Alexander, Christopher
Notes on the Synthesis of Form
Cambridge: Harvard U. Press
1964

7 / THE REALIZATION OF THE PROGRAM

Finding the right design program for a given problem is the first phase of the design process. It is, if we like, the analytical phase of the process. This first phase of the process must of course be followed by the synthetic phase, in which a form is derived from the program. We shall call this synthetic phase *the realization of the program.*[1] Although these notes are given principally to the analytical phase of the process, and to the invention of programs which can make the synthesis of form a reasonable task, we must now spend a little time thinking about the way this synthesis or realization will work. Until we do so, we cannot know how to develop the details of the program.

The starting point of analysis is the requirement. The end product of analysis is a program, which is a tree of sets of requirements. The starting point of synthesis is the diagram. The end product of synthesis is the realization of the problem, which is a tree of diagrams. The program is made by decomposing a set of requirements into successively smaller subsets. The realization is made by making small diagrams and putting them together as the program directs, to get more and more complex diagrams. To achieve this we must learn to match each set of requirements in the program with a corresponding diagram.

The invention of diagrams is familiar to every designer. Any pattern which, by being abstracted from a real situation, conveys the physical influence of certain demands or forces is a diagram.

The famous stroboscopic photograph of the splash of a milk drop is, for certain purposes, a diagram of the way the forces go at the moment of impact. If you want to study these forces, this photograph, by abstracting their *immediate* physical consequences from the confusion of what you usually see when a milk drop falls, tells you a great deal about them.[2]

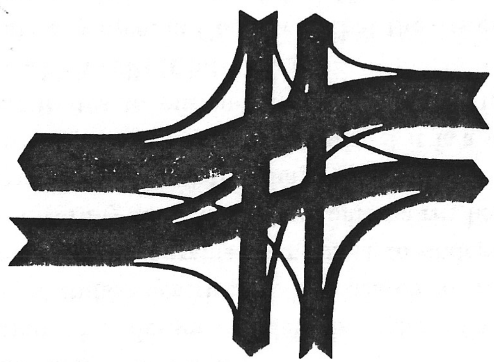
Le Corbusier's *ville radieuse* is a diagram, which expresses the physical consequences of two very simple basic requirements: that people should be housed at high overall density, and that they should yet all have equal and maximum access to sunlight and air.[3]

The sphere is a diagram. It expresses, among other things, the physical implications of the need to enclose as large a volume as possible within as small a surface as possible. It also expresses the implication of the requirement that a number of things be equidistant from a single point.[4]

The texture of bathers on a crowded bathing beach is a diagram. The evenness of the texture tells you that there are forces tending to place family groups as far as possible (and hence at equal distances) from one another, instead of allowing them to place themselves randomly.

An arrow is a diagram, of course, which conveys direction. Many flow problems contain requirements which can be summarized by means of arrows.[5] Very occasionally the form called for turns out to be physically arrow-shaped itself; like the case where the aerodynamic needs of a fast aeroplane are embodied in a swept-wing design.

is that today's traffic can flow without congestion. The requirement diagram, therefore, consists basically of information about how much traffic flows in various directions at different times of day. It is possible to present this information in a nonconstructive diagram by simply tabulating the flow numerically for each of the twelve possible paths, for different times of day. It is also possible, however, to present this same information in the condensed graphic form shown below.



Here we have a street map with arrows of various widths on it, representing the number of vehicles per hour flowing in various directions at peak hours. In this form the diagram indicates directly what form the new intersection must take. Clearly a thick arrow requires a wide street, so that the overall pattern called for emerges directly from the diagram.[9] It is both a requirement diagram and a form diagram. This diagram is a constructive one.

The constructive diagram is the bridge between requirements and form. But its great beauty is that it goes deeper still. The same duality between requirement and form which the constructive diagram is able to express and unify also

appears at a second level: the duality is itself characteristic of our knowledge of form.

Every form can be described in two ways: from the point of view of what it is, and from the point of view of what it does. What it is is sometimes called the formal description. What it does, when it is put in contact with other things, is sometimes called the functional description.

Here are some formal descriptions. A raincoat is three feet long, made of polythene $\frac{1}{2}$ mm thick, its sleeves cut in such and such a way, and so on. A salt crystal is a cubical arrangement of alternating sodium and chloride ions. A human body contains a heart, of such and such a size, in this position in the chest, a pair of kidneys rather lower and further back, and so on again. These descriptions specify size, position, pattern, material.

The corresponding functional descriptions tell you what happens when these objects are put in various contexts in the world. The raincoat is impervious to rain, and melts when heated. The salt crystal is transparent, conducts electricity slightly, dissolves in water but not in oil, shatters when hit hard with a hammer, and so on. The heart beats faster at high altitudes, the kidneys work when the body is fed.

In many of these cases we should find it hard to relate the two descriptions to one another, because we do not understand the objects thoroughly enough, and do not know, say, how the arrangement of atoms in a crystal relates to the solubility of the crystal in different solutes. However, for some very simple objects, there is virtually no rift between formal and functional descriptions. Take a soap bubble for instance, or a soap film on a wire frame. The behavior of soap films is so thoroughly understood that we know the economy of notation.[16] Like a hypothesis, it cannot be obtained by deductive methods, but only by abstraction and invention. Like a hypothesis, it is rejected when a discrepancy turns up and shows that it fails to account for some new force in the context.

The constructive diagram can describe the context, and it can describe the form. It offers us a way of probing the context, and a way of searching for form. Because it manages to do both simultaneously, it offers us a bridge between requirements and form, and therefore is a most important tool in the process of design.

In all design tasks the designer has to translate sets of requirements into diagrams which capture their physical implications. In a literal sense these diagrams are no more than stages on the way to the specification of a form, like the circulation diagram of a building, or the expected population density map for some region under development. They specify only gross pattern aspects of the form. But the path from these diagrams to the final design is a matter of local detail. The form's basic organization is born precisely in the constructive diagrams which precede its design.

What we must now see is that the constructive diagram is not only useful in probing the more obvious, known aspects of a problem like circulation, but that it can also be used to create the newly discovered implications of a new problem. We have seen that the *extension* of any problem may be captured by a set of requirements; and that by the same token any new set of requirements may be regarded as the definition of a new problem. Going one step further, the *intension* (or physical meaning) of a known problem may be captured by a

diagram; and by the same token the intension of any new, hitherto unconnected, set of requirements may be captured by a new diagram.[17]

The problem is defined by a set of requirements called M. The solution to this problem will be a form which successfully satisfies all of these requirements. This form could be developed, in all its important details, as a single constructive diagram for the set M, if it were not for the complexity of M's internal interactions (represented by L), which makes it impossible to find such a diagram directly. Can we find it indirectly? Are there some simpler diagrams which the designer *can* construct, and which will contribute substantially to his ability to find a diagram for M? There are; and the program tells us how to find them.

The program is a hierarchy of the most significant subsets of M. Each subset is a subproblem with its own integrity. In the program the smallest sets fall together in larger sets; and these in turn again in larger sets. Each subset can be translated into a constructive diagram. And each of these subsets of M, because it contains fewer requirements than M itself, and less interaction between them, is simpler to diagram than M. It is therefore natural to begin by constructing diagrams for the smallest sets prescribed by the program. If we build up compound diagrams from these simplest diagrams according to the program's structure, and build up further compound diagrams from these in turn, we get a tree of diagrams. This tree of diagrams contains just one diagram for each set of requirements in the program's tree. We call it the realization of the program.

It is easy to bring out the contrast between the analytical nature of the program and the synthetic nature of its realiza-

Now, as we know already, the set M consists of all those possible kinds of misfit which might occur between the form and the context; in the case of the kettle–metropolitan U.S.A. ensemble, this set includes specific economic limitations, technical requirements of production, functional performance standards, matters of safety and appearance, and so on.[1] To be exact, each element of M is a variable which can be in one of two states: fit and misfit.[2] It is important to remember that the state of this variable depends on the entire ensemble. We cannot decide whether a misfit has occurred either by looking at the form alone, or by looking at the context alone. Misfit is a condition of the ensemble as a whole, which comes from the unsatisfactory interaction of the form and context.

Take capital cost. The variable's two states are "too expensive," which represents misfit, and "OK," which represents fit. If a kettle is too expensive, this describes a property of the kettle plus its context — that is, of the ensemble. Out of context, the kettle's price either exceeds or does not exceed various figures we can name: nothing more. Only its relation to the rest of the ensemble makes it "too expensive" or "all right." In other words, it depends on how much we can afford. Again, take the kettle's capacity. If we look at the kettle by itself, all we can say is that it holds such and such a quantity of water. We cannot say whether this is enough, until we see what the context demands. Again, the fact that the kettle does not hold enough water, or that it does, is a property of the form plus context taken as a whole. This fact, that the variable describes the ensemble as a whole, and never the form alone, leads to the following important principle. In principle, to decide whether or not a form meets a given requirement, we must construct it, put it in contact with the

context in question, and test the ensemble so formed to see whether misfit occurs in it or not. You can only tell whether a kettle is comfortable enough to hold by picking it up. In principle, you can only decide whether a road is wide enough to drive down by constructing it, and trying to drive a car down it under the conditions it is supposed to meet.

Of course we do not stick to this principle in practice; it would be impossibly inconvenient if we had to. If we know the maximum width of cars to be used on the highway, and also know that for comfortable driving and adequate room for braking at a certain speed you need an extra 2′6″ on either side, we can tell in advance whether or not a given roadway is going to cause this kind of misfit or not. We can do so because the measurable character of the property "width" allows us to establish a connection between the width of the roadway and the likelihood of malfunction in the ensemble. What we do in such a case, to simplify the design task, is to establish a performance standard — in this case specifying that all roadways must have a minimum lane width of 11′0″ perhaps, because large cars are 6′ wide. We can then say, with a reasonable amount of confidence, that every road which meets this standard will not cause this misfit in the ensemble.

We can set up such a performance standard for every misfit variable that exhibits continuous variation along a well-defined scale. Other typical examples are acoustic separation of rooms (noise reduction can be expressed in decibels), illumination for comfortable reading (expressed in lumens per sq. ft.), load-bearing capacity required to prevent danger of structural failure (safety factor times maximum expected load), reasonable maintenance costs (expressed in dollars per the property "comfortable to hold" for kettles. There is no objectively measurable property that is known to correlate well enough with comfort to serve as a scale of "comfortableness." However, such a misfit variable can still be well enough defined. We can set up communicable limits which a group of experts can understand well enough to agree about classifying designs. We can certainly explain what we mean by comfort clearly enough, in commonsense language, for a group of people to learn to agree about which kettles are comfortable to hold, and which are not. This makes comfortableness an acceptable variable, for the purpose of the present analysis.

We shall treat a property of the ensemble (quantifiable or not), as an acceptable misfit variable, provided we can associate with it an unambiguous way of dividing all possible forms into two classes: those for which we agree that they fit or meet the requirement, which we describe by saying that the variable takes the value 0, and those for which we do not agree, which therefore fail to meet the requirement, and for which the variable is assigned the value 1.

This brings us to three questions, which may seem hard to answer.

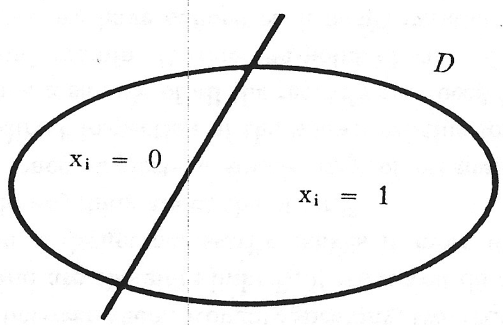
1. How can we get an exhaustive set of variables M for a given problem; in other words, how can we be sure we haven't left out some important issue?
2. How do we know that all the variables we include in the list M are relevant to the problem?
3. For any specific variable, how do we decide at what point misfit occurs; or if it is a continuous variable, how do we know what value to set as a performance standard? In other words, how do we recognize the condition so far described as misfit?

These questions have already been answered, substantially, in Chapter 2. Let us remind ourselves of the fundamental principle. *Any state of affairs in the ensemble which derives from the interaction between form and context, and causes stress in the ensemble, is a misfit.*

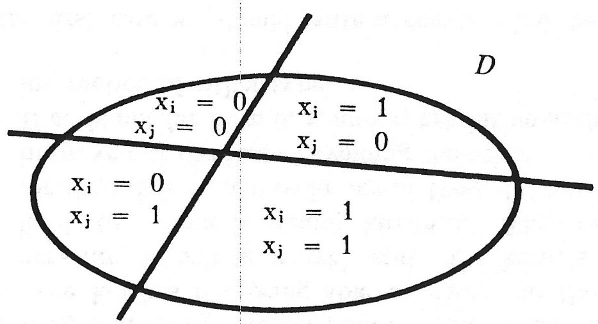
This concept of stress or misfit is a primitive one. We shall proceed without defining it. We may find precedents for this in the practice of common law, psychiatry, medicine, engineering, anthropology, where it also serves as a primitive undefined concept.[7] In all these cases, stress is said to occur wherever it can be shown, in a common-sense way, that some state of affairs is somehow detrimental to the unity and well-being of the whole ensemble. In design too, though it may seem hard to define the concept of stress in theory, it is easy in practice. In architecture, for example, when the context is defined by a client, this client will tell you in no uncertain terms what he won't put up with. Again, it is obvious that a kettle which is uncomfortable to hold causes stress, since the context demands that it should be comfortable to hold. The fact that the kettle is for use by human hands makes this no more than common sense. At the opposite extreme, if somebody suggests that the ensemble is stressed if the kettle will not reflect ultraviolet radiation, common sense tells us to reject this—unless some special reason can be given, which shows what damage the absorption of ultraviolet does to the ensemble.

This principle that stress or misfit is a primitive concept has the following consequences. First of all, it is clearly not possible to list all the types of stress which might occur in an main is imaginary, but serves to anchor the idea of intervariable connections. We should think of it as the totality of possible forms within the cognitive reach of the designer. In other words, it is a shorthand way of talking about all those discriminable forms which a designer can imagine and design.[8]

Now, we know by postulate, that we can in principle decide, for each one of the forms in D, which requirements it meets, and which it fails to meet. This means that each misfit variable x_i cuts the domain D in two: into a set of those forms which fit, and a set of those which don't. Schematically we show this:



From two variables we get four sets, in which the forms take values as shown below.



If we superimpose all m variables, we get a division of the domain D into 2^m mutually exclusive classes, each labeled by a different pattern of values for $x_1 \cdots x_m$. We shall call the proportion of forms in D which do not satisfy requirement x_i the probability of the misfit x_i occurring. We write this $p(x_i = 1)$. (Naturally $0 \leqslant p(x_i = 1) \leqslant 1$.) In the same way we define the probability of avoiding the misfit x_i as $p(x_i = 0)$; and the probability of avoiding both x_i and x_j simultaneously as $p(x_i = 0,\ x_j = 0)$, and so forth.

If the variables $x_1 \cdots x_m$ are all pairwise independent then it is an axiom of probability theory that we may write $p(x_i = 0,\ x_j = 0) = p(x_i = 0) \cdot p(x_j = 0)$ for all i and j. And similarly if the variables are also three-way, four-way and n-way independent, then these independence relations hold for the conditional probabilities, and we write, for example, $p(x_i = 0,\ x_j = 0 \mid x_k = 1) = p(x_i = 0 \mid x_k = 1) \cdot p(x_j = 0 \mid x_k = 1)$ conditional on $x_k = 1$ and so on.[9]

Wherever the variables are not independent, the above relations break down. Essentially, then, we speak of a dependence among two variables wherever $p(x_i = 0,\ x_j = 0)$ is markedly unequal to $p(x_i = 0) \cdot p(x_j = 0)$, and similarly for more than two variables. Formally, we describe these dependences by means of the correlation coefficients.[10] The simplest correlation coefficient is that for two variables:[11]

$$c_{ij} = \frac{p(x_i=0,x_j=0) \cdot p(x_i=1,x_j=1) - p(x_i=0,x_j=1) \cdot p(x_i=1,x_j=0)}{[p(x_i=0)p(x_j=0)p(x_i=1)p(x_j=1)]^{\frac{1}{2}}}.$$

For any pair of variables x_i and x_j, then, we may distinguish the following three possibilities.

1. If c_{ij} is markedly less than 0, x_i and x_j conflict; like "The kettle's being too small" and "The kettle's oc-

as soon as we begin to think of other materials and designs. This conflict certainly does not exist for all imaginable kettles.

Clearly we want to avoid muddling this last kind of case with the other two. If we were to accept the linkage it suggests, then together with the essential logic of the ensemble we should also be freezing in its most temporary incidentals. We are interested in those links between variables which hold for all forms we can conceive (that is, for the whole of D). Any sample based on those possible solutions which happen to have been constructed is heavily biased toward the past. To avoid the bias we should need either to examine all the members of D exhaustively or to find a theory which offers us a way of sampling D unbiasedly. Neither of these is practicable today.

However, we may overcome the bias by another means. Instead of just looking for statistical connections between variables, we may try to find causal relations between them. Blind belief based only on observed regularity is not very strong, because it is not the result of a seen causal connection. But if we can invent an explanation for inter-variable correlation in terms of some conceptual model, we shall be much better inclined to believe in the regularity, because we shall then know which kinds of extraneous circumstances are likely to upset the regularity and which are not. We call a correlation "causal" in this second case, when we have some kind of understanding or model whose rules account for it.

For example, the molecular and crystalline structure of materials gives us good reason to believe that the thermal conductivity of a material is the same in any two opposite directions, and hence that the need to heat a kettle quickly conflicts with the need to keep the water hot once it has

boiled. In this case, because we "understand" the connection between the two variables, we call it causal, and give it much greater weight — because we are convinced that it holds for almost all conceivable possibilities.

The search for causal relations of this sort cannot be mechanically experimental or statistical; it requires interpretation: to practice it we must adopt the same kind of common sense that we have to make use of all the time in the inductive part of science. The data of scientific method never go further than to display regularities. We put structure into them only by inference and interpretation.[14] In just the same way, the structural facts about a system of variables in an ensemble will come only from the thoughtful interpretation of observations.

We shall say that two variables interact if and only if the designer can find some reason (or conceptual model) which makes sense to him and tells him why they should do so.[15]

Again, as with the definition of the variables, this introduces a personal bias, and reminds us that L, like M, is a picture of the way the designer sees the problem, not an objective description of the problem itself. If the designer sees a conflict between the need to have sufficient capacity in a kettle and the need to conserve storage space, he does so because he has certain preconceptions in mind about the kinds of kettle which are possible. It is true that there are conceivable devices, not yet invented, for boiling water as it comes out of the faucet, and that these might take very little storage space. But until the designer understands this possibility, there is no point in telling him that the conflict is spurious; as far as he is concerned, there really is a conflict, which needs to be resolved, and therefore needs to be included in L and taken

From this matrix we define the set L as a set of links associated with the variables of M, as follows.[16] For every pair of variables x_i and x_j, there are $|\nu_{ij}|$ distinct elements of L which join x_i to x_j. These elements bear the same sign as the index ν_{ij}, negative for conflict, and positive for concurrence.[17] The sets M and L together, completely define the graph $G(M,L)$.[18]

The definitions we have given so far still leave certain practical questions about the sets M and L unanswered. Does it matter, for instance, if two variables are very close in meaning, though slightly different? How specific or how general must they be? What do we do about three-variable interaction? The answers to these questions depend on three important formal properties of the system $G(M,L)$, which we shall now explore.

First of all, if the graph $G(M,L)$ is to give us an accurate picture of the variables' behavior, it is necessary that the set L describe *all* the interaction between variables which there is. Since the elements of L are links which represent two-variable correlation, this means that the variables must be chosen to be free from three-variable and higher-order correlations. The mathematics of Appendix 2 is also based on the assumption that the higher-order correlations vanish.[19] If this is not so, any analysis based on M and L alone is sure to give misleading results.

Second, even the two-variable correlation $\delta\nu_{ij}$ must be small, for each pair of variables. Specifically, as far as the mathematics of Appendix 2 is concerned, we must have $l\delta \lessdot 1$, where l is the total number of links in L.[20]

Third, the analysis in Appendix 2 is also based on the assumption of a certain simple symmetry among the variables

of M. It demands that $p(x_i = 0)$ should be the same for all i.[21] Again, if this is not so, the analysis will be invalid.

Let us now consider the practical implications of these three formal properties which the system $G(M,L)$ must have. We take the last one first. It demands that $p(x_i = 0)$ should be the same for all i, or that the proportion of all thinkable forms which satisfy a requirement should be about the same for each requirement. What this amounts to, in common-sense language, is that all the variables should be roughly comparable in their scope and significance.

We cannot admit "economically satisfactory" as one requirement, and "maintenance costs low enough" as another. Plainly these have different degrees of significance, because the second is part of the first, while the first is not part of the second. Every design which is economically satisfactory must *a fortiori* have acceptable maintenance costs. But the reverse is not true. There are far more possible designs which meet the second than the first, because the first is much wider in scope and significance; their probabilities of occurrence are very unequal. In this case the inequality is especially clear because the second requirement is, as it were, contained in the first. But the difference would be just as great if we replaced the first by "functionally satisfactory." This is again wider in scope and significance than "maintenance costs low enough" even though it does not contain it. If we want to use "maintenance costs low enough" as one requirement, then we must break down "functionally satisfactory" into smaller, more specific requirements, comparable to it. The first step in constructing the set M is to make all its variables approximately equal in "size" or scope.[22]

Let us take the second of the three formal properties next.