Chapter 13

The First Experiment

THIS CHAPTER TELLS HOW TO RUN GREAT EXPERIMENTS, NOT IN THE DRAMATIC breakthrough sense, but in the methodical craftsman-who-knows-his-tools sense. Experiments are the analytical activist's central tool, the one used the most to penetrate to the very core of how to solve difficult social problems. Process can put you on the road to solution. Modeling can lead dazzling insights, key hypotheses and elegant solution strategies. But only experimentation can provide the actual proof that a road about to be taken will lead to success instead of disaster.

A great experiment is one that takes problem solvers a giant leap closer to their goal. The greatest leap of them all is your very first experiment, because once you or your organization takes this step, they have opened the door to the same hundredfold productivity increase that all of science experienced when it took the same step back in the 17^{th} century. To take this first step, all you have to do is run The First Experiment.

The purpose of The First Experiment is to help people make the mental leap to thinking in terms of process driven cycles of hypothesis generation, model growth, and experimentation. This is the very heart of the new paradigm. The rest of this book will assume you have bought into the new paradigm, and that your work is now driven by the Analytical Method and the System Improvement Process, or similar processes.

As many chapters in this book have demonstrated, structural thinking can be used to generate a comprehensive set of hypotheses about the causes of a social problem and the solution. This collection of hypotheses is best expressed as a formal model of behavior, using causal flow diagramming or simulation modeling. But how do we know the model is correct?

That's where experimentation comes in. Each node in a model is a hypothesis of how the inputs to that node cause the outputs from the node. In other words, each node is a cause and effect hypothesis. Science knows of only one way to determine whether a hypothesis is true: experimentation.

If you have worked on activist problems much at all, you have probably noticed a huge glaring problem: *most activists do not test their key assumptions*. Instead, they come up with one new scheme after another on how to solve a problem, and then optimistically rush out and try it. Hope springs eternal. Sometimes it works. Usually it does not. When it works they assume they were smart. When it fails they start over with another intuitive solution that is sure to work this time. Or they try the same solution again, but harder and somehow better.

It is time for classic activists to recognize the truth. When a solution works on difficult problems they were lucky, and when it fails there is a reason for that: the solution was based on one or more untested hypotheses.

There is only one way to correct this pattern of failure. It requires a whole new way of thinking, which begins by learning:

How to Run the First Experiment

This is about as simple as a social experiment can be and still collect meaningful results. Here is a one paragraph overview of how it works:

The experiment consists of a 5 page handout testing a single factor: whether or not a brief one page introduction to the Truth Test makes a difference. Half the handouts contain the Truth Test page and the others don't. All contain actual political statements. Each statement is followed by multiple choice questions about how the reader feels about the politicians who made the statements. When everyone is done the answer sheets are handed in. The results are then are calculated, announced, and discussed. All this takes about two hours.

The First Experiment is a controlled experiment. Let's explain what controlled experiments are and then how to run The First Experiment in more detail:

The most important thing about an experiment is the hypothesis it is designed to test. All experiments revolve around their hypothesis, just as the planets revolve around the sun. To understand or design an experiment you must start with its hypothesis, which for this one is:

> *Even a very brief exposure to the Truth Test can raise a person's ability to detect political deception.*

A hypothesis has two parts: the cause(s) and the effect(s). These become the independent and dependent variables in the experiment.

An **independent variable** is a factor that varies. Here it is whether or not a person is exposed to the Truth Test (presented on page 41). This is accomplished by using different second pages in the handouts.

Half are a brief introduction to the Truth Test and how to use it to spot fallacious arguments. The other half contain a "dummy page" which at a glance looks like the Truth Test, but is actually about a neutral subject related to political statements. The reason for the dummy page is so that both groups take about the same time to complete the questionnaire and both groups pause to read something new just before they are exposed to fallacious political statements.

A **dependent variable** is a measure of how subjects respond to the experimental conditions, which include the independent variable. The response is measured using the data in the answer sheets.

A **controlled experiment** divides the subjects into two groups: the treatment group and the control group. The **treatment group** receives the experimental treatment, which in this case is the handouts with the Truth Test. The **control group** receives no special treatment, which is the handouts with the dummy page.

If subjects are randomly assigned to the treatment and control groups, you have a **controlled experiment**. If the results of the two groups are significantly different, the difference must be due to the factor that varied between groups. This assumes there was no experimental error, such as experimenter bias, random assignment errors, or data collection errors.

Everything you need to run The First Experiment is at Thwink.org, including these 4 PDF files: [103]

1. **Handout with Truth Test -** Print 5 of these for every 10 people.
2. **Handout without Truth Test -** Print 5 of these for every 10 people.
3. **Additional Material -** Print 10 of these for every 10 people.
4. **Experiment Results -** Print 2 per experiment. Send a completed one to Thwink.org.

A Sample from the Handout

Politician B – "In the last few years government has gotten bigger and bigger; industry has gotten larger and larger; labor unions have gotten bigger and bigger; and our children have been the victims of mass education. We must make this next century the century of the individual. We should never forget that a government big enough to give us everything we want is a government big enough to take from us everything we have. The individual worker in the plants throughout the United States should not be a small cog in a big machine."

To what extent do you agree with this statement and its implications?

___ Strongly agree
___ Somewhat agree
___ Neither agree or disagree
___ Somewhat disagree
___ Strongly disagree

Why do you feel this way? ___________________________

A Sample from the Additional Materials

Statement by politician B – From President Ford's closing statement, in the first Carter-Ford presidential debate on September 23, 1976. Ford's assertion that "We should never forget that a government big enough to give us everything we want is a government big enough to take from us everything we have." implies that big government is bad, without saying how big "big" really is. Also, the conclusion that big government will "take from us everything we have" does not follow from the premise. The conclusion is also gross exaggeration, because there has never been a government that has taken everything from its citizens. Ford is creating a false enemy here by raising the false fear that government itself is bad.

By contrast, in his debate statements Jimmy Carter did not engage in nearly as many fallacious appeals. For example, in his closing statement he said "I don't claim to know all the answers. But I've got confidence in my country. Our economic strength is still there. Our system of government—in spite of Vietnam, Cambodia, CIA, and Watergate—is still the best system of government on earth." Instead of pushing the fear hot button, Jimmy Carter did just the opposite. He appealed to the people's intellect by saying that he didn't have all the answers, and implying that he could figure them out as he went along. Thus he avoided making the false promises so common to political campaign speeches, where numerous false promises are made to woo supporters.

* * *

The additional materials explain the fallacies, which can be quite involved. The explanations, the Truth Test, and the Dueling Loops provide the basis for a fascinating and potentially illuminating discussion after the experiment is completed. If you do a good job of preparing for the discussion and it goes well, the eyes of some of the activists in the group will begin shining with a certain spark that will soon ignite the flame of Analytical Activism.

Here are the steps to run the experiment:

Step 1 – Preparation

Print out half of the handouts with the Truth Test and half without it. Staple each set together, so that there are 5 pages in each set. The all important page 2 is not visible. Shuffle the sets of pages to be handed out ahead of time. Hand them out in any manner desired. If the experimenter is not the one who shuffled them, or shuffled them in a manner such that they lost track of which was which, then we have a double blind experiment and randomly selected treatment and control groups, and therefore a valid controlled experiment. This is important.

It is crucial that the test subjects be unaware of the two types of handouts and the existence of the Truth Test, so do not discuss this beforehand. Just tell everyone you will be running an interesting social experiment.

Step 2 – The Actual Experiment

The group silently reads the questionnaire and answers the questions. This takes about 30 minutes.

Step 3 – Determining the Results

When all are done, everyone hands in their answer page and keeps the other pages. The experimenter then hands out the additional material pages, which reveal who said what, where the fallacies are, and explain what the experiment is trying to do. While the group is reading this, the experimenter tallies up the answers, adds up the results, and prepares the summarized outcome of the experiment. This is very simple to do.

Step 4 – Discussion

After everyone has read the material, the experimenter leads a discussion. This can be very insightful and educational. Emphasis should be placed on how it was a double blind, randomly assigned groups, high quality controlled experiment. This is the gold standard of experimentation.

The results only hold for the total population if the full group was randomly selected, which is not the case if your group is friends or colleagues. But in spite of this the results will be useful because these groups are still somewhat representative.

After people understand the experimental design the results of the experiment should be presented.

The hypothesis is that those subjects with the handouts educating them about how to detect fallacious arguments will be less easily deceived. How the results support or do not support the hypothesis should be discussed. If the results support the hypothesis, then how this new scientific knowledge can be used to improve the way the human system is engineered is a lively topic, and a possible good note to end the formal discussion on. On the other hand, if the results do not support the hypothesis, that too is a fruitful topic. If the hypothesis is false, does that mean the Dueling Loops model is flawed? Or is it the experiment that has a problem? Or was the experiment okay but the group was so small that a statistical fluke occurred that is not representative of the population?

You should discuss the Dueling Loops and the high leverage point the experiment is testing. Please study The Dueling Loops of the Political Powerplace chapter so you can do this well. If people can understand how the Dueling Loops work then they will be able to see how their participation in this experiment can help greatly in solving the problem, and they will be more motivated to become experimenters themselves.

At some point you may want to discuss what logical fallacies are and which are the most common. Study up on this. A well run experiment should itself result in a small bit of training the population to be more immune to the immense power of political deception.

Step 5 – Report the Results

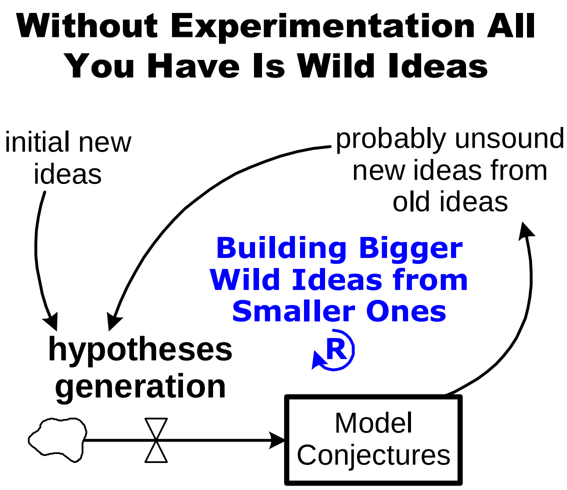
Report the results to Thwink.org by sending us a completed copy of the Experimental Results page, along with any comments and suggestions.

The Importance of Experimentation

That's all there is to running The First Experiment. But experiments alone will not solve difficult problems. They are but a single step in a greater process.

This process was modeled earlier on page 180 as The Memetic Evolution of Solutions to Difficult Problems. The key loop is ***Building the New from the Old***. It is only by building the new upon the old that knowledge advances.

But classic activists have not been using experimentation to build their mental models of how to solve the sustainability problem. Thus their process looks like the one shown below:



Classic Activism leads to ***Building Bigger Wild Ideas from Smaller Ones***, because the ideas are never tested. The loop starts with the initial new ideas. These are the first hypotheses, which causes hypotheses generation to begin. These are haphazardly assembled into the many mental models that classic activists are using to solve problems. Because they are untested they are Model Conjectures. As the models of how to solve the problem grow, they are used to create new ideas. But because they are based on conjecture and not proven fact, what is created is probably unsound new ideas from old ideas. These are used for more hypotheses generation and the loop starts over again.

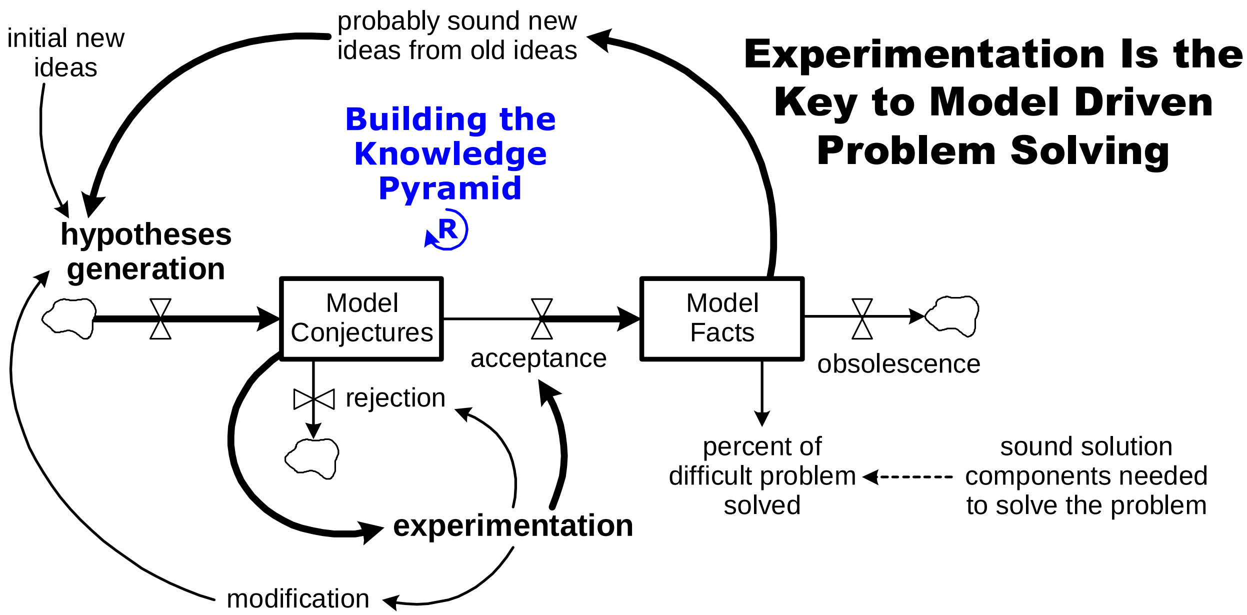
The stock of Model Conjectures can grow quite large, as indeed many surefire solutions to the sustainability problem have. But none of them have worked, because a bigger wild idea is no better than a smaller one.

There is, however, a better way. It begins by realizing that the limitation to the above process is the hypotheses are never formally tested.

In the 17th century society's problem solvers invented the Scientific Method. When combined with the Industrial Revolution in the late 18th and early 19th century, this caused scientific knowledge and technology to begin growing exponentially. It was as if a dam had split wide open and now nothing could stop problem solvers from solving one difficult problem after another. The dam that burst was the invention of what was holding back the above process, which is what all of science and industry had been using up to that point. The roadblock was there was no reliable way to test new ideas to see if they were true or false. What removed this barrier in one swift stroke was the invention of experimentation.

Today it is intuitively obvious that new ideas must be tested before they can be accepted and used as the basis for creating more scientific knowledge. But long ago it was not, because the definition of sound knowledge was what could *logically* be shown to be true. The problem with this is that while some logic is better than none, logic alone is unreliable when it comes to creating new cause and effect knowledge. It is too easy to make an error, particularly if you are biased and hope your hypothesis is true.

The historic result was the ***Building Bigger Wild Ideas from Smaller Ones*** process. But with the addition of experimentation, the process changed completely to the one shown on the next page.



This diagram shows how experimentation is the key to model driven problem solving. At last knowledge creators have a reliable way to move hypotheses from conjecture to fact: experimentation. The main loop consists of the thick arrows.

Recall that the steps of the Scientific Method are:

1. Observe a phenomenon that has no good explanation.
2. Formulate a hypothesis.
3. Design an experiment(s) to test the hypothesis.
4. Perform the experiment(s).
5. Accept, reject, or modify the hypothesis.

Experimentation is steps 3, 4, and 5. The diagram shows how once step 4 is complete, a hypothesis flows to one of three places: If it has been proven true, acceptance occurs and it flows on to Model Facts. If the hypothesis has been proven false and is not worth trying to modify, rejection occurs and it dies. But if false and it appears to still have some potential if it can be improved, modification occurs and it flows back to hypotheses generation.

Now as the stock of Model Facts grows, we no longer get probably unsound ideas from old ideas. Instead we get probably sound new ideas from old ideas. This increases the likelihood that the next round of hypotheses generation will contain hypotheses that are true or can be modified slightly to be true once experimentation occurs. In other words, the process is now producing *high quality hypotheses*.

As the stock of Model Facts accumulates, something that is not possible with Classic Activism occurs. The percent of difficult problem solved starts to rise substantially and rapidly. The percent is calculated by comparing the solution components on Model Facts to the sound solution components needed to solve the problem. In the real world these are initially unknown and can only be discovered by this process. The stock of Model Facts grows until the percent of difficult problem solved reaches 100%, at which point the problem is solved.

But there is more of the process to consider, because in the real world we must consider solution component obsolescence.

The only thing that will happen for certain is change. This is especially true for difficult complex social system problems. As they change certain Model Facts become out of date and die due to obsolescence. This requires problem solvers to continually apply the process to keep the Model Facts high enough to keep the problem solved. If they do not do this, then Model Drift will begin and soon there will be a Model Crisis. This explains why a good solution includes a self-managing component that insures that dominant social agents have strong incentives to keep the problem solved.

Building the Knowledge Pyramid

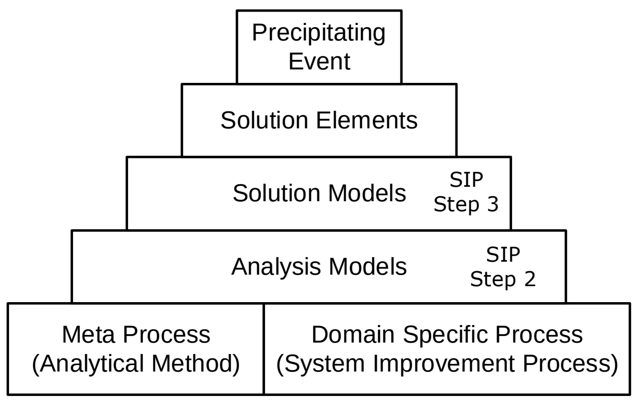
The main loop is called ***Building the Knowledge Pyramid*** because to solve difficult problems, a stack of layers of sound knowledge must be built. This stack is conceptually represented in the pyramid shape shown on the next page.

The top four layers are models. The top two layers are particular model components that actually solve the problem. The very top layer is what kicks off the chain of engineered events that leads to solution of the problem. In this book the top of the pyramid is the precipitating event. In other problems it might be the initial solution elements.

Each layer is heavily dependent on the layer beneath it. If a layer is weak then the layer above it will be even weaker, which will cause the pyramid to come crashing down long before the problem is solved, as it becomes painfully obvious that the solution being constructed is not going to work. For the environmental sustainability problem this has already happened thousands of times at the local level and dozens of times at the international level.

The <u>Model Facts</u> stock represents the top four layers of the pyramid. The ***Building the Knowledge Pyramid*** process as a whole is the bottom layer.

The Knowledge Pyramid



If you have read parts one and two of this book and have made the personal decision to switch to Analytical Activism, then you have essentially decided to begin building a Knowledge Pyramid for yourself and your organization. This book contains everything you need to get started on that Herculean task. We wish you the very best of luck. But then again, it is not luck that is going to make the difference. It is something else.

The big three of Analytical Activism are process, modeling, and experimentation. The pyramid shows how a process that fits the problem drives the entire work effort. The process drives your modeling, which does two main things: First, it represents the core of your understanding of how the system with the problem behaves, including how various solutions should work. As the ***Building the Knowledge Pyramid*** diagram shows, modeling also generates a steady stream of new high quality hypotheses as the model evolves to reflect your increased understanding. Experimentation tests the crucial hypotheses. If they are correct they stay in the model. If not the model is revised. *If experimentation keeps up with model growth then there is nothing to stop the model from eventually becoming complete and correct enough to solve the problem.*

The Limits to Solving a Difficult Problem

The three limits to solving a difficult complex social system problem are process maturity, model maturity, and experimentation capacity. *Whichever has the lowest limit determines how far and how fast your problem solving efforts will go.*

Once your organization decides to reengineer itself along the most efficient and effective lines possible, at first most of your reengineering effort will go (or at least should go) to process maturity, which is the first layer of the pyramid. This is because in the beginning process maturity will be your limiting constraint, since you have probably been using Classic Activism, which has a very low process maturity.

Then, since you are working on a difficult social problem, the process will guide you toward the second layer of analysis models. Once that is solid the third layer of solution models can be built. During this phase, as you and others are learning system dynamics, modeling is your limiting constraint.

As the second and third layers begin to grow and your modeling skills mature, a number of untested key assumptions will stand out like cardinals on a pine tree. It is at this point that you will find that experiment throughput capacity is now your limiting constraint, because modeling goes much faster than experimentation.

After this experiment capacity will remain the limiting constraint indefinitely. *Thus experimentation is ultimately the most important element in your total process.* This is not only because it is the long term bottleneck. It is also because as your experiments move up from logical (in the model) to artificial (on groups of people in an artificial setting) to real world experiments (on real subsystems like a city), they have the ability to scale up smoothly to the actual global solution. There is no definite line between when an experiment becomes part of the actual solution, just as there is no discernable line as one species evolves into another. This is because solutions to difficult problems are the result of memetic evolution.

To summarize, once an organization commits itself to Analytical Activism or a similar process, at the strategic level the key to success is process maturity. But at the tactical level it is modeling and experimentation.

Over time, generally within a few years, the limiting constraint to how long it takes to solve a problem is how well can experimentation keep up with the key hypotheses generated by modeling. Therefore:

When It Comes to Solving Difficult Social Problems, High Experimental Efficiency Is the Crux

Environmentalists are problem solvers. When Thomas Edison famously remarked that genius is 1% inspiration and 99% perspiration, what he was saying is the labor of a problem solver is 1% hypotheses generation and 99% experimentation, and that experimentation is the hard part. If you don't achieve high experimental efficiency, you are not going to achieve your goal.

Edison's goal was a continuous stream of breakthrough inventions. When he opened the world's first invention factory in Menlo Park, New Jersey in 1876, he promised he would "invent some minor thing every ten days and some big thing every six months." And he did it.

Environmentalists have a different goal: solving the global environmental sustainability problem. It is a much harder problem than any of the problems Edison solved. But if you look at the sum total of all of Edison's inventions, they were just as hard. What he did was create the Age of Electricity. He invented the world's first practical light bulb. He invented and implemented the world's second electric utility company, which by 1994 was serving 508 customers with 10,164 electric lights. All told, he patented 1,097 inventions, nearly all of them electrical based. It was this flood of new discoveries that precipitated the Age of Electricity.

It will be a similar flood of successful experiments that will precipitate the Age of Sustainability. Each experiment will prove or disprove a model conjecture. As the experimental output grows, so too will the stock of model facts, until the Knowledge Pyramid is tall enough to solve the sustainability problem.

Edison succeeded only because he was able to harness the power of high experimental efficiency. Environmentalists will succeed only if they can do the same. If they can't then the pyramid will grow too slowly to solve the problem it time.

The Goal of a Massive Experimental Effort

Let's be clear about the goal of the massive experimental effort that will be required to solve the sustainability problem. The goal is to first build an *analysis model* that is so sound that once it is built, how to solve the problem will be as obvious as how to eat an ice cream cone. The low leverage points that problem solvers have been mistakenly pushing on will pop into focus. So will the high leverage points they need to be pushing on instead. After that, further experimentation on solution element candidates and model evolution will lead to a *solution model* that, once implemented, will solve the problem. The solution model shows how pushing on the high leverage points will solve the problem.

The goal is these two models.

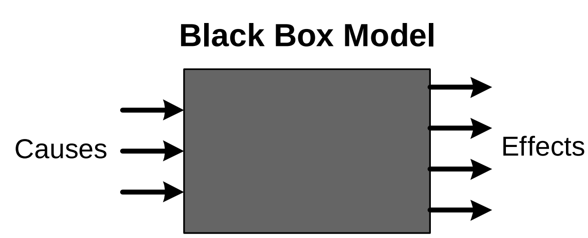
The General Model of System Behavior

Now then, where do the analysis and solution models fit into the greater scheme of things? *They allow us to create a glass box model of the human system process that we are improving.*

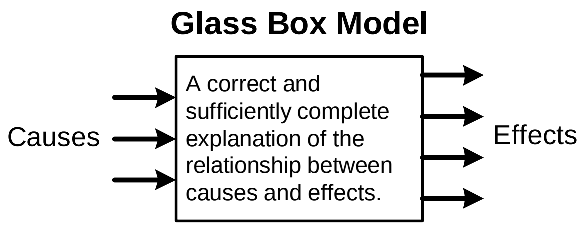
This perspective views the human system as a gigantic ongoing process. We will not be improving all of it—only the portion it takes to solve the problem. This perspective allows powerful insights into how to best approach experimentation, because to improve the process, all we have to do is find which inputs make the critical difference, and change them to the values that will solve the problem.

The goal of experimentation is to more deeply and correctly understand the cause and effect relationships between a system's inputs and outputs. The better this is understood, the better we can decide how to change the inputs so as to improve the outputs. We touched on this earlier when discussing black box and glass box models on page 91. Let's review those concepts and then take them a little further.

A black box model of a system has no idea why the causes lead to the effects. It only knows that they do. For example, few classic activists know why they sometimes make modest progress on solving the sustainability problem. They only know that they do, which causes them to keep trying the same problem solving process over and over.



The superior alternative is a **glass box model** explaining why cause and effect occurs. For example, once classic activists have a model that explains why they are failing to solve the sustainability problem, they will be able to change their approach to one that works. The change can be fast, in a few years, but it must be gradual and evolutionary, because no one knows exactly what new approach will work best. We must experiment to find the best new approach. Overall it may be something like Analytical Activism. But what should the details be like? Which key process elements are the most important?



To answer questions like these we need a general conceptual model of how systems work from the perspective of experimentation. A suitable one, adapted from the large amount of literature available, is shown below. [104] The causes and effects of the black box and glass box models have been replaced by the independent and dependent variables.

The four types of arrows represent what to consider when designing an experiment. **System inputs** flow into the system, such as the natural resources the human system acquires from the greater system it lives within, the environment. **System outputs** flow out of the system, such as the wastes emitted by the human system into the environment. They are the **dependent variables**.

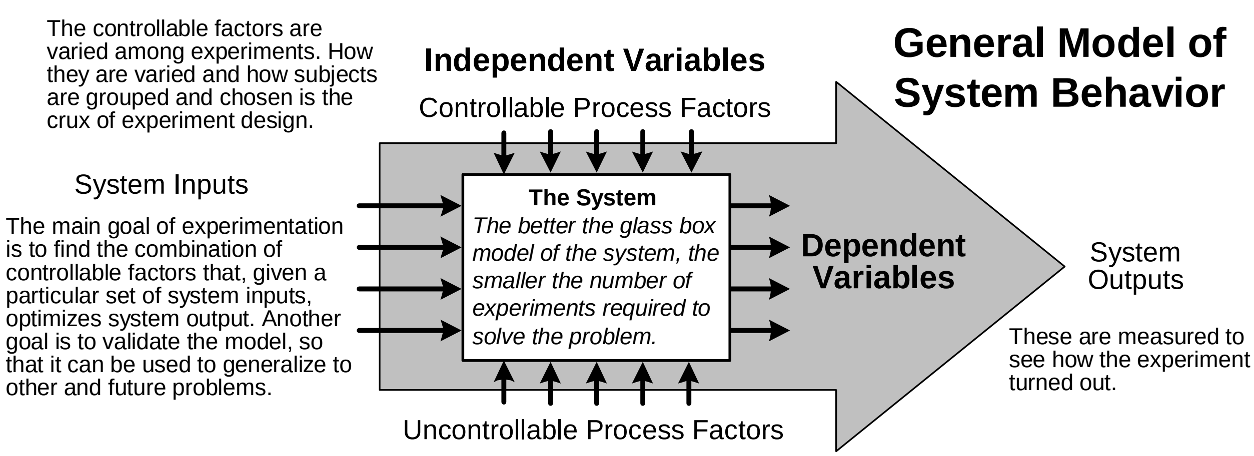
When trying to understand and improve their performance, complex systems are best viewed as elaborate ongoing processes. A **system process** consists of the process factors that turn system inputs into outputs. A **factor** is anything affecting the relationship between system inputs and outputs. In a complex social system problem the factors of interest are the social structure that is in place.

Consider the human system as an example. If we are trying to solve the sustainability problem, most human system factors are beyond our control, and are **uncontrollable process factors**. We cannot change people's normal life spans or the laws of physics. But we can change things like legislation, published articles, and the priority people give certain issues. What we can change are the **controllable process factors**. These are the **independent variables** we can vary in different experiments.

Notice the center of the diagram, which says that *the better the glass box model of the system, the smaller the number of experiments required to solve the problem*. This is the golden rule of advanced problem solving. The better the model, and the better your understanding of it, the more likely you will be able to find a small handful of key assumptions that, if proven to be true, would prove that the entire model is sound. *This can lead to ultra high efficient experimentation*, because instead of hundreds of experiments and decades of effort, you can prove the model to be sound with less than ten experiments, which can probably be run in less than a year. Each time the entire model is proven to be sound, it can then be used for a new, much deeper cycle of analysis or solution. Finally, when the last cycle occurs you are done.

For example, Einstein's general theory of relativity, first published in 1905, was an immensely complex model. In 1911 Einstein published a related paper calling on astronomers to test two key predictions of the theory. One was that gravity can bend the path of light, in an effect called gravitational lensing. On May 29, 1919 Arthur Eddington, then Secretary of the Royal Astronomical Society, supervised observational experiments of the bending of starlight as it passed near the sun during a total eclipse. The results matched those predicted by Einstein's general theory of relativity. This single experiment, combined with the elegance and explanatory power of the glass box model, was hailed as conclusive proof of general relativity over the Newtonian model.

As we proceed with this chapter on experimentation, notice how we are unable to talk about experimentation without talking about models. The two are



inseparable. Models generate hypotheses, and experiments test these hypotheses, which allow the models to grow and generate still more hypotheses, in a never ending cycle that drives the advance of science and civilization.

Einstein's models and those of the hard sciences tend to be equations that offer very fine grained predictive power. Those of the soft sciences tend to be general relationships, such as the way an ecological niche influences the evolution of those species competing to fill the niche. These more general relationships can be expressed in a large body of principles and facts, as seen in the fields of ecology and psychology. Or they can be expressed as mathematical equations if they can be assembled in a suitable model, such as system dynamics models. This allows the much greater predictive accuracy associated with the hard sciences to be applied to the soft sciences.

Using the General Model of System Behavior to Design Experiments

We will not go far into the statistical side of experimental design here. That is best left to those with specialist training. Instead, here we will cover the strategic design of experiments, which is where the battle is won or lost.

The general model allows us to intuitively see how experiments can be designed most efficiently. There are three main tactics to achieve high efficiency:

1. Test more than one controllable factor at a time – The goal of our collection of experiments is to find the optimum set of controllable process factors, because a complex problem will always require changing multiple factors to solve it. Therefore if we can figure out how to test multiple factors at once, instead of one at a time, it will take far fewer experiments to find the optimum combination. This can be accomplished by what are called factorial experiments.

A **factorial experiment** is an experiment in which two or more factors are varied at the same time. Looking at the tables of simulation runs in this book, each run was an experiment. Usually only one factor was changed from one run to the next, for ease of presentation and greater understanding of the model. But the most efficient way to find the optimum combination of factors to solve the problem would have been a collection of factorial experiments.

In The First Experiment only one factor is varied: exposure to the Truth Test. This is the old fashioned "one factor at a time" type of experiment. This is how most experimentation was done in the past, before factorial experimentation was invented by Sir Ronald Fisher in the 1920s at the Rothamsted (agricultural) Experimental Station in Hertfordshire, England.

Here's an example of how much more efficient the factorial approach is: Suppose we have constructed a model containing 11 key factors we suspect make the difference. Each factor can have two values, low and high. A "change one factor at a time" approach would take 2,048 experiments to find the optimum combination of the 11 factors, since the solution space size is 11^2 = 2,048 combinations. But using a factorial approach where we change multiple factors at a time, it takes only 16 experiments to do a rough first pass of the solution space. The data from the first pass can then be used to "home in" on the optimum combination with only a few dozen more experiments. This example shows how, for models with many factors, the factorial approach is an order of magnitude more efficient.

2. Use controlled experiments to eliminate the contaminating effect of all the uncontrollable factors and the controllable factors you are not testing – A huge problem facing experimenters is all those arrows on the top and bottom of the diagram. In the real world there are millions of them. Unless we can design experiments where only the factors we are testing matter, we will be unable to draw sound conclusions from the results, because the results will be contaminated.

To "control" the many factors that don't matter, our only recourse is to use what are called *controlled experiments*, as discussed earlier in this chapter.

3. Use the right sample size: not too many and not too little – The third strategic aspect of efficient experimental design is setting the required sample size (number of subjects) for statistical significance. We can never be 100% sure an experiment has proved anything. But we can be 90%, or 95%, or even 99.99% confident a hypothesis is true if certain conditions hold. These conditions depend on the sample size and the variation of the results. The larger the sample size or the smaller the variation, the higher the confidence level.

The general model helps us visualize the three things affecting system output. These are the system inputs, controllable process factors, and uncontrollable process factors. They are not uniform but vary continuously. For example, even the amount of solar radiation reaching the Earth varies by plus or minus 0.1%, due mostly to the 11 year sunspot cycle.

If there is no variation then a single measurement will tell how a set of controllable factors affected output with 100% confidence. But because there is always at least a little bit of variation, and sometimes a lot, multiple measurements are needed. This is why the sample size for an experimental group must be much greater than one. For example, in an opinion poll of a large population, a sample size of 1,000 is needed to give an error of plus or minus 3% at a confidence level of 95%.

If we do the above three things well, our experiments will lead to a correct analysis and a successful solution as fast and inexpensively as is humanly possible. But this will be true only when our experiments are guided by high quality hypotheses.

An Example of What Happens When There Is No High Quality Model for Guidance

High quality hypotheses can only come from high quality models. A **high quality hypothesis** is one that has a high probability of being true, makes a decisive contribution to an analysis or solution, and is based on reasoning that is easily inspected and highly convincing. The last trait makes it easy to pick which hypothesis to test next. High quality hypotheses cannot come from lucky guesses, because these have a low probability of being correct.

The Kuhn Cycle (presented on page 170) tells us that the models problem solvers use for guidance are continually decaying, as they drift from the Normal Science stage to the Model Crises stage. *This is where the model used by sustainability problem solvers is now, because the model is failing to show how to solve the problem.*

The Model Crisis stage is also where the model used by those attempting to end the Great Depression of the 1930s was. All initial attempts to fix the Great Depression failed, because there was no high quality model for guidance. No matter what governments did the problem deteriorated, such as in the United States, where: (Italics added)

> "In the nation as a whole, residential construction fell by 95%. Nine million savings accounts were lost. Eighty-five thousand businesses failed. The national volume of salaries dwindled 40%; dividends 56%; wages 60%.
>
> "And the worst of it, the most depressing aspect of the Great Depression, was that there seemed to be no end to it, no turning point, no relief. In 1930 the nation manfully whistled 'Happy Days Are Here Again,' but the national income precipitously fell from $87 to $75 billion. In 1931 the country sang 'I've Got Five Dollars'; meanwhile its income plummeted to $59 billion. In 1932 the song was grimmer: 'Brother Can You Spare a Dime?' National income had dwindled to a miserable $42 billion.
>
> "By 1933 the nation was virtually prostrate. The income of the country was down to $39 billion. On street corners, in homes, in Hoovervilles, 14 million unemployed sat, haunting the land. It seemed as if the proud spirit of hope had been permanently crushed out of America.
>
> "It was the unemployment that was hardest to bear. The jobless millions were like an embolism in the nation's vital circulation; and while their indisputable existence argued more forcibly than any text that something was wrong with the system, *the economists wrung their hands and racked their brains and called upon the spirit of Adam Smith, but could offer neither diagnosis nor remedy.* Unemployment—this kind of unemployment—was simply not listed among the possible ills of the system; it was absurd, unreasonable, and therefore impossible. But it was there." [105]

The economists wrung their hands and racked their brains because they had no model that could offer diagnosis or remedy. This forced leaders to take desperate, wild guesses at what to do, which made the depression even worse. In the end it was not a correct solution that solved the problem. It was luck. In the late 1930s along came World War II, and with it a sudden injection of massive spending into the system to fight the war. It was not until later that economists fully realized that this spending not only fought the war—it also fought the depression, and ended it with a proper dose of what is now called fiscal stimulus.

A new model explaining what had happened and what should be done emerged in the mid 1930s, when John Maynard Keynes published *The General Theory of Employment, Interest and Money*. But governments were skeptical about the model's remedy, even though its diagnosis was beyond question. In the US alone, business investment had fallen from $15 billion in 1939 to an appalling $886 *million* in 1932, a drop of 96%. The main diagnosis was that something had to prime the money flow pump until business investment returned to normal levels. There was no one else who could do this besides the government. (This is a contentious diagnosis, which shows how immature the model remains even today, 70 years later.)

But the new model was immature. Although it could tell that ending the depression required governments to spend more, it could not tell *how much* more. Nor did it say how the money supply should be managed. The tragic result was that, for example, the American Federal Reserve System turned a small depression into a big one by *decreasing* the money supply by one-third from 1930 to 1931, according to a modern analysis. [106] And since the model didn't say exactly how much government spending was required, governments spent too little. In the US, government spending rose from $10 billion on 1929 to only $15 billion by 1936. Europe and North America, where the Great Depression hit hardest, did not return to normal until World War II forced them to prime the pump harder than it had ever been primed before. Government spending in the US rose to the unheard of level of $103 billon, which totally eliminated unemployment and brought the US economy roaring back to life.

There was no shortage of suggested solutions when the Great Depression hit. But none were based on a proven model, so none were high quality hypotheses. The result was wild experimentation at the level of nations. Nothing worked. Only the fortuitous appearance of the Second World War solved the problem.

Today there is no shortage of suggestions on how to solve the sustainability problem. But to my knowledge, none are based on a proven model. The result is exactly what you would expect: wild experimentation while the problem grows worse. It is possible that the surprise appearance of a "wakeup call catastrophe" could wake the world up and cause an instant solution. But this is unlikely, because such a catastrophe would also weaken the human system so much, introduce so many new competing mega-problems, and delay the solution so long that critical environmental threshold points will have been passed, that solution of the sustainability problem would now be impossible.

The Role of Micro Social Experiments

How then are we to construct the high quality model we need to solve the sustainability problem?

This book is one long example of how that can be done. First you do what science did 400 years ago: switch to a process that fits the problem. Then you use the tools it takes to build the models you need. For difficult complex social system problems, the only known tool that can do this is system dynamics. Then you use that tool to build analyses and solution models. Finally you use experimentation to test these models as they grow. Once the models are mature they can be confidently used to solve the problem.

The limiting factor to this process is experimental efficiency. High efficiency can be achieved using two main strategies: (1) Find the key hypotheses that, if true, confirm the entire model. (2) Use *micro* social experiments instead of *macro* experiments.

Presently the common meaning of "social experiment" is **macro social experiments**. These test whether a particular government program (also called a policy intervention) will work on a major subsystem or the system as a whole. For example, the book *Social Experimentation and Public Policymaking,* by Greenburg and others, 2003, page 4, states that "social experiments" require that "treatment group members face changes in economic incentives, opportunities, or constraints" and that this excludes "trials that do not test policy interventions."

The drawback is macro experiments are terribly inefficient, because they only test final solutions. They cannot test the smaller assumptions that make up the model. The result is no glass box model for guidance on what solution will probably work best on what problem. Without a glass box model you can only guess at what to try, based on what did and did not work before. This is how evolution works, by trial and error. As evolution has shown, it eventually leads to amazingly powerful solutions. But as evolution also shows, this approach takes a very long time. We do not have that long to solve the sustainability problem.

Here's an example of how slow macro social experiments are:

> "The New Jersey Income Maintenance Experiment is generally regarded as the first large-scale social experiment. During 1968 and 1969, 1,300 low income families in five cities were randomly assigned to treatment or control groups, and the treatment group received negative income tax payments for the next 3 years. The effects of the experimental program on family members' employment and earnings, educational attainment, marital stability, and other behavioral outcomes were measured by the difference in subsequent outcomes between the two groups." [107]

This experiment took five years from start to finish. Now suppose there over 100 factors to test to solve the sustainability problem. Without a model, how many experiments is it going to take to stumble on the right combination of controllable factors that works? Remember now, each experiment will take an average of about five years. Because a trial and error solution space search technique is being used, the experiments must be fairly sequential. Suppose we get very lucky and find the solution in only ten experiments. That will take 50 years. That is so long that by the time the experiments are finishing up, the problem will have changed so much that the experimental results will apply to the old problem, but not the new problem. The problem will never be solved.

But it will be if a faster approach is used. If a model based approach is used and micro experiments are used to test the model as it grows, progress can be very rapid. A **micro social experiment** is a social experiment performed on a small group of people to test a particular direct (non-emergent) cause and effect relationship. The First Experiment is a typical example.

Micro social experiments can be performed in hours or days, unless retention is under consideration, which would change this to months. Thus micro experiments will average several weeks, instead of the several years of macro experiments. *This is a fifty-fold difference, and shows why problem solvers must take a model driven approach.*

Sadly, conventional wisdom is to *guess and test the guess*. Better would be to model and test the model, because solutions are an emergent property of the model.

The Incredible Power of Emergent Solutions

A key point to understand in this chapter is the incredible power of emergent solutions.

An **emergent property** of a system is a behavior that cannot be predicted from logical inspection of the parts. For example, no amount of inspection of the human system in 1928 predicted the Great Depression of 1929. Nor can how well an entirely new product will sell be reliably predicted from inspection of the product, consumers, and the rest of the system. The reason is that once a system passes a certain level of complexity, its behavior becomes greater than the sum of its parts, due to the many interactions between the parts.

An **emergent solution** is one whose success cannot be predicted by inspection of the problem. It is the success (or failure) of the solution that is the emergent property.

Sustainability is an emergent property of the human system. No amount of inspection will predict whether a growing society will be able to solve its sustainability problem before it reaches its limits. Likewise, no amount of inspection will lead to the exact solution that will work on the first try once a society has entered into unsustainability. Thus sustainability is an emergent solution.

This leaves problem solvers in a quandary. If inspection won't lead to solution, then what will?

Whenever you need to be able to predict emergent behavior, you need to build a model of the system involved and use the model for prediction. There is no other way.

This is exactly what system dynamics models do for complex social systems. Once you have defined the structure of the problem, calibrated the model with measurements from the real world, and verified that key assumptions in the model are true by the use of micro experiments, all you have to do to predict the future is run the model. Each simulation run is a cheap and quick **model experiment**. You can run thousands of them as you first find the root cause of the problem, then the high leverage points that need to be pushed on to solve it, and then how to best push on those points. Every model experiment tests how system structure and process factors affect the emergent behavior of the system.

The emergent solutions a model recommends must still undergo final testing and refinement. But the probability of them working the first time or needing very little refinement to work is very high. In fact, once the model is mature, it will churn out solutions to new problems with high confidence every time, just as the models of physics, chemistry, and the rest of the hard sciences have been doing for a long time.

Emergent solutions derive their incredible power from the fact that usually only a small number of factors need to be changed to solve difficult problems. For example, humans and chimpanzees share about 95% of their DNA. Another example is the way economic recessions and inflation can be tamed by pushing mostly on two points in the human system: control of the money supply and the federal funds discount rate. That this take very little effort shows these are very high leverage points.

The best emergent solutions push on a very small number of ultra high leverage points. Because of such elegant simplicity, they are easily managed and maintained over a long period of time.

The Ultimate Scenario: Using Micro Experiments to Build the Model Until the Emergent Solution Appears

The role of micro experiments and the power of emergent solutions explain why the long term strategy to solve the sustainability problem must center on using micro experiments to build the model until the solution emerges. We cannot predict how long this will take. But we can predict it will not take too long, because once a group of problem solvers deeply understands this strategy and the tools involved, and they have the financial means to support their efforts, their work can go as fast (or faster) as the project that put the first man on the moon in ten breathtaking years. Here is a scenario of how their work might go:

Suppose we have a guiding coalition and several dozen well managed non-governmental and governmental organizations around the world committed to solving the sustainability problem using the equivalent of Analytical Activism, the System Improvement Process, system dynamics, and ultra high efficiency experimentation. This will happen once the transformation problem is solved.

Some members of the coalition will focus on managing the overall project. The rest will play a multitude of specialist roles, just as NASA's many contractors developed and built the many elements required to put a man on the moon. Mega projects take grand visions. They also take one central growing thing to hold the entire project together. For NASA this was the glamour and excitement of space flight. For the sustainability coalition it will be a collection of computer simulation models whose explanatory and predictive power will stun the world, not with their ability to reach another planet, but to save this one.

The first of these models was born on June 29, 1970, when Jay Forrester,

> "...attended a meeting of the Club of Rome in Bern, Switzerland. [Their] orientation is activist—that is, they wish to do more than study and understand. They wish to clarify the course of human events in a way that can be transmitted to governments and peoples to influence the trends of rising population, increasing pollution, greater crowding, and growing social strife." [108]

Out of this meeting came World1, the first model to ever simulate "The Predicament of Mankind." It was created in a burst of insight and enthusiasm by Forrester. World1 soon became World2, as published by Forrester in *World Dynamics* in 1971. World2 was the starting point for World3, which was the basis for *The Limits to Growth*, which appeared in 1972. This book, and the model in it, went on to become the best selling environmental book of all time, with about 30 million copies sold to date. (The next closest book is *Silent Spring* with ten million copies.)

But World3 only *identified* the problem. It convincingly showed that if business as usual continues, environmental, economic, and population collapse is inevitable.

Some time before 2010 the Sustainability Coalition formed and launched its first project. *The goal was to first find a conclusive diagnosis for why, despite enormous effort by millions of environmentalists for decades, the sustainability problem was as unsolved as ever.* Why was the human system exhibiting such strong change resistance?

Sponsored by a number of sustainability oriented transnational corporations, the coalition contracted with eight organizations. Two of them (Let's call them team A and B) worked on competing models to see who could arrive at the most satisfying diagnosis. The rest launched a long series of micro experiments to test the fundamental assumptions of the models.

Team A had the great foresight to start by going all the way back to the Limits to Growth (LTG) project to determine why it was so successful. They quickly concluded that about 80% of the value of the project came from the World2 model, 10% from the World3 model, and 10% from the book. All the LTG team had basically done was to refine Jay Forrester's model into a bigger, more detailed, more accurate model that allowed more plausible projections, and communicate the model to the world in a well written book.

The team defined their strategy as "taking up where Limits to Growth left off." LTG had identified the sustainability problem. The next step, as the coalition had mandated, was to diagnose why the human system was unable to self-correct itself.

Buried in the back of the second edition of Forrester's *World Dynamics*, published in 1973, was a new chapter on *Postscript–Physical Versus Social Limits*. In it Forrester pointed out that: [109]

> "*The Limits to Growth* veered away from social and political factors to stress the more tangible physical aspects of the world environment. The debate about growth has centered on resources, pollution, and agriculture. But the most important issue is not the ability of technology to continue pushing back the physical limits. The question can be better stated, 'Assuming technology can continue to push back the physical limits of the earth, should society want to do so?' Relying on technology to solve the problems created by growth is to evade the question of how to slow growth."

The chapter also included a small model showing how social limits could be added to a model with only physical limits. From the model and the chapter the organization concluded they should, if possible, get Forrester to perform a second miracle: to sketch some foundational structures for the diagnosis and looking ahead, the cure.

By then Forrester was 90 some years old, but still mentally spry. He relished the idea of finally doing a serious job on the social side of the problem. In an intense weeklong workshop, Forester and five modelers pounded out several models that looked very promising. The group was elated. The models made gigantic leaps forward, leaps so large that the modelers admitted this would have taken them five or ten years. Some of the insights they would have never seen.

Meanwhile the other modeling team, team B, found an obscure paper on *The Dueling Loops of the Political*

Powerplace, first published on December 7, 2005 and later issued as a small book. It never caught on. The team wondered why, because it explained so much and offered an entirely new promising direction. One member pointed out the model's extreme novelty was precisely why it had failed to catch on—it was such a radically new paradigm that established environmental organizations could not see its potential, because they were so enamored by their own paradigms.

Working closely with the six experimental organizations, team B proceeded to subject the Dueling Loops model to a battery of grueling, high speed micro experiments. At first the model fared poorly in some areas. But a quick series of model evolution and experimentation cycles fixed the model's shortcomings. In a few months the team surprised itself. The model was starting to hold up quite well. It showed the root cause of systemic change resistance to be not just the basic Dueling Loops structure, and the Battle for Niche Succession between *Homo sapiens* and the modern corporation, but also the inability of cultural fabric to evolve fast enough to react correctly to a phenomenon they called *social stress*.

Getting a little excited now, the team presented their diagnostic results to the guiding coalition. The coalition faulted them for straying slightly from the process, but the team fired back that what they had done was *improve* the process by using a lower level of confidence to confirm initial hypotheses. This allowed them to explore a broader range of the solution space more quickly. The coalition was impressed, and the process was amended and improved on the spot.

The coalition then astonished the team by revealing that the other modeling team, team A, had come to one of the same conclusions that team B had. Social stress was a term Jay Forrester had used 40 years earlier! Team A had several promising models bubbling along that, while not as advanced and final as team B's model, were pursuing an entirely different line of analysis. The coalition would not say what they were, so that the two teams would continue their independent lines of inquiry.

It was now only six months into the project. The coalition put together a status report for the sponsors. It highlighted the rapid progress being made by the modelers and experimenters, the breakthrough insights gleaned from the week with Forrester, and the emerging model structures. Sponsor representatives requested and then sat patiently through an hour long presentation of the actual models and key experiments. They then peppered the coalition with questions and frank remarks that, in the words of one observer, "made it clear who really knew what was going on. The corporate sponsors had sent their best researchers, who had studied the status report beforehand. These were some of the sharpest scientists in the world. In a matter of minutes they pinpointed a number of flaws in the design of several experiments, and a notable omission in one of the models. I looked over and saw the eyes of one the presenters pop right open. He just about had a heart attack! But he also had the great presence of mind to thank them and ask if they would like to personally meet with the organizations involved."

They met. The corporate researchers explained the problems they had spotted to each of the eight contract organizations. This worked so well they arranged for frequent reviews at the team level instead of the coalition level. This process improvement effectively doubled team productivity, because now the teams had a new defect prevention loop that nipped bad ideas in the bud and shined a light on the ones that might have otherwise received little attention.

At 14 months into the project the diagnosis reached the clear and proven milestone. The two modeling teams had reached fairly identical conclusions, using what were mostly entirely different models. This was seen as very strong proof of a correct diagnosis. It was like having two different doctors examine a patient and agree on the same diagnosis.

The biggest surprise was that "more of the truth" was not the only low leverage point. Forrester's concept of social stress and the use of memetic infection had led to a much more subtle and powerful insight: Due to the way activists were framing the issue, it was seen as only "scientifically critical." But it was not seen as "practically critical" by the public. It was by the activists, especially those who understood the Limits to Growth model. But the sense of the sustainability problem being clearly critical enough to make it everyone's top priority was not present in the public, and therefore not in the media, and thus not in the minds of the world's politicians. All this was clearly shown in the models and proven by the experiments.

It was a flurry of micro experiments that had led to this amazing conclusion, one nobody had expected. Guided by a model structure with a loop called ***Rationalize Now to Make Doom Later Appear Irrelevant***, the experimenters had started with a concept linguistics professor George Lakoff called the improper metaphor. Employing Lakoff for his linguistic expertise, Dawkins for his memetic skill, and a few other luminaries, the experimenters had cracked open one of the greatest mysteries of social science: Why do so many societies end up marching to their collective doom, even when they are full of Cassandras pointing out that disaster is certain unless the course in changed in time? The question and its related high leverage point were labeled the Cassandra Mystery.

The micro experiments allowed the model's memetic nodes to be calibrated with remarkable precision. The model scenarios became fascinating excursions into high speed, controlled cultural evolution. The key run, affectionately labeled "the engineer's dream," showed how seemingly small structural changes caused new loops to become dominant in a multi-stage phase transition chain. One little group of memes grew and spread, and caused other groups to grow and spread, and so on. The memetic cascade, as the experimenters liked to call it, gently transformed the dominant memeplexes controlling the course of human civilization from a cacophony of destructively competing ones into a luxuriant, uplifting symphony of constructively competing ones. The emergent effect was the human system now wanted to aggressively optimize itself for the common good of all, instead of those with the highest current competitive advantage. The system didn't know how to do it at first, since it had never been done. But that didn't matter. The system always figured out a way. The effect was so dazzling to watch the experimenters ran the model over and over, in an insightful hypnotic trance.

The next step on the project plan was the Solution Convergence step. Because the high leverage points were so easy to understand, and because experimentation had already shown how the system would respond when the points were pushed on, this step took only a few months. Next came the Implementation step.

Like all great mysteries once they are solved, the answer to the Cassandra Mystery was so simple and desirable that even a ten year old could understand it and wanted to do it. Thus it was extremely easy to convince educators to revise curriculums to inject the memetic solution into society's meme stream at a young age. Injection into the adult meme stream was even easier. A few hurried press conferences arranged by the coalition were all it took to take the sustainability problem from "this is important" to "this is the only important problem we have" in the public's eye. This and parallel work on raising the public's ability to detect political deception quickly overcame the mountain of change resistance that had plagued environmentalists for decades. After that, the coalition hardly had to do a thing, because the human system had entered the Age of Transition to Sustainability.

One year later the coalition disbanded. Their own project post mortem showed that, just as predicted, the solution was an emergent property of the models they had been building. Everyone had thought that the solution convergence and implementation phases would take years or decades. But as soon as they had found the ultra high leverage point that the ***Rationalize Now to Make Doom Later Appear Irrelevant*** loop had led them to, the problem solved itself so fast they barely had time to blink.

Announced at a small press conference attended mostly by a tired but happy bunch of modelers, experimenters and their friends and families, the post mortem report attracted little attention. It was a quiet footnote to a quiet revolution.

Summary and Conclusions

This scenario ties together the essential elements it will take to solve the sustainability problem as fast as is humanly possible. These elements appear to be:

1. A **guiding coalition** to manage the large project.
2. An appropriate formal **process** driving the entire work effort. This is the Analytical Method, its domain specific subprocess, the System Improvement Process, and the many smaller processes, such as the Building the Knowledge Pyramid loop.
3. A formal **model** whose growth reflects the project's progress and whose growth builds the Knowledge Pyramid.
4. An unusually large amount of **experimentation** to turn Model Conjectures into Model Facts. Most of this is micro experiments.
5. Shifting gears into the new mode of **ultra high experimental efficiency**.
6. Reliance on the power of **emergent solutions**.

The last element is the end game masterstroke. A solution to a new problem can never be proven to work until after it has actually worked. But a modeled solution can be proven to have a high probability of working if the key assumptions in the model can individually be proven to have a high probability of being true. This proof is attainable only by experimentation.

Suppose there were ten key assumptions of equal importance. For the solution to have a 95% chance of working, the assumptions must each have a 99.5% probability of being true. [110] Achieving this very high level of confidence, 99.5%, is why it takes so much experimentation to build the model.

The sustainability problem will be solved only after a long series of ingenious, highly efficient experiments are used to build a model whose emergent behavior solves the problem.

All this begins with The First Experiment.