constants. Thus I got a reasonable simulation of tennis without "spin" on the ball.

Had it been other than a public amusement I would have done a lot more. I could have hung a tennis ball on a string in front of variable strength fan and noted carefully the angle at which it hung for different wind velocities, thus getting at the drag, and included those for variously worn tennis balls. I could have dropped the balls and noted the rebound for different heights to test the linearity of the elastic constants. If it had been an important problem I could have filmed some games and tested that I could reproduce the shots that had no spin on them. I did not do any of these things! It was not worth the cost. Hence it was my sloppiest simulation.

The major part of the story, however, is what happened! As the groups came by they were told what was going on by some assistants, and shown the display of the game as it developed on the plotting board outputs. Then we let them play the game against the machine, and I had programmed the simulation so that the machine could lose. Watching the entire process from the background, human and machine, I noticed, after a while, that not one adult ever got the idea of what was going on enough to play successfully, and almost every child did! Think that over! It speaks volumes about the elasticity of young minds and the rigidity of older minds! It is currently believed that most old people cannot run VCRs but that children can!

Remember this fact that older minds have more trouble adjusting to new ideas than do younger minds since you will be showing new ideas, and even making formal presentations to, older people throughout much of your career. That your children could understand what you are showing is of little relevance to whether or not the audience to whom you are running the exhibition can. It was a terrible lesson I had to learn, and I have tried not to make that mistake again. Old people are not very quick to grasp new ideas - it is not that they are dumb, stupid, or anything else like that, it is simply that older minds are usually slow to adjust to <u>radically</u> new ideas.

I have emphasized the necessity of having the underlying laws of what ever field you are simulating well under control. But there are no such laws of economics! The only law of economics that I believe in is Hamming's law, "You cannot consume what is not produced." There is not another single, reliable law in all of economics that I know of that is not either a tautology in mathematics, or else it is sometimes false. Hence when you do simulations in economics you have not the reliability you have in the hard sciences.

Let me inject another story. Some years ago the following happened at U.C. Berkeley. About equal numbers of males and females applied to graduate school, but many more men were accepted than women. There was no reason to assume that the men were better prepared on the average than were the women. Hence there was obvious discrimination in terms of the ideal model of fairness. The President of the University demanded to know which departments were guilty. A close examination showed that no department was guilty! How could that be? Easy! Various departments have varying numbers of openings for the entering graduate school, and various ratios of men to women applying for them. Those with both many openings and many men applying are the hard sciences, including mathematics, and those with the low ratios of acceptance and many women applying, are the soft ones like literature, history, drama, social sciences, etc. Thus the discrimination, if you can say it occurred, came from the fact that the men, at a younger age, were made to take mathematics which is the preparation for the hard sciences, and the women could or could not take mathematics as they chose. Those who avoided mathematics, physics, chemistry, engineering, and such, were simply not eligible to apply where the openings were readily available, but had to apply where there was a high probability of rejection. People have trouble adapting to such situations these days!

Here you see a not widely recognized phenomena, but one that has been extensively examined in many of its appearances by statisticians; the combining of data can create effects that are not there in detail. You are used to the idea that combining data can obscure things, but that it can also create effects is less well known. You need to be careful in your future that this does not happen to you - that you are accused, from amalgamated data, of what you are not guilty. Simpson's paradox is a famous example where both sub-samples can favor A over B and C, but the combined data favors B over A.

Now you may say that in the space flight simulations we combined data and at times made the whole vehicle into a point. Yes, we did, but we knew the laws of mechanics and knew when we could and could not do it. Thus, in midcourse corrections you get the vehicle pointed in exactly the right direction and then fire the retro or other rockets to get the corrections, and during such times you do not allow the people to move around in the vehicle as that can produce rotations and hence spoil the careful directing of the rockets. We thought we knew enough of the background theory, and we had had years of experience in the matter, so that the combining of all the details into one point mass still gave reliable simulation results.

In many proposed areas of simulation there are neither such known experiences nor theory. Thus when I was occasionally asked to do some ecological simulation I quietly asked for the mathematically expressed rules for every possible interaction, for example given the amount of rain what growth of the trees would occur, what exactly were the constants, and also where I could get some real live data to compare some test runs. They soon got the idea and went elsewhere to get someone more willing to run very questionable simulations which would give the results that they wanted and could use for their propaganda. I suggest that you keep your integrity and do not allow yourself to be used for other people's propaganda; you need to be wary when agreeing to do a simulation!

If these soft science situations are hard to simulate with much reliability, think of those in which humans by their knowledge of the simulation can alter their behavior and thus vitiate the simulation. In the insurance business the company is betting that you will live a long time and you are betting that you will die young. For an annuity the sides are reversed, in case you had not thought about that point. While, in principle, you can fool the insurance companies and commit suicide, it is not common, and the insurance companies are indeed careful about this point.

In the stock market, if there were any widely known strategy for making lots of money, the very knowledge of it would ruin the strategy! In this case people would alter their behavior to vitiate the predictions you made. Not that some legally permissible strategy could not exist, (though I am pretty sure that it would have to be a fairly nonlinear theory to do much good above the normal stock market rise), but it would have to be kept very private. The basic trouble is that the stock market is crooked. The insiders have knowledge which according to the explicitly stated laws they may not act on, but they do so all the time! If you do not use inside information then you have little chance against those who do, and if you do act on inside information you are acting illegally! It is a bad business either way, and the insiders are resisting all attempts to automate the trading by machine that would eliminate some of the inside deals that they now profit on. It is known that they do but it is apparently not provable in court! Furthermore, false "inside information" is constantly circulated in the hopes that the outsiders will think they are inside and act on it to the profit of the originators of the rumors.

Thus beware of any simulation of a situation that allows the human to use the output to alter their behavior patterns for their own benefit, since they will do so whenever they can.

But all is not lost. We have devised <u>the method of scenarios</u> to cope with many difficult situations. In this method we do not attempt to predict what will actually happen, we merely give a number of possible projections. This is exactly what Spock did in his baby raising book. From the observations of many children in the past he assumed that the future (early) behavior of children would not differ radically from these observations, and he predicted not what your specific child would do but only gave typical patterns with ranges of behavior, on such things as when babies begin to crawl, talk, say "no" to everything, etc. Spock predicted mainly the biological behavior and avoided as much as he could the cultural behavior of the child.

In some simulations the method of scenarios is the best that we can do. Indeed, that is what I am doing in this set of Lectures; the future I predict cannot be known in detail, but only in some kinds of scenarios of what is likely to happen, in my opinion. More on this topic in the next Lecture.

I want to return to the problem of deciding how you can make realistic estimates of the reliability of your simulations, or those that are presented to you in the future. First, does the background field support the assumed laws to a high degree? How sure are you that some small, but vital, effect is not missing? Is the input data reliable? Is the simulation stable or unstable? What cross checks against known past experience have you available for checking things? Can you produce any internal checks, such as a conservation of mass, or energy or angular momentum? Without redundancy, as you know from the talks on error correcting codes, there can be no check on the reliability.

I have not so far mentioned what at first will appear to be a trivial point; do the marks on the paper that describe the problem get into the machine accurately? Programming errors are known to be all too common.

Let me tell another story that illustrates this point that there are things one can do about this problem. One time the chemistry department was considering a contract to examine, for the Federal Government, the chemistry of the upper atmosphere immediately after an atomic bomb explosion. I was asked only to supply advice and guidance. Upon looking into the problem I found that there would be in each case that was to be computed somewhere around 100 ordinary differential equations to be solved, depending on the particular chemical reactions they expected.

I did not think that they could get the various sets of these equations into the machine correctly every time, so I said that we would first write a program that would go from the punched cards, one card describing each particular reaction with all its relevant constants of interactions, to the equations themselves, thus insuring that all the terms were there; no errors in the coefficients not being the same for the same reaction as it appears in different equations, etc. By hindsight it is an obvious thing to do; at the time it was a surprise to them, but it paid off in effort on their part. They had only to select those cards from the file that they wanted to include in the particular simulation they were going to run, and the machine automated all the rest, including the spacing of the steps in the integration. My main idea, besides the ease and accuracy, was to keep their minds focused on what they were best able to do - chemistry - and not have them fussing with the machine with which they were not experts. They were, moreover, in charge of the actual computing. I made it easy to do the bookkeeping and the mechanics of the computer, but I refused to relieve them of the thinking part.

In summary, the reliability of a simulation, of which you will see many in your career since it is becoming increasingly common, is of vital importance. It is not something you can take for granted just because a big machine gives out nicely printed sheets, or displays nice, colorful pictures. <u>You are responsible for your decisions, and cannot blame them on those who do the simulations, much as you wish you could. Reliability is a central question with no easy answers.</u>

Let us return to the relationship of analog to digital computers. The point sometimes arises in these of days of <u>neural nets</u>. The argument is made that the analog machines can compute things that the digital version cannot. We need to look at this point more closely - it is really the same as was made years ago when the analog computers were being displaced by digital computers. In these Lectures we now have the relevant knowledge to approach the topic carefully.

The basic fact is tht the Nyquist sampling theorem says that it takes two samples for the highest fequency present in the signal (for the equally spaced points on the entire real line) to reproduce (within roundoff) the original signal. In practice most signals have a fairly sharp cutoff in the frequency band; with no cutoff there would be infinite energy in the signal!

In practice we use only a comparatively few samples in the digital solution and hence something like twice the number that Nyquist required is needed. Furthermore, usually we have samples on only one side and that produces another factor of two. Hence, something from seven to ten samples for the highest frequency are needed. And there is still a little aliasing of the higher frequencies into the band that is being treated, (but this is seldom where the information in the signal lies). This can be checked both theoretically and experimentally.

Sometimes the mathematician can acccurately estimate the frequency content of the signal (possibly from the answer being computed), but usually you have to go to the designers and get their best estimates. Competent designer should be able to deliver such estimates, and if they cannot then you need to do a lot of exploring of the solutions to estimate this critical number, the sampling rate of the digital solution. The step by step solution of a problem is actually sampling the function, and you can use adaptive methods of step by step solution if you wish. You have much theory and some practice on you side.

For accuracy the digital machine can carry many digits, while analog machines are rarely better than one part in 10,000 per component, if that much. Thus analog machines cannot give very accurate answers, nor carry out "deep computations". But often the situation you are simulating has uncertainties of a similar size, and with care you can handle the accuracy problem.

With the passage of time we have developed wider band width analog computers, but we have used this to speed up the computations rather than use the implied band width of the circuits for accuracy. In any case, the fundamental accuracy of the analog parts limits what you can do with an analog machine. The old mechanical computers, like the RDA #2, took about half an hour per solution; the electrical computers derived from the gun directors, which still had some mechanical parts, took minutes; later an all electronic one took seconds, and now some of them can flash the solution on the screen as fast as you can supply input.

In spite of their relatively low accuracy analog computers are still valuable at times, especially when you can incorporate a part of the proposed device into the circuits so that you do not have to find the proper mathematical description of it. Some of the faster analog computers can react to the change of a parameter, either in the initial conditions or in the equations themselves, and you can see on the screen the effect immediately. Thus you can get a "feel" for the problem easier than for the digital machines which generally take more time per solution and must have a full mathematical description. Analog machines are generally ignored these days, so I feel that I need to remind you that they have a place in the arsenal of tools in the kit of the scientist and engineer.