Chapter 10

The System Improvement Process

The System Improvement Process (SIP) was developed from scratch to solve difficult large-scale social system problems. The process was developed incrementally as this author went about trying to solve the sustainability problem. Since this was such an incredibly difficult problem the three steps of process driven problem solving were followed:

1. Identify the problem.
2. Choose or develop a suitable process for solving this type of problem.
3. Execute the process.

There was considerable iteration between steps 2 and 3. Whenever I found myself failing for long in step 3, I went back to step 2, improved the process, and then continued. One rule in a process driven approach is to always be executing particular steps of the process. Another rule is to always be improving the process. *The better a process fits the problem, the more productive it can be.* But a good process is not a panacea. Quality of execution is just as important, because:

quality of process x quality of execution = quality of results

Any process that fits the sustainability problem will do. SIP is a mere example, one that may give you ideas for an even better process.

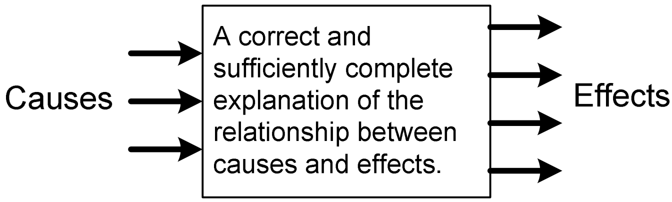
SIP is a *glass box method*. Inside the analysis is a model that lets you see why the inputs cause the outputs. This is infinitely more productive for difficult problems than *black box methods*. These leave the inside of the analysis box empty or in the case of *gray box models*, incomplete.

SIP allows problem solvers to avoid committing The Six Deadly Sins of the Wrong Process (as listed on page 95) and more. It does this by problem decomposition into a structured series of steps that produces a glass box model and considers root causes, high leverage points, and change resistance. The process allows advanced problem solvers to implement the following two step strategy:

1. Proper problem decomposition, so you can build:
2. A sufficiently complete model of system behavior.

Achieving number two gives you a glass box model of how the portion of the system with the problem works. This changes everything because you now have "A sufficiently complete explanation of the relationship between causes and effects" as shown on the next page. Night will become day. Problematic dragons will flee or fall by the wayside as you step confidently forward with a solution that, because others can also see in the glass box, will be accepted by your fellow problem solvers, backers, funders, and decision makers. One by one, they will join you in your march to success because they can all *see* your solution will work.

Glass Box Model



How will the Sir Lancelots and Lady Guineveres of the 21^{st} century build what's inside the glass box? To be honest, it's not as easy as the tales of the Knights of the Round Table would have you believe. One has to select the right armor, train diligently, find the right horse, and many other essentials. SIP offers a simple, repeatable way to do this.

The four main steps of SIP

The System Improvement Process is derived from the natural way people go about solving problems. They intuitively apply these four steps:

The Universal Problem Solving Process

1. Identify the problem
2. Understand the problem
3. Decide how to solve it
4. Implement the solution

The process of solving any kind of problem consists of first noticing the problem, then understanding the problem, then deciding what to do, and then doing it. People routinely zip through these steps so fast they solve little problems in seconds, like the sudden appearance of a puddle of spilled milk. Bigger problems take longer, but the basic process is always the same.

This is so universal it applies to all types of problems. It's the core on which all problem solving processes are built. Everyone learns this process at some point in their childhood, usually unconsciously.

Imagine you're five years old again. You're walking along when all of a sudden you feel your left shoe is loose (step 1). You look down to understand the problem better (step 2) and notice your shoe is untied. This level of understanding is so complete that you immediately decide all you've got to do is stop and tie your shoe (step 3). You do that (step 4) and continue on your way.

Thirty years later you're an orthopedist. A patient complains of back pain (step 1). You ask her a series of questions and run some tests to understand and diagnose the problem (step 2). The results come back, you study them and then consult with the patient. Together you decide what to do (step 3). Then the patient receives the treatment (step 4). Some of the physical therapy doesn't quite work, so you return to step 2 and run a few further tests and ask a few more questions. It turns out the patient's occupation is aggravating the problem. The patient decides how to solve that part of the problem (step 3) and then takes action (step 4). A few months later, her back pain is gone and does not return.

This is the process everyone has used millions of times to solve problems. We do it so fast we rarely consciously enumerate the steps. But what if we're working on a problem where it doesn't work? Then what process should we use? We can't just wing it. All that will lead to is more solution failure.

Instead, we need to stop, improve the process to where it's capable of solving the problem, and only then return to solving the problem. Once we do this, we have broken through into the mindset of seeing problem solving as consisting of two distinct aspects: process design and process execution. *In difficult problems the limiting factor is usually maturity of process design*, rather than the dozens of other things we frequently blame, like lack of money, time, personnel, skills, contacts, strong opposition, deceitful opposition, competing problems, etc. Another way to look at this is:

process maturity x quality of effort = quality of results

If millions of smart activists are trying as hard as they can to solve a problem whose solution would improve the common good and are failing repeatedly, then the bottleneck cannot lie in quality of effort. It must lie in process maturity.

Using the universal problem solving process as a starting point, let's rename the above steps to nouns representing major project milestones. This gives us the four main steps of SIP.

The strategy is to continually amplify the value of your work as it flows through these steps, until by the time it gets to the last step it's so productive it has a high probability of solving the problem.

The Four Main Steps of SIP

1. Problem Definition
2. Analysis
3. Solution Convergence
4. Implementation

This amplification is accomplished by following the guidelines in each step.

This is not a "waterfall" process. Frequent iteration is the norm. When a step bogs down it's usually due to an error in a previous step. Back up, do the portion of that step again that is causing the later step to fail, and then continue. Here's a summary of each step:

1. **Problem Definition** – Definition is done once for the entire problem rather than for each subproblem since the definition of each subproblem is reusable. (The subproblems are explained later.) Problem definition identifies the problem in terms of the desired goal state of the system, which is the state a solution will move the system to. The standard problem definition format is "Move system A under constraints B from present state C to goal state D by deadline E with confidence level F." A problem is solved when a solution is created that will move the system to goal state D by deadline E with confidence level F and keep it there.

2. **Analysis** – Using some type of modeling, here you observe, measure, and experiment to capture the system's relevant structure. Analyze the system until its key cause and effect relationships are well understood. In particular, find the *root causes* of the problem and the *high leverage points* (HLPs) for resolving those causes. For difficult problems be prepared to spend about 80% of your time in this step.

3. **Solution Convergence** – A solution element pushes on a HLP. Using knowledge from the previous steps, converge on a set of solution elements by artificial world or small-scale real world experimentation and further modeling as necessary. If the system is well understood this step will be relatively straightforward.

4. **Implementation** – Implement the solution by scaling up the experiments used to produce the solution elements. In a traditional intuitive problem solving process this step is the hardest. But in SIP it's the easiest, because by the time you get this far your cause and effect knowledge is so sound that how the system will respond to solution intervention is fairly predictable.

These four steps give the basic SIP grid:

The four main steps of SIP

1. Problem Definition	**The System Improvement Process (SIP)**
2. Analysis Spend about 80% of your time here. The problem solving battle is won or lost in this step, so take the time to get the analysis right.	
3. Solution Convergence	
4. Implementation	

For difficult problems this is not enough. SIP assumes a problem needs further decomposition. Over the years this has led to what SIP calls:

The three subproblems of the main problem

SIP strategically decomposes problems into the three sequential subproblems present in all difficult social problems:

A. How to overcome change resistance

B. How to achieve proper coupling

C. How to avoid excessive model drift

The decomposition uses three powerful, carefully designed abstractions. These allow problem solvers to approach the overall problem more productively. The three subproblems are:

Subproblem A – Change resistance has already been discussed in *Sin 5. The process is blind to change resistance*, on page 102. **Change resistance** is the tendency for a system to continue its current behavior when force is applied in an effort to change that behavior. Change resistance, whether high or low, must be overcome before the system will accept proper coupling solutions.

It is critical to differentiate between systemic and individual change resistance. The quote about Kurt Lewin's research on page 102 stated that "…it is possible for the resistance to be sited within the individual, but it is *much more likely to be found elsewhere in the system*." Systemic change resistance originates from a system, while individual change resistance originates from individual people and organizations. Usually systemic change resistance is ignored when solving a social problem, due to the strong tendency to make Fundamental Attribution Error, as described on page 269. In difficult social problems the root causes of change resistance are almost always systemic. Intermediate causes, however, are usually individual.

Subproblem B – **Proper coupling** occurs when the behavior of one system affects the behavior of one or more other systems in a desirable manner, using the appropriate feedback loops, so the systems work together in harmony in accordance with design objectives. For example, if you never got hungry you would starve to death. You would be improperly coupled to the world around you. In the environmental sustainability problem the human system is improperly coupled to the greater system it lives within: the environment.

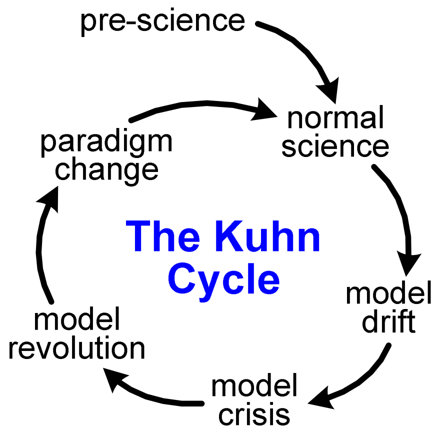
In difficult social problems, change resistance is the crux. It must be solved first because until change resistance (whether it's high or low) is overcome proper coupling is impossible. But that's not how problem solvers are working. For example in the environmental sustainability problem, in all cases I know, activists are attempting to solve the proper coupling part first and are treating change resistance as a minor hurdle that can somehow be overcome. This is being done without realizing it. Hence the need for the right decomposition.

Proper coupling is a crucial abstraction, one that activists must add to their mental models of the world if they are to have any rational hope of achieving their objectives. The abstraction allows you to more clearly differentiate between proper coupling and change resistance, and to avoid falling into the trap of seeing proper coupling as *the* problem to solve. It almost never is in difficult social problems, because high change resistance is what makes nearly all social problems difficult. This is because if change resistance is low, the system will "want" to be properly coupled. Solving the problem in this case is relatively easy, because there are usually a vast variety of proper coupling mechanisms (proper practices) available.

Subproblem C – Next we explain a more subtle concept: model drift. All conscious decisions are based on mental models of the world around us. As individuals and groups develop solutions to problems, we develop mental models (frequently augmented with physical models) of the problem and solution. When the model is sufficiently mature we implement the solution. If something goes wrong we go back to the model, improve it, and try again. Large mental models are a synonym for paradigms.

Model drift occurs when situations appear that the solution model cannot handle and the model is not or cannot be patched up to accommodate them. If the exceptions are relatively small, the model is still useful and model drift is said to have occurred. But if the exceptions accumulate and become major, then the model is now a hindrance to those using it. Excessive model drift has occurred and the model is broken. It's so useless the field is in crisis. This may or may not be noticed by some or even the majority of model users, who often erroneously claim the present model still works.

In order to more fully understand model drift we need to review the Kuhn Cycle. This was described by Thomas Kuhn in 1962 in his magnum opus, *The Structure of Scientific Revolutions*. Kuhn showed that all scientific revolutions go through a predictable pattern consisting of one pre-step to initiate the cycle and four steps for each revolution of the cycle.



The chief result of each scientific revolution is paradigm change, so the cycle is called the cycle of scientific revolutions or more commonly, paradigm change. According to Kuhn, "A **paradigm** is what members of a scientific community, and they alone, share." However, the cycle applies to more than scholarly science. It applies to any shared paradigm used by a group to achieve its mission.

To make the cycle more applicable to SIP we've added a fifth step, model drift. This gives the cycle five steps and leads to a key insight: *If excessive model drift can be prevented, such as by continuous self-improvement of a solution, then model crisis will never occur.*

Step 0 – Here's how the Kuhn Cycle works: Before a paradigm first exists it's in pre-science. According to Kuhn, **pre-science** is "disorganized and unstructured activity characterized by total disagreement and constant debate over fundamentals, such as optics before Newton."

Step 1 – But once that crystallizes into a clear vision of how to move forward together, the paradigm is born and the cycle moves into the normal science step. To Kuhn **normal science** is "structured activity that is directed by a single paradigm, which is uncritically accepted by the vast majority of the scientific community." [61]

Step 2 – Paradigms/models are rarely stable because social systems are always evolving and new knowledge is always being discovered. Particularly when a paradigm is young, **model drift** will soon begin. Situations and phenomenon will appear the model cannot handle. This is common. If these anomalies are small the model can be patched up, which takes the state of the model back to normal science. But over time, the exceptions the model can't handle accumulate. When they become excessive, the model crisis step begins.

Step 3 – In the **model crisis** step the model can no longer serve as a reliable guide to decision making (or in the case of a self-managing solution, as a driver for system behavior). This throws those using the model into confusion and crisis, because now they have nothing to base rational decisions on. They are intellectually lost when it comes to interpreting the world and deciding how to solve problems, and are forced to either guess or do nothing. As the crisis grows, new models are proposed that may or may not accommodate the exceptions, which Kuhn calls anomalies. Once the effort to find a new model that works begins to dominate debate, the next step, model revolution begins.

Step 4 – The **model revolution** step is revolutionary because it takes radical change to conceive of the totally new ideas necessary to accommodate all the accumulated anomalies. This step can be quite tumultuous and prolonged. It can take years, decades, or even centuries for innovators to arrive at a new model that successfully integrates most of the anomalies into a new conceptually complete model. This step ends when the new model is agreed upon by the innovators in a field. The new model may not yet be mature, but it's clearly better than the old one. This causes the next step to begin.

Step 5 – In the **paradigm change** step the innovators begin spreading the new model to others. People's mental models change from the old to the new paradigm. This is usually difficult for most individuals and groups due to long habit, social norms, and invested egos. In many cases change is impossible, so this step remains incomplete until most influential believers in the old model have died off and the new model has been taught to the new generation. But in other cases intense pressures may hasten paradigm change, such as if catastrophic failure faces a field if it cannot solve a central problem.

Step 1 – The Kuhn Cycle completes when the new paradigm becomes the new **normal science**. It's the normal way most of a field's members look at the world. The old paradigm has been tossed on the rubbish heap of history. The field is now productive again, but this time even more so. The cycle then continues because *Homo sapiens*, knowing man, is always learning.

Here's why the Kuhn Cycle is so useful for design of SIP: It helps us understand attempts to solve large social problems. Generally solutions go through a series of stages of maturity. Each stage is a Kuhn Cycle. The solution works at first, so it looks like a good new model. But as its shortcomings appear, and the system starts reacting to the new solution, model drift begins. It grows until model crisis happens: the solution is obviously no longer working. This throws problem solvers into a crisis of activity as they struggle to salvage the old solution model, which, after fifty piled on fixes, they discover is no longer humanly possible. When that undeniably bleak conclusion works its way up into a field's consciousness, the model revolution stage begins. What new model might work? How can it cope with all the things the old model could not? When these questions are answered and consensus emerges on what the new solution model should be, the cycle enters the paradigm change stage. There the new model is taught to practitioners. Selling the idea that the new model is better than the old occurs. Because of paradigm change resistance, especially among those who have used the old model for a lifetime, the paradigm change stage frequently takes a generation or more, because of the need to wait for the old generation to die off and take their old models with them to the grave.

The tremendous problems encountered in moving rapidly through the cycle are explained in this passage from the Wikipedia entry on Kuhn's work:[62]

> According to Kuhn, the scientific paradigms preceding and succeeding a paradigm shift are so different that their theories are incommensurable — the new paradigm cannot be proven or disproven by the rules of the old paradigm, and vice versa. The paradigm shift does not merely involve the revision or transformation of an individual theory, it changes the way terminology is defined, how the scientists in that field view their subject, and, perhaps most significantly, what questions are regarded as valid, and what rules are used to determine the truth of a particular theory. The new theories were not, as the scientists had previously thought, just extensions of old theories, but were instead completely new world views.
>
> Such incommensurability exists not just before and after a paradigm shift, but in the periods in between conflicting paradigms. It is simply not possible, according to Kuhn, to construct an impartial language that can be used to perform a neutral comparison between conflicting paradigms, because the very terms used are integral to the respective paradigms, and therefore have different connotations in each paradigm. The advocates of mutually exclusive paradigms are in an invidious position: *"Though each may hope to convert the other to his way of seeing science and its problems, neither may hope to prove his case. The competition between paradigms is not the sort of battle that can be resolved by proof."*

Hence the need for a self-managing solution, so that excessive model drift does not occur and we can avoid the problem of incommensurable new and old solution models. If we cannot build robust self-managing solutions, then large social problems will continue to go through endless cycles of solution, model drift, and solution failure.

If the model drift subproblem is not solved then the overall problem will eventually recur. If you are dealing with multiple problems, which all large social systems contain, multiple problem recurrence will eventually exhaust the resources available to solve problems.

This is the state many nations find themselves in today. They are barely able to cope with chronic old problems and a steady stream of new problems, and are forced into triage. Such nations are highly vulnerable to sudden collapse, because when one more new problem appears it can overwhelm a nation's problem solving capacity. This fate is not limited to poor nations. It also occurs in developed ones due to hitting various system limits and/or the cyclic re-emergence of dominant races to the bottom.

Let's review how the three subproblems work. Change resistance occurs when social agents refuse to adopt the proper practices needed to achieve proper coupling. There is always at least a little change resistance because otherwise the system would be unstable. Therefore subproblems A (how to overcome change resistance) and B (how to achieve proper coupling) are sequential and must be solved in that order. This is especially important for difficult problems, because high change resistance is usually what makes them difficult. Subproblem C (how to avoid excessive model drift) can be solved anytime and is usually solved last so as to achieve proper coupling as soon as possible. Thus in general the three subproblems are solved in the order given. *All three must be solved for the problem to be fully solved.*

In the short term change resistance is the crux. But in the long term the crux is model drift. Thus SIP views proper coupling as the relatively easy part of solving difficult social problems. It's easy because once change resistance is overcome the system will "want" to be properly coupled. It will stay that way indefinitely if excessive model drift does not occur. Change resistance and model drift solutions form a stable sandwich that holds proper coupling solutions where they should be. It's a tasty sandwich, once assembled.

In any difficult social problem there will be at least one change resistance, proper coupling, and model drift subproblem. The power of the three subproblems arises from the way they are far more easily solved separately. If difficult social problems are not decomposed in a manner similar to the one presented here, they are Gordian knots of insolvability.

Adding the three subproblems to the basic SIP grid gives us this one:

The four main steps of SIP

The three subproblems of the main problem

1. Problem Definition	The System Improvement Process (SIP)		
Subproblems	A. Change Resistance	B. Proper Coupling	C. Model Drift
2. Analysis Spend about 80% of your time here. The problem solving battle is won or lost in this step, so take the time to get the analysis right.			
3. Solution Convergence			
4. Implementation			

Let's take a look at each of the four main steps in detail:

Main Step 1. Problem definition

The entire problem is defined once. The three subproblems have reusable definitions, so they don't need to be defined.

Solving a formidable social problem begins with defining the problem in a manner that enhances all subsequent effort. The guiding principle of this step is: *The more correctly a problem is defined, the less work required to solve it.*

Difficult complex *s*ystem problems are best defined starting with this standard format:

> *Move system A under constraints B from present state C to goal state D by deadline E with confidence level F.*

The letters are the variables that are filled in to define a problem. Variables D, E, and F are the **solution goal**, the most important part. Constraints B include human resources, budget, conflicting goals, uncertainty, authority limitations, and so on.

Once you've filled in the blanks you have clear targets on which to focus your work in the remaining steps.

An example of defining a problem using this format is, "Given you are downtown and have lost your wallet and car keys, get home by 6:00 PM today with a 90% probability." Here system A is implied. It is the normal world you live and work in.

The **goal state** is the preferred state of the system, as opposed to the undesirable **present state**. The **state** of a system is the current values of the system's elements, such as your current location and who you know downtown. Implied is once the system enters goal state D, it must stay there.

In the above example the goal state is a Boolean value. A **Boolean** value is either true or false. Either you are home or you are not.

But in many cases goal state D must be expressed as a **range**, with lower and upper bounds. For example, you may want to keep from $1,000 to $5,000 in your checking account. Anything less is too risky, because who knows what expenses might appear suddenly? Anything more is losing money, because you could put it in an investment account and draw more interest.

One must be careful to avoid defining a problem that cannot be solved, or if solved creates other problems. Please note the format is flexible.

A problem is "solved" when a solution is created that will move the system to goal state D by deadline E with confidence level F. Solution occurs in stages, using the steps of SIP. Precisely when a problem is solved depends on the definition of "solved." I feel a problem is solved when you can predict with confidence level F that it will be solved, regardless of which solution stage that occurs in. The more mature and predictable your problem solving process is, the earlier in the solution stages a problem can reliably said to be solved.

Implied is the system will remain in the goal state indefinitely or until the end of the system's natural life. If the goal state needs to be maintained only temporarily, then modify the problem definition format. Use the format "…move to goal state D for X period of time…"

If your problem does not fit the standard format well then devise a suitable format. The important thing is to have a written, unambiguous, *measurable* definition of solution success. This allows subsequent work to be much more focused and efficient.

Main Step 2. Analysis

Design of SIP faces one final puzzle. Step 2, analysis, contains too much of a magical leap. How exactly do you find the root causes and high leverage points? How do you build a simulation model in an efficient manner? This is anything but obvious, so the analysis step contains five substeps:

A. Find the immediate cause of the problem symptoms in terms of the system's dominant feedback loops.

B. Find the intermediate causes, low leverage points, and symptomatic solutions.

C. Find the root causes of the intermediate causes.

D. Find the feedback loops that should be dominant to resolve the root causes.

E. Find the high leverage points to make those loops go dominant.

The goal of analysis is to understand the system's structure so well that its behavior becomes obvious. This will cause two supremely powerful insights about the problem to emerge. The first is that because the system's *intermediate causes* and *low leverage points* are now so clearly revealed, it becomes perfectly obvious why we have been failing to solve the problem. The second is that because we can now see where the *root causes* and *high leverage points* are, how to solve the problem becomes a relatively simple matter of determining how to best push on the correct high leverage points.

This is why if we do a good job of step 2 then the remaining steps, solution convergence and implementation, are relatively easy. That's why step 2 is also called *analysis*. This strategy forms the very heart of why the System Improvement Process is so powerful.

To execute this strategy, problem solvers should spend approximately 80% of their time in the analysis step. They need not spend most of their of time in implementation (as they usually do), because of the way public interest activist problem solving works. If initial solution success occurs then government assumes solution responsibility. After that activists can move on to other problems. Once most of these are solved they can lean back and enjoy the quality of life they have helped bring to the human system. This, rather than endless problem solving and tinkering with the system, is the goal of true altruists and true self-actualization.

The behavior of complex social systems, from families all the way up to world civilizations, is driven by feedback loops. *It follows that if you don't understand a social system's feedback loops, then you don't understand the system.* Normally these loops are invisible, which is what makes solving social problems so challenging. We are trying to fathom and influence something we can't see. It's as if we were up against an invisible opponent who can see us but we can't see him. The outcome of such a lopsided contest is a foregone conclusion.

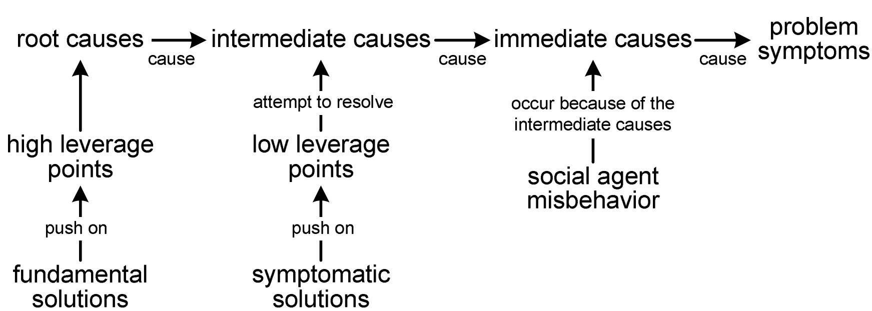
But that would change if we could see those feedback loops and how they work. Here's how the five substeps allow us to do this:

Substep A. Find the immediate cause of the problem symptoms in terms of the system's dominant feedback loops. – Using your problem definition, list the problem symptoms. Then create a model that reproduces those symptoms. Next study the model to see which loops are dominant and causing the symptoms. *Those loops are the immediate cause.* This completes step A.

This step differs dramatically from common sense approaches to solving social problems. For example, it's commonly assumed that the immediate cause of climate change is rising greenhouse gas emissions. But SIP says no, that's a superficial non-structural viewpoint, so shallow it will not help much in solving the problem. Better is to identify the feedback loops causing those emissions to rise, which is exactly what the World3 model in *The Limits to Growth* did. That's why that model was considered so profound.

Substep B. Find the intermediate causes, low leverage points, and symptomatic (superficial) solutions. – If we want to find out what to do right, it helps to first find out what we've been doing wrong. That's what step B does, as shown in the diagram below:

Expanded Root Cause Chain



Analysis starts at problem symptoms and works backward. The diagram shows how solution failure is due either to symptomatic solutions or poor implementation of fundamental solutions. The former is usually the case.

The purpose of step B is to find what problem solvers are doing wrong so it can be replaced by doing the right thing. This need is always present in difficult problems, because previous solution failure is what defines a problem as difficult. (The rare exception occurs when a mature process is used, and a normally difficult problem is solved correctly the first time. This is routine for organizations using mature processes, such as with full Six Sigma or level five of software's Capability Maturity Model.)

An **intermediate cause** sits on the causal chain between root causes and immediate causes. **Immediate causes** are the immediate source of symptoms. Take the case of a polluted river. Pollution is the symptom. The social agent misbehavior of polluting is the immediate cause. Why did that occur? That leads to the intermediate cause, such as pollution was the cheapest way for a farmer to grow crops. Why did that occur? If this line of questioning is rigorously followed *and modeled as you go* it will lead to the root cause.

Once the intermediate causes are found, next we find the attractive but low leverage points problem solvers have been pushing on and the symptomatic solutions they are using to do that. This is necessary, because part of the final solution will be to educate problem solvers to stop pushing on those low leverage points and switch to pushing on high leverage ones instead.

As your analysis technique becomes highly structured, such as by using a causal flow model or a simulation model, performing step B becomes easy. Like physicists, chemists, biologists, and ecologists, you are now thinking through the system's structure so thoroughly and correctly (We hope!) that the symptomatic solutions are blindingly obvious. So are the low leverage points they are pushing on. So are the intermediate causes those low leverage points are trying to resolve in vain. Because step B is usually so obvious it's the easiest step in the analysis.

Pushing on low leverage points usually makes the problem better in the short term (which explains its intuitive appeal) but worse in the long term. Pushing on low leverage points fails because **symptomatic solutions** treat the symptoms rather than the root cause. Here are three examples:

1. The early welfare programs of the United States increased dependency on the program, instead of decreasing it. The root cause was it was more financially attractive to not have a job, or not have a husband, etc, and to depend on welfare than it was to do the opposite.

2. Another example was early solutions to the urban decay crisis in the United States in the 1960s. An analysis by Jay Forrester of four of the top solution revealed that none were making the problem better and some were making it worse. The most favored solution of them all, lost cost housing, was the worst of them all. It made the urban decay problem much worse.

3. A third example is the way environmentalists have been pushing on the intuitively attractive but low leverage point of more efficient technology, such as the green revolution, more efficient cars with higher gas mileage, and renewable energy. But more efficient technology only raises the carrying capacity of the planet by reducing consumption per person and impact per unit of consumption (the AT in the IPAT equation). These are intermediate causes of environmental impact so this is a symptomatic solution.

More efficient technology does not decrease population (the P in the equation). More importantly, it does nothing to decrease the social forces causing the PAT factors to increase. The fundamental pressures causing impact growth have thus not been resolved, which will inevitably cause impact to rise. Thus in the long run, all more efficient technology does is *delay* overshoot and collapse. Furthermore, when collapse finally does come, it's much bigger. As the old saying goes, "The bigger they are, the harder they fall." Therefore in the long run more efficient technology makes the problem worse, not better. [63]

A **negative solution** makes a problem worse. A **neutral solution** has no effect. A **positive solution** makes a problem better.

The above examples are *negative solutions*. Due to the counterintuitive behavior of complex social systems, problem solvers routinely jump to the wrong conclusions about what will work. Jay Forrester, inventor of the type of simulation modeling used in this book, encountered this pitfall time and time again: [64]

> Commonly in complex systems a vicious cycle develops in which the action erroneously assumed to be corrective makes the problem worse and the worsening calls forth still more of the presumed remedial action, which only further aggravates the situation.

The way out of this trap is to perform a high quality analysis and find *positive solutions*.

Example of the Five Whys of Kaizen

Question 1: Why did the machine stop?
Answer: Because the fuse blew due to overload.

Question 2: Why was there an overload?
Answer: Because of inadequate lubrication.

Question 3: Why was it inadequate?
Answer: Because the lubrication pump was not functioning right.

Question 4: Why wasn't the pump working right?
Answer: Because the pump axle was worn out.

Question 5: Why was it worn out?
Answer: Because sludge got in.

The cause of negative solutions is usually misperception about what will work and what won't. If you can persuasively show activists promoting negative solutions why they won't work and how other positive solutions could, they will quickly change course. A classic example was Jay Forrester's work on urban decay.[65] By innovative use of system dynamics modeling, Forrester was able to educate urban planners and leaders about why popular solutions to the urban decay problem were negative or neutral, and how positive solutions could work. At first he encountered change resistance to his new way of thwinking. But he persevered and it was eventually accepted.

Substep C. Find the root causes of the intermediate causes. – Starting with the intermediate causes, employ Kaizen *and the model* to find the root cause(s). Don't, as many do, use just Kaizen.

Kaizen is gradual unending process improvement based on asking why. It teaches "problem solvers to ask why not once but five times. Often the first answer to a problem is not the root cause. Asking why several times will dig out several causes, one of which is usually the root cause." In the example shown in the box, "it was possible to identify the real cause and hence the real solution: attaching a strainer to the lubricating pump. If the workers had not gone through such repetitive questions, they might have settled with an intermediate countermeasure, such as replacing the fuse." [66]

A **root cause** (defined at length on page 98) is that portion of the model's structure that explains why the system's present behavior produces the problem symptoms. As the example of The Five Whys of Kaizen shows, you must dig deep to find the real root cause(s). As you dig you model. As you model you come to understand the system correctly more and more. This *analytical based* understanding gradually replaces the unsound knowledge that resulted from your earlier *intuitive based* understanding.

Digging for root causes is so important, so tricky, and so fraught with potential for hair pulling, frustration and failure, that the System Improvement Process offers the *three reusable subproblems* presented earlier. Each subproblem has enormously different dominant loops, root causes, low leverage points, and high leverage points. Finding these separately is several orders of magnitude easier than finding them by examining one big jumbled mess of a problem, which for difficult problems like sustainability and war has proven to be impossible.

A trap to avoid is treating social problems as technical problems with easy quick technical solutions. Social problems have *social* root causes that require *social* solutions. As you go digging for root causes ask questions like these: What are the dominant social agent types? What patterns of memetic transmission and infection are driving social agent behavior? What strategies for maximizing competitive advantage are social agents using?

The root causes of a difficult complex system social problem *must be systemic* because "Only a system level cause can actually be considered the root cause of a problem…." [67] Recall that for social problems, **systemic** means originating from the structure of the system in such a manner as to affect the behavior of most or all social agents of certain types, as opposed to originating from individual agents. Whenever you see most of a system's social agents misbehaving in the same manner, what you have is a systemic problem with systemic root causes.

Finding the right root causes is the key step in the entire process. Get it right and the problem is 80% solved because you'll be able to push on high leverage points with solution elements that are relatively easy to create.

Note what this step is doing. It's not finding the root causes of the entire problem, which is a hopelessly broad question. Instead, the process fires a laser beam by asking: What are the root causes of the intermediate causes of a single subproblem. This is a decidedly easier problem to solve by several orders of magnitude.

Because SIP tends to lead the analyst to the right root causes it can be very productive. Philip Bangerter, *the* first innovator and influential supporter of the concepts at Thwink.org, sees the output of the process as evidence of its strength. Here's how he describes SIP to others:

> The System Improvement Process is the engine in the car driving us on the road to the solution. We've got to get researchers to pop the bonnet, look at the engine, and open the door and get in the car. Then they will drive it away to solve the problem.

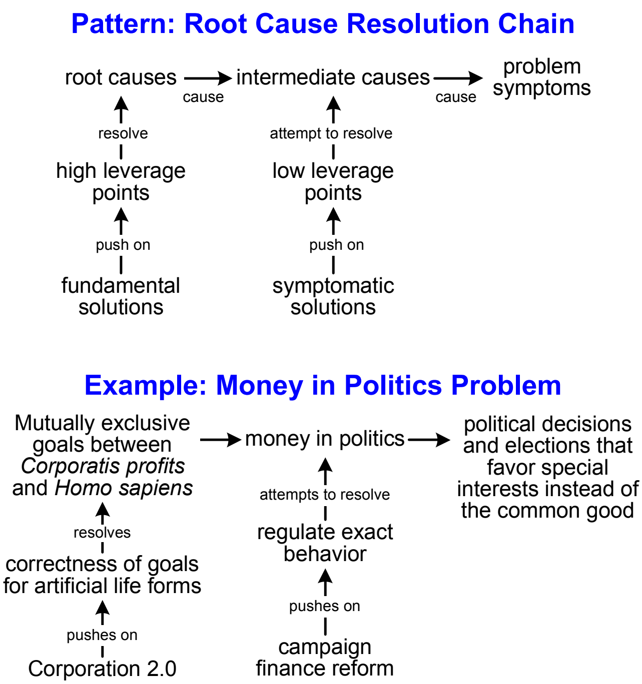
What a delightful metaphor!

The Principle of Root Cause Resolution

There's a fundamental principle enshrined in science, business, medicine, engineering, and a host of mature fields. This principle drives these fields to success. Without it they would be as lost as environmentalists are today:

> **Difficult problems can be solved only by resolving their root causes.**

This is the main principle SIP is built on, since SIP is nothing more than advanced root cause analysis. The principle explains why popular solutions to the sustainability problem or other difficult common good problems usually fail. It's because they do not resolve root causes, as illustrated below.



In the Pattern diagram, root causes cause intermediate causes which cause the problem symptoms. Correct fundamental solutions push on high leverage points to resolve root causes.

In difficult problems root causes are hard to find, so problem solvers tend to intuitively focus on resolving intermediate causes with symptomatic solutions. These are symptomatic because they try to make the problem symptoms go away by fixing immediate causes of the problem symptoms. This is done by pushing on low leverage points related to the immediate causes. Pushing on low leverage points requires much more effort to solve the problem because symptomatic solutions are not the only force trying to change the intermediate causes. So are the root causes.

A root cause has a far stronger and longer lasting influence on system behavior than a symptomatic solution. This explains why, if problem solvers hope to solve difficult social problems, they must first find the root causes, then find the high leverage points for resolving the root causes, and then develop the correct solutions for pushing on the high leverage points.

The Example diagram is an example of solution failure due to a symptomatic solution. The problem to solve is (falsely) assumed to be money in politics. The logic behind the symptomatic solution of campaign finance reform is that too much money in politics leads to too many political decisions and elections that favor special interests instead of the common good. Too much money in politics should not be the decisive factor. In a democracy, the correctness of the ideas various candidates put forth should be the decisive factor. If it's not then we have a broken political system. Therefore, the logic goes, if we could reform campaign finance the money in politics problem would go away.

But this seldom happens. Even if a reform bill can be passed against strenuous opposition, it is invariably circumvented, weakened, or eventually rolled back by powerful special interests. In particular, they continually find new way to use money to influence elections and decisions. This indicates that money in politics is an intermediate cause. There must be a deeper cause that causes special interests to behave the way they do.

If one goes looking for that deeper cause armed with a formal definition of what a root cause is and a suitable process, eventually you will find the root causes of the problem. Here the root cause is mutually exclusive goals between *Corporatis profitis* and *Homo sapiens*. This is the root cause of column B in the *Summary of Analysis Results* on page 202.

Compare the popularity of the (hopefully) correct solution of Corporation 2.0 to the symptomatic solution of campaign finance reform. This illustrates how hard it is to find the root causes of difficult social problems.

The two diagrams also explain the growth of solution complexity seen in most sustainability solutions, such as the mind boggling complexity of implementation of nation's Kyoto Protocol targets. The main reason most solutions are too complex is they attempt to substitute the endless complexity and inefficiency of command-and-control for root cause resolution. An ever growing mountain of regulations tries to command social agents to behave in desired ways, by pushing on the very popular regulate exact behavior low leverage point in the second diagram. After a solution fails either a new one replaces it or the complexity of the solution is increased in a vain effort to solve the problem. This fails over and over, since popular solutions treat only the symptoms of the real problem, i.e. the intermediate causes. They don't treat the root causes, which is the real problem to solve. The mountain of complex symptomatic solutions grows ever higher and more complex. This increases the size and complexity of the bureaucracy involved until finally both the solution and bureaucracy become unmanageable.

A few simple laws that resolve root causes are preferred to a mountain of complex laws that only partially fix an intermediate cause. This principle is embodied in the old British saying: "Do not judge us by the laws we have written, but by the ones we did not have to write." [68]

Substep D. Find the feedback loops that should be dominant to resolve the root causes. – In this step you find the feedback loops that, if you could change them to be dominant, would resolve the root cause and solve the problem. These loops usually already exist but are weak. In some cases these loops may not exist at all, such as the way the voter feedback loop did not exist before the invention of democracy.

If you have developed a model that mimics the problem symptoms and correctly pinpoints their root causes, the loops that need to go dominant are usually obvious.

Substep E. Find the high leverage points to make those loops go dominant. – Here you find the high leverage points that when pushed will make the feedback loops found in step D go dominant and solve the problem. This is the primary output of the analysis step so be sure to take the time to get it right. Step E exploits the **Principle of Social Problem Leverage**: *In problems where problem solvers are in the minority or lack governance of the system, root causes can be efficiently resolved only by pushing on their related high leverage points (HLPs).*

This step takes some ingenuity and contemplation, because you have many points to choose from. *The trick is to take your clues from where the loops that need to go dominant are.* They should be close by. If not, you are veering toward a solution strategy that is going to be tenuous, because it will depend on too many structural connections from high leverage points to dominant loops. Your modeling assumptions behind each connection have to be correct. The connections have to hold throughout the life of the solution. The path of connection has to be such that it is not easily circumvented by clever agents or destroyed by disruptions to the system. The longer the path, the more likely it is that model behavior will differ from that in the real world, and the more work you will have to do to reduce that uncertainty to an acceptable level. So try to pick short paths from the high leverage points to the loops that need to go dominant to solve the problem.

If your model is sound the right high leverage points should be surprisingly easy to find. They are much easier to find than root causes. If you have found the true root cause then your model either already has the high leverage points or will need only a modest amount of modification to add them.

The five substeps are a process for replacing a defective mental model with a sound one. This goal is the same one Jay Forrester has in mind: [69] (Italics and comments added)

> *Enhancing Mental Models* – Because of errors of dynamic interpretation in *mental models*, policy changes have often led to ineffective results, or worse, to the opposite of the intended results. [For example] A policy giving the opposite of the intended result was identified in *Urban Dynamics*. Economic distress in declining American cities in the 1960s generated symptoms of high unemployment and deteriorating housing.
>
> It appeared natural enough [due to a defective mental model] to combat such symptoms by government intervention to build low-cost housing. But the modeling study showed, as events have since confirmed, that such urban areas already have more low-cost housing than the economy of the city can sustain. Public policy to build more such housing merely occupies land that could instead have been used for job-creating businesses, while at the same time the housing attracts people who needed jobs. A low-cost housing program introduces a powerful double force for increasing unemployment, both by reducing employment while at the same time attracting people seeking work.
>
> Low-cost housing programs in inner cities became a social trap. The policy of building low-cost housing actually creates poor and unemployed people, rather than alleviating personal hardship. *The lesson here is to avoid attacking symptoms of difficulty until the [root] causes of those symptoms have been identified, and a high leverage policy has been found that will cause the system itself to correct the problem.*

That last sentence says everything this chapter has been trying to say. SIP allows problem solvers to "avoid attacking symptoms of difficulty" by finding the root "causes of those symptoms" and then the "high leverage policy" for resolving those root causes. If the process is executed well enough, the solution will "cause the system itself to correct the problem." The improved system will be in the right self-managing mode indefinitely, which is the strongest and most stable solution possible.

Main Step 3. Solution convergence

Using the analysis, this step converges on solution elements that can push on the high leverage points effectively. By comparison to step 2, step 3 goes quickly. In a complex social system there are countless possible high leverage points. But there are only a few realistic ways to push on each of them. High leverage points are high level solution strategies. Solution elements are tactical plans for implementing those strategies.

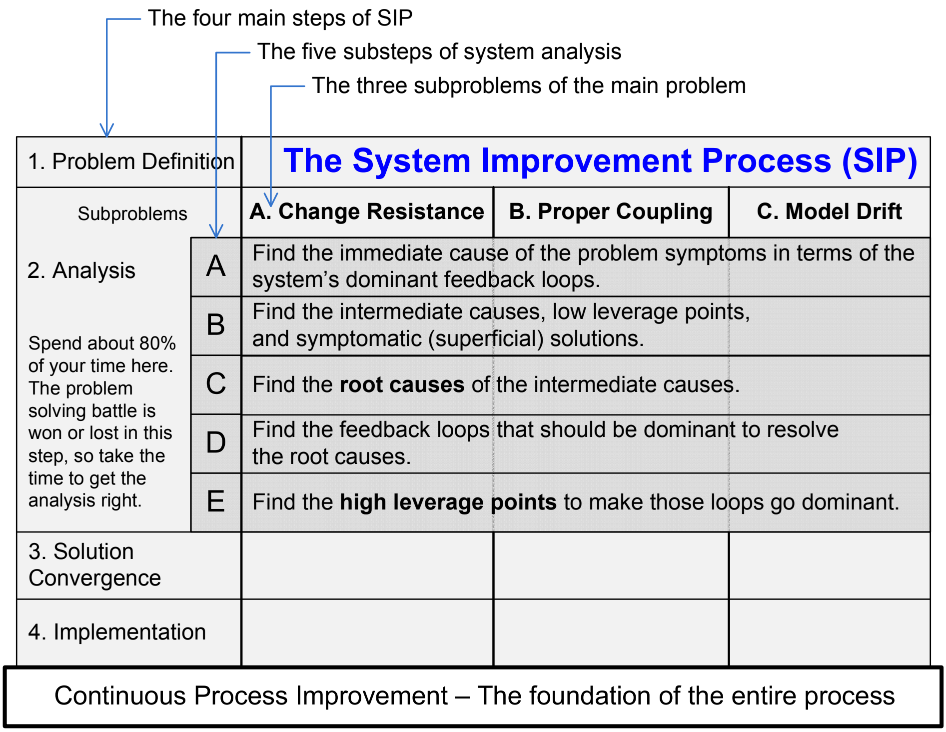
Solution element hypotheses require extensive testing. This can be done by expert opinion, modeling, social experimentation (laboratory), and real world experimentation (pilots). Much iteration with the analysis step will be required. It's worth keeping the analysis model up to date so that as solution model drift occurs, which it will, the solution can more easily be corrected.

Main Step 4. Implementation

Somewhere in the above hypotheses testing implementation informally begins. For large social problems, implementation formally begins when policy recommendations are made to government. If this is done using the above process, implementation should go smoothly and efficiently, perhaps astonishingly so. That is the payoff for using a process that fits the problem.

The Complete SIP Grid

With addition of the five substeps of step 2 and continuous process improvement we arrive at the complete SIP grid. It looks like this:



This is a powerful, complete process. It decomposes one big problem into three smaller subproblems, each of which is an order of magnitude easier to solve because we are no longer trying to solve them all simultaneously without realizing it. The process has four main steps. The first step, problem definition, applies to the entire problem. The other three steps apply to each subproblem. The second main step, analysis, is further decomposed into five substeps.

The complete grid contains 23 steps. The first step is problem definition. Each column has 7 steps, which gives 21 steps. Adding continuous process improvement gives a total of 23 steps.

The 23rd step, continuous process improvement, is the most important step of all. This step is what has taken SIP, the analysis, and the solution elements to where they are today. Continuous process improvement is *the* foundation of any highly productive process, so it sits at the bottom.

Conclusions

The purpose of this chapter is to show that for difficult problems, unless a solution is based on a process that fits the problem the solution will probably fail. It follows that if a solution avoids all Six Deadly Sins of the Wrong Process (listed on page 95) and is well managed the solution will probably succeed.

We therefore hypothesize that the set of solution elements presented in Part 3 will solve the global environmental sustainability problem if reasonably well refined and implemented because:

1. The solution is based on a **glass box model** of system behavior. We can see with our own eyes why the system is behaving unsustainably and why the solution will cause it to inherently behave sustainably.

2. The solution is based on **deep analysis**. Our defective mental models have been replaced with effective mental models.

3. The solution resolves the **root causes** of all three subproblems. Because the problem has been fully diagnosed, the treatment can fully cure the patient.

4. The solution pushes on the right **high leverage points**. Problems solvers will thus have sufficient force to make implementation work.

5. Solution strategy is based on full understanding of the very high **change resistance** present. That's why the strategy of striking where change resistance is low will work. (A later chapter describes this strategy.)

6. Because **the process fits the problem**, attempts to pound a round peg into a *round* hole will now succeed.

Number six is the heart of this hypothesis. The System Improvement Process is young. It may or may not be mature enough for a satisfactory fit. Further evolution is probably necessary. But unless this chapter contains grave errors, SIP appears to be a much better fit than traditional processes. It is to these that we now turn our analytical eye.

Introduction to Next Three Chapters

Past effort to solve the environmental sustainability problem may be divided into two types: research and application. **Research** attempts to use systematic investigation to establish facts, principles, and other knowledge that will be useful in solution development. **Application** is the design, implementation, improvement, and on going management of solutions based on knowledge from a variety of sources. Application requires research. Holding all else equal, the better the research the more successful the application.

Most fundamental social research on the sustainability problem consists of commons and economics research. Commons researchers view sustainability as what happens when a social/ecological system is qualitatively running smoothly. The emphasis is on quality of people's decisions. Economists view sustainability as what happens when an economic system is quantitatively allocating scarce resources efficiently, with emphasis on efficiency of transactions. The two approaches overlap, but each is attracted to very different aspects of the system for analysis.

The next three chapters examine three groups trying to solve the sustainability problem: commons researchers, economics researchers, and public interest activists taking an applied research approach. Each group uses a different process. *We conclude that none of these groups follows a process that fits the problem. This explains their history of continual failure.* There's a serious lesson to be learned here because:

> Progress, far from consisting in change, depends on retentiveness. When change is absolute there remains no being to improve and no direction is set for possible improvement: and when experience is not retained, as among savages, infancy is perpetual. Those who cannot remember the past are condemned to repeat it.

George Santayana ~ *The Life of Reason* ~ 1905