Optimisation by Evaluation in the Appraisal of Buildings

by Thomas A. Markus

MA MArch (MIT), ARIBA, Professor of Building Science,
Department of Architecture and Building Science,
University of Strathclyde, Glasgow, C. 1.

Abstract

The paper describes the application of optimisation techniques based on cost benefit analysis. The cost of provision, maintenance and 'failure'-to-provide are used. The use of indifference sets and production functions for design decisions is demonstrated. There is a model of the design process which relates these techniques to those of analysis and design synthesis. The relationship of parts of the system to the whole, and of constraints, is examined in this context.

1 Introduction

This series of articles* has so far explored a number of concepts relating to value, cost, price and a number of techniques by which the designer can attempt to reach the 'best' solution — the one of maximum value. In this paper these ideas are developed in a somewhat different direction. First, a simple model of design decision-making is proposed, in which appraisal is identified as part of the decision process, and evaluation as a part of appraisal. Then there is a brief description of the total system over which the designer has freedom of choice. Next the idea of optimising the state of this complex, interactive system is explored. And finally, some practical and theoretical techniques which could be relevant are described.

2 The Design Decision Process

Most descriptions of design procedure — whether theoretical or based on empirical studies — recognise two basic patterns in the process[1]. (Figure 1). The first is a management process, according to which one phase follows another in time. It divides up the total time available to make the decision into phases which develop from the general and abstract to the detailed and concrete. The RIBA Plan of Work[2], stages A to E, is a well-known example of such a structure. The need to complete by a deadline is emphasised by rules that prevent going back. The attendant failures that would be involved in making this process an iterative one are obvious: continuous re-call of consultants, re-constitution of design teams, inability to key firmly into financial, legal and government action; inability to predict design time and costs; and the arbitrariness of any decision as to how many times to iterate. But the paradoxes of making this a linear process are also obvious: it is a deterministic procedure for an essentially non-deterministic problem. If a design solution can always be novel it is akin to exploration; neither the end of the process nor the resources needed can be predicted. It is rather as if the first trans-Atlantic explorer had been told "We know that there is a Continent to the West, but we do not know how far away it might lie. Here is a ship and 100 men; in 6 months time you much reach it." One could instruct him to reach his goal and offer the necessary and unlimited time to do so; or instruct him to sail for 6 months and report his position. But to attempt to specify both resources and achievements is to programme a paradox — to sail into unknown territory pretending that the journey is on a charted map. This weakness of the "vertical" process (morphology) is to some extent remedied by using an open-ended, iterative decision process at each of its phases — the "horizontal" process.

This decision process appears to have three main parts to it[3].

(i) Understanding of the problem (Analysis)
This includes the gathering of all relevant information; the establishment of relationships, constraints, objectives,

criteria — indeed the imaginative structuring of the problem. Often, well done, this leads to good and imaginative solutions.

(ii) Producing a design solution (Synthesis)

The problem structure may suggest part or whole solutions. There is a great body of literature and experience suggesting a rich variety of rational, intuitive, ordered and random processes which may be appropriate to different problems and different personalities. The process may result in a single design; or a variety of different designs; or a cluster of similar designs, being variants of a basic type.

(iii) Establishing the performance of the solution (Appraisal)

This is a retrospective act by which the designer establishes the quality of his solution. There are three basic steps in appraisal (Figure 2):

(a) Representation. The solution is modelled in any suitable way. The model might be verbal, mathematical, visual or even 'full-scale' (in this sense a building-in-use is a full and complete 'model' of the design).

(b) Measurement. This is a neutral activity in which the performance of the model is obtained on as wide a variety of counts as necessary. Costs, environmental conditions, flexibility, space utilisation, human response, are amongst those that suggest themselves easily. Such aspects of performance as depend on unpredictable human response — e.g., judgements of the formal qualities of an object, can be obtained by simulation and recording direct responses.

(c) Evaluation. The measured results are now evaluated; cost-benefit analysis; subjective value judgements; comparison with ideal, average or statutory performance standards found in the analysis; conformity to constraints recorded in the analysis — all these and other techniques are appropriate.

In the light of the appraisal the designer may wish to re-design, or even to re-examine and change or develop his analysis. He will iterate these sequences as often as necessary and as time allows and reach a decision.

The entire vertical and horizontal structure, with its iterative feedback loops, is simply illustrated in Figure (1); but the overall sequence followed by a designer may be represented by at least three routes which arise from the paradox of the supposed linearity of the morphology. This creates another, the "double pyramid", paradox. The development of a total design through its sub-systems, components and down to fine details implies, at the analysis stage, a decomposition process; i.e., breaking up the problem into finer and finer detail. At the synthesis stage it implies a clustering or recomposition process; i.e., a building of the total solution from its components. The three possible routes through these two pyramids are indicated by Figure (3).

In the first, the understanding of the problem is carried through all phases, from general to particular, before a design synthesis (with its appraisal) is sought. In this mode the problem of a city is analysed down to its last detail; and the first components to be actually designed would be the most detailed; doorknobs; the last, the complete land use and transport systems.

In the second route each phase of the problem is analysed and solved before proceeding to the next; this mode is the one most often associated with the structure of the design process illustrated in Figure (1).

In the third route, after complete analysis, the synthesis process starts at the general, strategic level and descends to the detailed component.

In each of these an appraisal follows a synthesis, although its nature will vary. Clearly the arrival of industrialised systems of hardware — in which components down to the last detail have already been designed and manufactured — has relevance to the choice of the most appropriate route.

3 The System being Optimised

Organisations which need buildings to house their activities have definable objectives to meet. In order to meet these, they need a certain pattern of activity and behaviour. The setting or environment for this behaviour will influence the degree to which it succeeds in meeting its objectives and will itself be the outcome of decisions about the physical form of the building. Thus the designer is dealing with a complex and interactive system which is simply illustrated in Figure (4)[4].

The objectives of the organisation may be social, economic or cultural (e.g., to do with health, education, commerce or leisure). The activity pattern required to meet these objectives may include activity directly related to production, or communication, control, identification (i.e., establishing the image and self identity of an organisation) and other definable, formal categories. Dependent on the nature of the objectives there will generally also be a host of informal activities which are often vital to the survival of the organisation, its morale and its communication system.

The environmental system consists of two main parts, the spatial environment and the physical environment. The former can be described and measured in terms of the size, number, form, type and linkages between, spaces; and the latter in terms of the visual, thermal, aural and other physical characteristics of the spaces. The building or hardware system can be conveniently described under the constructional system (fabric, structure, components etc.), the services system and the contents system (furniture, furnishings, equipment, fittings etc.).

Whilst the relationship between the four parts are complex they are generally causal in one direction, from left to right on the figure and derivative from right to left, except at the interface between environment and activity where they are mutually interactive and where continuous adaptation takes place in the sense that environments are constantly adapted to activity needs and activity modes are continuously influenced by or adapted to environmental realities.

Also on Figure (4) is included a cost system, for each of the four major parts has a cost, revenue or value attached to it. The hardware costs something to provide; the environment costs something to maintain (energy, cleaning, repair, maintenance); the activities cost something to provide (labour, materials, advertising, waste, etc.); and the objectives have a value in being partially or wholly met. If it were possible to establish all the functional relationships between all the parts of this system and to attach valid cost figures throughout, then the evaluation of any proposed design by cost-benefit analysis, as proposed by Fleming[5], would be feasible and a true optimum could be found. This possibility is discussed further below.

One other interesting propriety of this model and of this way of looking at buildings is that it shows how it is possible to talk about a whole building, or any, even minute, part of it, at any one of the four major levels. For example, it would be possible to describe the hardware of a light fitting giving details of the tubes, diffusers, dimensions and mounting. This would be a type I description. A type II description would specify the environmental output of the light fitting; e.g., its heat output, the emission of total luminous flux or its intensity in a given direction. Clearly this would leave the way open to the selection of alternative fittings with the same environmental characteristics. A type III description would describe the effects of the fitting in terms of human activity or behaviour. It could, for instance, describe a fitting producing illumination in which the maximum error in a standard task would be a specified percentage; or it could describe the score on a standard satisfaction or gloom scale; or the frequency distribution of discomfort glare judgement made by a panel of observers. Since these behavioural responses would also be affected by many other aspects of the environment, clearly the description would now become very open-ended and could lead to a wide variety of combinations of physical hardware sub-systems. At the fourth level one could simply describe, in objective terms, a lighting system; for instance, one that produced a given level of happiness, stability or education.' If appropriate language is used at each of these four levels, then the job of writing a specification — that is, producing a <u>pre</u>scription — and the job of satisfying that specification by a <u>de</u>scription — becomes relatively simple. The person or group needing a building starts from level four until they meet and overlap at one of the intermediate levels, the organisation supplying the hardware. To what extent the designer's role lies near one or other end of the spectrum is a question irrelevant to the present paper and obviously highly complex[6].

The last point to note about the description of buildings in these terms is that the statement of constraints, statutory or otherwise, can be made in suitable form at any one of the four levels or at the fifth level in terms of cost limits.

The use of this model, incomplete as it must be, can allow all present and future costs, revenues, resources, effects and values to be placed at the appropriate point in the system.

4 Optimisation

Although the word optimum is often loosely used, it must be clearly understood in the present context to refer to a single best solution. This in turn requires a single criterion for optimisation and in order to fit this article into the remainder in the series, the argument will be based on the use of money as a criterion and on maximising the value or minimising the cost of a solution as an objective. An equally strong argument could be developed in terms of using a physiological or psychological response — say satisfaction — as a single criterion. In raising this issue one should add, perhaps, that although some of the previous authors have treated money as an objective measure of costs or revenues but satisfaction as a subjective or intangible one, this distinction is not entirely valid. Money has simply become a convenient measure by which effort, sacrifice, human values, labour, risk, danger and a whole host of other factors are measured. It is no more objective than desire, satisfaction or sensation. The fact that the price of copper piping for central heating increases overnight and thus alters cost planning for building is merely an outcome of some deep seated problems of human relationship, say in Rhodesia, concerning race segregation, the value of labour and the economic pressures of a country being internationally ostracised. This is no place to develop the history and theory of economic measures or of marginal utility.

It is however very necessary to point out to designers, builders and administrators in the building industry, that there is nothing more objective about money than about a number of other scales of value which could be used. For confirmation of this one has only to look at experiments in which people's evaluation of the worth of money itself is assessed[7] to see that the relationship is not only non-linear but varies from person to person and group to group and that it is not even static for any individual. Under conditions of risk or danger or poverty, an individual's judgement of the value of a certain increment of money will be quite different from his judgement in the absence of these conditions.

Using the model established in Section 3, it is possible to treat the costs of the first three parts of the system as one side of an equation and the value achieved in the fourth part as the other and to set out to make the value exceed the cost by the maximum possible amount — i.e., a maximum return on investment. Another approach, and that which will be followed here, is to consider the total costs of providing any solution as well as the total cost of not providing it, i.e., "failure" to provide, and by summation to establish the minimum cost, i.e., the optimum solution. A practical example of this technique is the well-known one of the decision as to the optimum thickness t_o of pipe insulation around a pipe of known diameter carrying a fluid at temperature T1 in an environment at temperature T2, the latter being lower than the former. If the cost of insulation per unit thickness is represented by Figure (5) (this is a non-linear relationship), and if the conductivity of the insulant is known together with the unit cost of heat generation, then Figure (5) will represent the two curves which describe the cost of provision and the cost of failure to provide (heat loss). The saddle shaped summary curve represents the total cost and yields the optimum thickness t_o at the point where its slope is equal to zero; provision of anything less or greater than this is uneconomic. Mathematical techniques for discovering o-slope points on curves, planes or hyperplanes (either minima or maxima) and for testing whether there are others lower or higher are available for 2, 3 and n-1 dimensional problems; their usefulness is however limited.

This example suffers from three defects in its applicability to most building design problems. The first is that it relies on a few well established functional relationships based on the laws of natural philosophy and economics, second that it deals with only a few variables, and third that it deals with static and not dynamic performance. Most building design problems are both more complex in nature in that they involve many interactive variables and systems, and are also less clear cut in that the functional relationships are not always known. Often, too, the performance of the system changes with time. Nevertheless, the principles of the technique can be applied to many areas of design decision-making and one of its lessons is that it is important to break a complex problem down into as many small "independent" problems as possible. There are few truly independent areas, for even in the case of the pipe the thickness of the insulation would determine the width of duct in which the pipe had to be placed, that in turn might influence the width of corridor in which the duct runs, and that in turn the layout of the rooms, the utilisation of the site and finally the form of the city! However, these links are sufficiently tenuous to allow one to proceed as if the original problem had been independent. Dynamic programming techniques can assist where the system performance varies with time.

In many instances whilst the effects of a decision in activity or organisational terms can easily be predicted, the cost or value of these effects is difficult to establish. This is a central problem, as has already been stated in the previous papers, in any full cost-benefit analysis, and an attempt is made in Section 5 to suggest both research and design techniques which, to some extent, might deal with this difficulty.

There are many cases where functional relationships between variables in the system to be optimised are all linear; all known; where there is a single criterion for optimisation, and where all constraints can be defined. The best known technique for finding the optimum solution in such cases is Linear Programming — a technique for sequential solution of the set of equations describing the criterion function; the functional relationships (expressed as functional constraints — i.e., conditions the solution must obey in order not to break laws of physics, for instance!) and regional constraints (constraints on resources, or freedom of action, arising from the nature of the problem). There is no room to develop this technique here[8]— sufficient to say that it has been successfully used for such problems as minimising the cost of the admixture of different dwellings on a site[9]; minimising circulation distances[10]; and maximising the production of cast concrete panels[11].

5 Evaluation of Effects on Human Responses.

It is often possible to establish a functional relationship between sets of environmental variables and sets of human response variables. These latter may be measured by the use of semantic scales, which attach magnitudes to the way people think and speak about environment or by means of direct observation of non-verbal behaviour. If the response variable is in the form of performance which is relevant to a functional or productive task, it may be a relatively simple matter to attach cost figures to the various levels of performance. For instance, if Figure (6) represents the relationship between illumination and visual performance on a standard task; Figure (7) between the standard task and a known industrial one; and Figure (8) between industrial performance and cost (in terms of time, error, waste etc.) then a similar curve in Figure (9) might reasonably represent the relationship between illumination and cost (of "failure"). Treating the illumination subsystem as independent, the minimum cost (optimum) level of illumination can be found as shown. Such techniques can be of use in considering noise and speech interference; thermal environment and production; corridor width and risk of accident.

In more typical cases the only measure of response that can be reliably obtained is an evaluative one: satisfaction, acceptability, friendliness, gloom, noise measured on ordinal or interval scales. Findings are generally expressed in one of two forms: either a functional relationship between a continuum of an environmental variable and a response variable as in Figure (10); or a frequency distribution of the response at a given point in the response scale. In the latter case it might be, say the mid-point between maximum satisfaction and maximum dissatisfaction on a scale; or it may be the distribution of one response where all are dichotomised into one of two alternatives — say "acceptable" or "not acceptable" as in Figure (11). If either mode is used, how can costs be assigned to various levels of (say) satisfaction or to the frequency distribution?

One way is to establish by reference to simulations or real life, the amount of money required to shift responses by a measurable amount or to turn a "not acceptable" judgement into an "acceptable" one. For instance, if the frequency distribution of "not acceptable" is turned into a cumulative frequency distribution and data on the cost of conversion to "acceptable" is available, then for a homogeneous group of people the sigmoid curve can be read directly as a cost curve. For instance, let us assume that Figure (12) represents the cumulative frequency of "not acceptable" for office workers' judgements of background sound pressure level. Assuming that it is known that an office worker will work in unacceptable noise conditions at a cost of £100 p.a. (say 10% of clerical salary) then the increase in costs with increased noise levels is as in Figure (13) where by the time 75 dBA is reached, all the workers require environmental compensation. Superimposed is the curve representing the cost of preventative measures (duct lining, double glazing etc.). The summation yields the optimum.

One problem which is fundamental to predictive design by cost is to find relationships between measures of hardware systems and their environmental performance, and the costs. For instance, some work has been done on the relationship between the sound attenuation of partitions and their costs. The data was scattered but sufficient for a clear correlation to emerge[12], and took the form shown in Figure (14). One now has a cost(of provision and maintenance) function for sound attenuation. If the behavioural effects of different levels of attenuation can be predicted, and the cost of these effects can be established realistically, then one is in a position to use the optimisation technique suggested by the 3 curves in Figure (14).

It may be objected that the first relationship is tenuous, and so dependent on a host of other factors — current technology, market pressures, materials supply — that it in no sense represents even a weak law. A technological breakthrough may, at any moment, completely nullify even such relationships as are found. This, coupled with the complexity of discovering behavioural effects and costing these, makes the case for predictive optimisation look pretty weak. The only alternative then, however, is to simulate a large number of solutions in every design, obtain accurate and up-to-date costs for each and also the latest data on environmental and human effects and, if possible, the costs of these latter. This, with the aid of computers and massive data structures would certainly be feasible. However, two reasons still make it seem worthwhile to pursue the more classical, predictive approach. The first, a short-term one, is that computing systems, data structures and personnel for the large and rapid simulation approach will be unavailable or in short supply for some years yet. The second, a long-term one, is that even then, the designer will need a <u>direction</u> in which to progress. He cannot generate solutions randomly, hoping for improved results; he must know why improvement is likely to result and hence have some predictive knowledge. Even more basically than this, unless he carries in his head at least some generally valid relationships of the kind illustrated in Figure (14) then in the multivariate and complex total system he is designing his initial direction will be totally random, and hence have a high probability of cost, performance or human failure which only a near-infinite set of simulations could cure. Relationships are merely formalisations of knowledge available to him and his experience and are discarded at his peril.

Often there is a choice between different combinations of desirable properties — two in the simplest case. Much economic theory and psychological experiment has been carried out into the nature of choices between such combinations. The value or 'utility' of the combination can be plotted in such a way that different combinations of equal utility are represented by an indifference curve. Figure (15A) shows a set of such curves (hypothetical) for different combinations of space heating standards and space in housing. Such a set is an indifference map. With three variables an indifference plane can be drawn; with n variables, represented in a matrix, a hyperplane with n-1 dimensions can be obtained. The curves are generally assumed to be convex in relation to the origin (according to the theory of diminishing marginal utility, now only accepted in the most general case). That is, a small increment of any good is seen as less valuable when added to a large amount than when the same increment is added on to a smaller amount.

If total resources — say for house purchase by an individual, or house construction by a housing authority or the nation — are limited, and can be divided in any way between combinations of two or more desirable properties, then all those combinations costing the same amount can be plotted on curves, planes or hyperplanes of equal cost, sometimes called the production function. Figure (15B) shows a set of equal cost curves (each for a house of different total cost) in which space heating standards and space are considered. These curves are concave in respect to the origin on the basis of the law of diminishing returns (again only true in a general sense) which is held to apply to resource utilisation. That is, if a large amount of something is made, to make an additional incremental amount of it costs less than if a small amount is made.

If these two sets of curves are combined on the same diagram, as on Figure (16), then we get a number of tangent points. Each of these represent an optimal decision; that is, if the satisfaction to be achieved is defined by curve S_2, then the least cost of production which will meet it is represented by point P_x on curve P_2. If the maximum resources available are represented by curve P_1, then the maximum value which can be achieved by them is represented by point P_y tangential to curve S_1. It will be noticed that this technique will not answer the question "what is the optimum amount to invest" — it merely enables the best result from a fixed investment or the least cost investment for a fixed result to be determined. But this is useful, as it represents the degree of freedom in many real situations.

An interesting adaptation of this technique comes from the theory of games. If there are conflicting parties to a decision, each of whom desires the maximisation of a different property, and if an increase in one can only be met by a decrease in the other, then conflict exists. A practical example is the competitive demand on skilled building labour (a fixed pool) for house building and all other types of building. Point P_1, on Figure (17) represents the allocation of labour to the two areas at time t_1. Production function curve F_1 represents the output of houses and other buildings which can be achieved by this pool of labour. As a result of political struggle, the house builders gain extra labour at the expense of the "other" builders, so that at time t_2, P_2 represents the new situation. It will be noticed that this lies on a higher production function curve, F_2, than P_1, that is some increased efficiency of labour has been achieved. As a result of a second conflict there is a reverse for the house builders, and P_3 is reached. Eventually P_5 is reached, on F_5, where limit of improved efficiency is reached. Both sides have made positive gains, in spite of temporary setbacks, as a result of positively resolved conflicts. But now, any shift must be along the locus F_5 — i.e., the final boundary of labour capacity. Thus P moves up and down this curve, gains for one party being losses for the other and no overall improvement for both taking place with time. Such situations result, usually, in open political conflict or in genuine technological innovation — i.e., the expenditure of energy in pushing the production curve to a new frontier, F_6, which enables conflicts to be resolved within a new area. Many examples of design, production or labour innovation are the result of this type of pressure.

It will have been noticed that of two or more goods or properties being compared to obtain indifference pairs, triplets or sets, one may be money itself. Thus, if pairs can be obtained in sequence (of independent goods and properties) until one can be costed (objectively or subjectively) then all can be costed. Where the goods or properties are interdependent, more complex ordering and ranking techniques are used to give, in the end, similar indifference sets.

Another technique for obtaining the relative values of benefits and disadvantages is to present subjects with simulated solutions in which variables are systematically varied, in combination with each other, and included is either money (e.g., salary) or some easily and objectively costable items. By expressing preferences, in pairs, or judging each solution on a scale, and subsequent variance analysis, the potency of each variable in determining the overall judgement can be obtained. This potency can be taken to be proportional to the percentage of the total variance accounted for by the particular variable; and if one is money, the others may be evaluated in relative terms. One such experiment presented panels of school teachers with verbal descriptions (simulations) of schools having different combinations of headmaster, children, efficiency of buildings, distance from home[13]. Figure (18) shows the pie chart resulting from the variance analysis. Teachers' salaries could easily be included in such an experiment and the cost of all or some variables can be established.

Often evaluations can be obtained from market or field observations. For instance the relative value of different aspects of housing environment may be obtainable from massive samples of sales or rentals, treated factorially and analysed for variance. Sometimes environment is negotiated for money directly — as in certain industrial occupations where the Unions have agreed to standard additional rates for environmental hazards such as heat or dust. On other occasions the sequence of choice — say of houses sold on an estate — will yield rank orders of priorities for various properties.

In business and military situations complex decisions sometimes have to be made in which the optimum involves choices partly determined by predictable outcomes (profit or loss of lives) and partly by subjective value judgements of the relative importance of various objectives and of the amount of risk which is justified in situations only partly predictable. Neumann and Morgenstern[14] formalised the theory of games and recently attempts have been made to apply some of these techniques to planning and design decisions[15]. In these participants may play specific roles, either conflicting or co-operating, and be asked to select from sets of solutions, containing different combinations of variables, the most-preferred. Often this is done sequentially, so that new choices can be made in the light of choices by the other participants, until a stable preference set is reached. Sometimes participants have a fixed resource to allocate between all the solutions (money or votes). If simple preferences and rankings are used, the value (or 'utility') of each solution can be measured by an ordinal utility index. If however it is necessary to compare the differences between various strengths of preferences, it is necessary to have a utility index which has measurable gaps between choices — i.e., a cardinal utility index. Newmann and Morgenstern showed how this could be obtained and the increase in the power of the technique thus resulting. The great difficulty in using such games for building design is the construction of simulations which are rich enough to represent realistically the differences between the solution sets, and yet quick enough for the game to proceed.

6 Constraints — Financial and Other

Constraints of any kind have two functions: first to limit the designer's area of search; and second to protect society against unscrupulous and ignorant designers. Constraints on resources simply express the limit on what is available. It is clear that often these constraints not only limit freedom of choice but sometimes force designers to adopt solutions which are far from optimal. A practical example will show this.

Let us assume that the design decision to be made concerns the area (A) of window to be used in a school classroom. The cost of provision and maintenance is shown by curve 1 in Figure (19) for a single glazed window of a certain proportion. This includes the variation in the heating plant and heating energy to maintain a fixed design temparature. It is assumed that the effects of noise are represented by the cost of "failure" curve, 2, which also increases monotonically with size. The effects of view, sunlight penetration and general visual character are shown by the monotonically decreasing curve 3. The summary curve 4 shows that the optimum (minimum cost) solution lies at size X_1. However, the 2% minimum daylight factor constraint limits the solution so that $A \geqslant X_2$. The requirement to limit background noise, from a busy street outside, places an upper constraint on $A \leqslant X_3$. Even the area between X_2 and X_3, well away from the optimum, cannot be searched as the cost limit of Y_1 determines $A \leqslant X_4$. Between X_2 and X_4, due to rationalisation of window dimensions, only one size exists — which is the solution! Such examples abound in daily practice; some are even more serious as the constraints are mutually exclusive and prevent a feasible solution from existing.

The need for constraints is obvious; there needs to be, however, careful thought about allowing as much interchangeability between resources as possible so that really economic solutions are not constrained out. Moreover, industrialised building and rationalisation of available components must ensure that the maximum amount of choice exists in the near-optimal areas.

Designers sometimes refer loosely to optimisation by several criteria. This is impossible — one criterion leads to one optimum, another to another. Criteria can be expressed in terms of one another, say by ranking and weighting them and expressing them as a single composite criterion. Alternatively the less important criteria can be turned into constraints around the key one. Since sub-optimisation of sub-systems can lead to a near "pessimum" overall solution unless the sub-system is independent of others, it is not only important to make sure that suboptima are only used in such situations (with their own criteria) but also that, as far as possible, constraints are applied to the least interdependent parts of the system.

The problem of optimising interactive sub-systems has been mentioned, and also that of optimising even a system with independent sub-systems each of which has a different criterion. However another problem which appears much more tractable is common and important. This is that of two or more independent sub-systems, with the same criterion, say cost, where the sum of the sub-optimal (minimum) solution exceeds a constraint on the whole system. (To the extent that an overall system constraint binds all the sub-systems, they are not truly independent.) A simple example will illustrate this.

Consider single storey shed-type factory construction. Three decisions are to be made:

(i) The span between columns
(ii) The amount of roof insulation
(iii) The amount of ceiling sound absorbent

In (i) the cost of the structure increases with bay span; the cost of "failure" in restricting production space decreases as shown in Figure 20; the total cost has a minimum value. The problem in (ii) is similar to the pipe insulation problem already dealt with; Figure 21 shows the cost of provision, the cost of "failure" (heat loss) and the total cost, with a minimum cost, optimum thickness, of insulation. In(iii) it is assumed that the lowering of noise levels resulting from increased absorbent, results in productivity and morale increases which have the costs indicated in the "failure" curve in Figure 22, which also shows the cost of provision of the absorbent and the total cost with its minimum value.

Assume now that if the three optimal values of costs of provision are added together, they exceed the overall cost limit for the factory and the designer is then left with having to choose which of the three decisions, if any, should be optimally decided, with the others shifted to lower, less-than-optimum, cost of provision points; or whether to manipulate all three, and each to what extent. In other words he wants to carry out a sensitivity study to see how he can depart from the three optima with least penalty.

Figure 23 shows a graphical solution based on a technique developed by Maver[16]. On it three curves representing the relationship between the incremental cost of provision and the incremental total cost are plotted with the three optimal solutions represented at the origin; for each of the three decision areas, as well as a fourth for the summated incremental costs of provision and summated total costs. From the origin, on the x axis, representing costs of provision, a sum equal to the difference between the true optimal total cost and the overall cost limit is plotted; this is point x_4. This is projected up onto curve 4 and then across onto curves 3, 2 and 1. From each of these three intersections a line is dropped onto the x axis giving points x_3, x_2 and x_1; these three points represent the amount by which each of the costs of provision of the three sub-systems should be underspent so as to depart least from the true optimum solution.

7 Cost, Responses, Values and Decisions

It may be useful to complete this paper by looking at one or two simple examples where essential information from various sources is brought together in a way for effective decisions.

Example 1. Optimum room size and shape[17].

Rectangular rooms with one window wall and a parallel corridor wall are to be designed, arranged in rows with party walls between them. This is a common built form in hotels, halls of residence, hostels and offices.

On Figure (24) equal area curves A_1-A_4 are drawn. All points of any of these represent x, y coordinate whose product is constant (i.e., they are hyperbolas). On the same Figure equal cost curves C_1-C_4 are drawn. Points along these represent rooms whose constructional, heating, maintenance and other costs are identical. These cross the A curves, on the basis that long narrow rooms with the window-wall on the long side cost more per unit area than those with the party wall on the long side on account of extra external wall costs, longer corridors, more extended service runs, higher heating, cleaning and maintenance costs etc. Also plotted on the Figure are curves PA_1-PA_4; these join points which yield room sizes which are perceived, under standard test conditions, to be of equal area. The hypothesis underlying their shape (but unverified) is that a shallow rectangular room, with large window area, appears to be larger than a deep room of the same shape and size with a window at one end. The slope of these curves is shallow, suggesting an effect stronger than that which determines the slope of the cost curves.

This set of data can be used in several ways. If the goal is to provide a specified subjective (perceived) or objective area, then the lowest cost solution can be found. If the cost is given, by a cost limit, then the maximum subjective or objective room area which can be built can be found. If on the graph one plots for each size and shape of room the number of furniture layouts it permits, and then draws curves to connect equal numbers, one can then find, for a given investment, objective size, or perceived size, the room size or shape giving the maximum amount of freedom (or "multi-modality") for the present or the future.

Example 2. Optimum window size[18].

Another example concerns the selection of the optimum size window for an office with four occupants. Figure (25) shows the provision and maintenance cost of a range of window sizes (x axis) with the weighted cumulative frequency responses of "not acceptable" for each size of window (y axis). Window shape is assumed constant and the subjective judgement is a composite of judgements relating to view, sunlight penetration and noise.

Any linear function drawn between the two axes represents an indifference line between subjective judgement and money. Thus, if it is assumed that the annual cost of providing "not acceptable" windows for the four office occupants (earning £1,000p.a. each) might be equivalent to the 1% of their combined salaries, then the line $x_1\ y_1$ represents combinations of dissatisfaction and cost where overall cost is constant. Hence any two windows falling on this line have equal total cost. If such an indifference line is tangential to the window curve at any point then, provided the evaluation of "not acceptable" is adhered to, the size of window at the tangent point is the least-cost solution. By altering the slope of $x_1\ y_1$ other evaluations of subjective judgements can be explored and the sensitivity of the curve to such iterative alterations can be explored.

8 Different Routes to "best" Solution

It will be clear from all that has been said in this and the previous articles that optimising the present and future state of a system as complex as a building, with all its services, environmental conditions and activities is no simple task. Obviously it can be achieved by a host of different techniques and usually has to be achieved by a mixture of them. In Figure (26) the starting point is the designer in search of the "best" solution which lies at the other end of the diagram. It would appear that he can follow one or any combination of 3 basic routes[19].

The first route is the predictive one. Here the designer uses all existing knowledge based on past research and experience, his own personal experience and the results of any specially commissioned research, to enable him to predict what combination of characteristics in his solution will best meet his goals. This kind of route is the appropriate one where the problem has no radically new features and where a lot of experience and research exists.

The second route is the simulation route or, putting it in a more homely manner, the "suck it and see" approach. If the predictive knowledge is inadequate or out of date and there is no time to mount special research, there will be two cogent reasons for making models or simulations of a wide range of solutions and, by iterative adaptations, arriving at a near best solution. The first reason is that some of the physical or cost data is too scattered for any predictable relationships to be feasible. Examples are complex heat flow, intricate lighting solutions or unusual circulation patterns. In all these cases a simulation by an analogue or by detailed computational means may enable the designer to quickly represent a sufficiently wide range of relevant solutions to come out not only with a feasible one but a good one if not the best (since by definition the best in this case is unknown). Many rules of search are available for this kind of exercise.

The second cogent reason for simulation is that in evaluation of the human effects and benefits of solutions there is often no short cut to actually presenting simulated solutions to relevant and representative people and measuring the effects upon them and also obtaining their evaluation of these effects. Measurement may be, for instance, by means of semantic scales on which numerical scores are obtained, and evaluation may be by assigning costs which people are willing to incur to avoid undesirable effects or to obtain desirable ones. Such judgements, whether carried out informally or in the more formal milieu of games, will involve increasing participation in the design process by all those who are likely to be affected by the outcome of decisions.

The third route is the most hazardous and is only taken if the predictive knowledge is unobtainable and if there is an insuperable difficulty about constructing an effective simulation. This route might be called the "multi-modal" route. It involves the attempt to recognise those dimensions of the solution which are likely to be most critical but whose desired properties are most difficult to predict. From this follows the design of a solution which is capable of use in many modes, i.e., it is not <u>a</u> solution at all but potentially a large number of different solutions. An infinitely adaptable solution will cost an infinite amount but is bound to embody within it the best solution. Even a small degree of adaptability, e.g., of spaces and services, may cost a large amount and therefore the decision to build it in is a risky one. Equally risky, however, is to pretend that one knows the answer from knowledge or simulation when one does not, because the chances of a unique solution, within a set of solutions with random probability of each being appropriate, being the best solution is very small indeed and the cost of failure may be high.

These 3 routes represented in Figure(26) from left to right, are the paths followed, most probably in combination, in almost all design work. Each has a corresponding feedback loop. For the first it is necessary to study the finished object to see whether it does in fact obey the predicted behaviour and if not to alter the body of theory which was used in the light of new experiments and experience. For the second route the feedback loop concerns tests on the validity of the simulations used by observing the behaviour of the real object and comparing it with the behaviour of the simulation. Not least on such validity tests are those designed to answer the question of people's responses to simulations compared to their responses to the real world. The resulting knowledge is added to the pool of simulation theory which will be used in the future. For the third route the feedback loop is that which continuously monitors the behaviour of multi-modal systems (and all real systems are multi-modal to some extent whether designed to be or not) and from this monitoring isolates the dimensions and the degree of adjustment and adaptation which people make to the system over its life. This knowledge is fed back into the pool of knowledge about the behaviour of multi-modal systems and is used in the future. These three feedback loops are properly speaking research activities and make clear that design decision-making and research are complimentary parts of a complete cybernetic cycle and without the support of the corresponding half each is lost. But that is the beginning of another story and serves as a suitable finishing point for this one.

*This article, in slightly amended form, originally formed part of a series in 'Building' and was published August 21st and September 18th, 1970.

References

1 Markus, T.A. 'The role of building performance measurement and appraisal in design method' Architects' Journal, 20.21.67, 1567-73. The paper summarises those models of design method which include the two basic patterns: Levin, Asimow, Pask, Archer, RIBA, Broadbent.

2 Royal Institute of British Architects 'Handbook of architectural practice and management', London, 1967, The Institute.

3 Markus, T.A., op.cit.

4 Based on a model by the Building Performance Research Unit, University of Strathclyde, and first published in Markus, op.cit.

5 Fleming, M.C. 'Building decisions and economic appraisal techniques in practice' Building, 17.4.70, No.16, 73-78.

6 Whitton, D.J.R. 'Performance specifications as contract documents for building work' MSc Thesis submitted to University of Strathclyde, 1970.

7 See for instance Mosteller, F. and Nogee, P. 'An experimental measurement of utility'(1951) and Davidson, D., Suppes, P. and Siegel, S. (1952)'Decision-making; an experimental approach', both in Edwards, E. and Tversky, A. (Ed.) Decision-making selected readings, Penguin Books, Middlesex, 1967.

8 For a general introduction see Churchman, C.W., Ackoff, R.L. and Arnoff, E.L. 'Introduction to operations research' Wiley, New York, 1957, especially Chapters 11, 12 and 13.

9 Krejcirik, M. and Spiler, V. 'Use of computers for determining the optimum development pattern for a residential area' Building Research Station Library Communication No.1235, May 1965, BRS, Watford.

10 Moseley, L. 'A rational design theory for planning buildings based on the analysis and solution of circulation problems' Architects' Journal, 11.9.1963, 525-37.

11 Sebestyen, G. 'Experimental computations connected with the allocation and dimensions of plants producing lightweight concrete medium wall-blocks for dwellings', Institute for Building Economy and Organisation, Budapest, Report No. 116/k-1.

12 Building Performance Research Unit, University of Strathclyde. 'The relationship between one performance criterion and the cost of some boundaries separating school teaching spaces', Report No. GD/27/DW/MM, November 1968.

13 Building Performance Research Unit, University of Strathclyde. Report on Exploratory Studies and Future Plans, Psychological Investigations, January 1968, Report No. GD/11/DC/MW.

14 Neumann, J. and Morgenstern, O. 'Theory of games and economic behaviour' Wiley, New York, 1964.

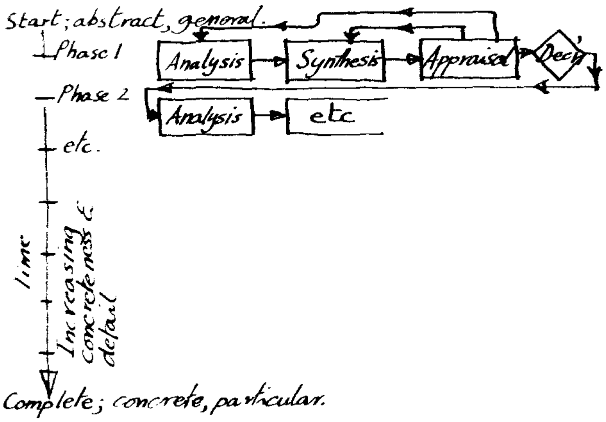
15 Duke, R.D. 'Gaming simulations in urban research' Michigan State University, 1964. Taylor, J.L. and Carter, K.R. 'Instructional simulation of urban development — a preliminary report' Journal of Town Planning Institute, 53, 10, December 1967.

16 Maver, T.W. 'Systems approach to engineering services' Chapter 4, Draft text for RIBA Textbooks series, 1970.

17 I am indebted to my colleague K. Taylor who first suggested to me the representation of equal room areas by the hyperbolas.

18 Based on Markus, T.A. 'The real cost of a window — an exercise in cost benefit analysis in building design' University of Strathclyde, Report No.62, Feb. [illegible].

19 Simplified from model in Markus, T.A. 'Design and Research — Cooperation not Conflict' Conrad, July 1969, 1, 2, 35-38.



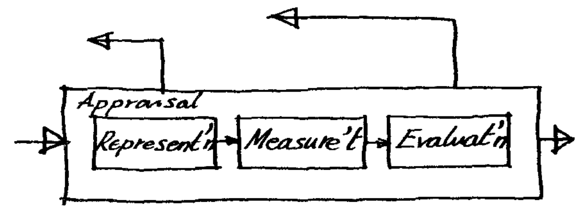


FIGURE 2.
The structure of appraisal

FIGURE 1.
The general structure of design

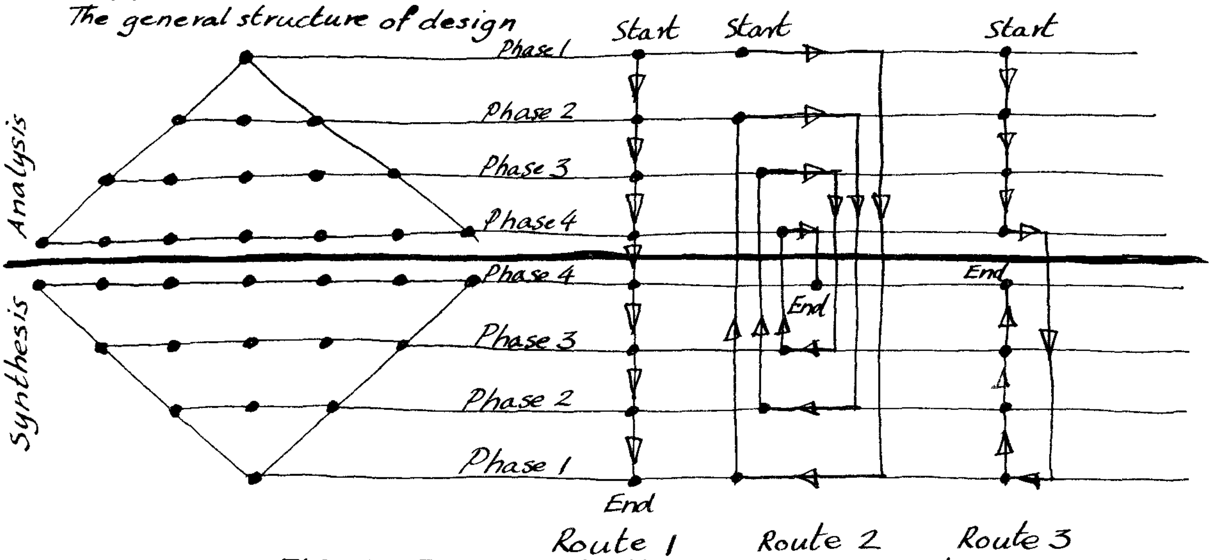


FIGURE 3. The double pyramid paradox

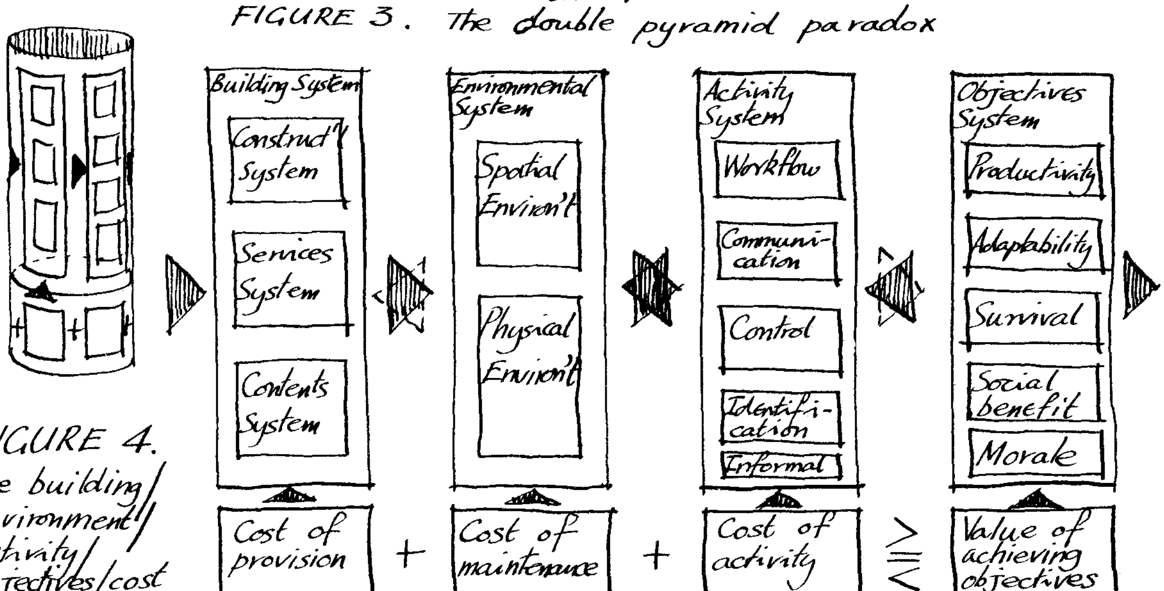


FIGURE 4.
The building/environment/activity/objectives/cost system

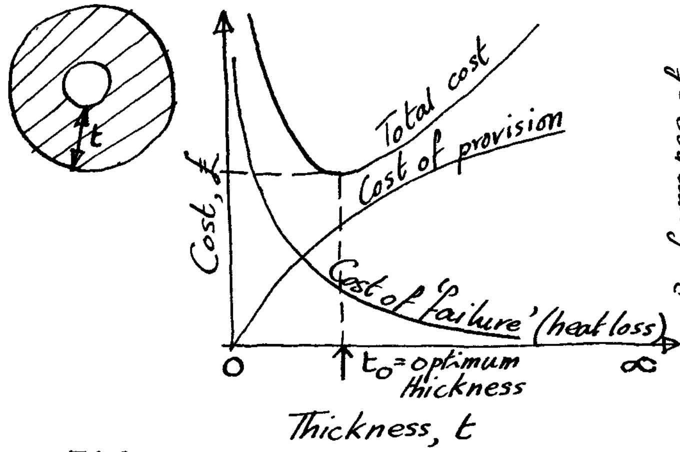


FIGURE 5.
Optimum thickness of pipe insulation

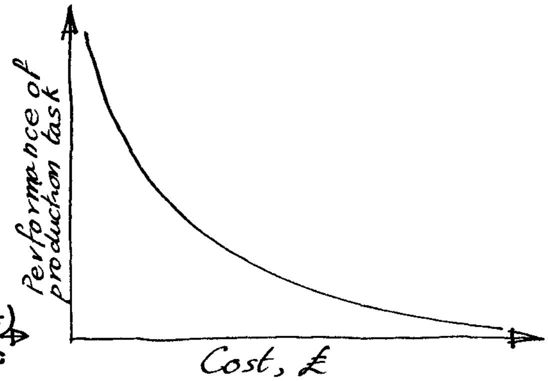


FIGURE 8.
Production performance and cost

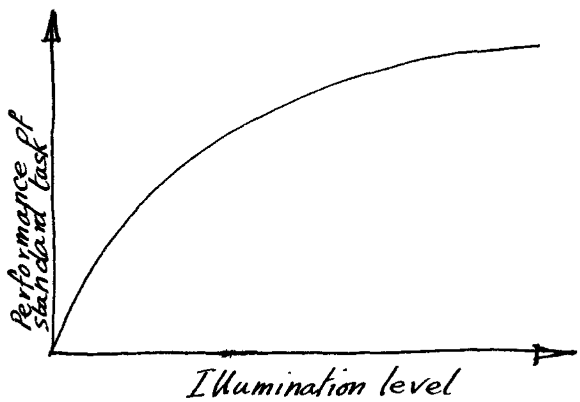


FIGURE 6.
Performance and illumination

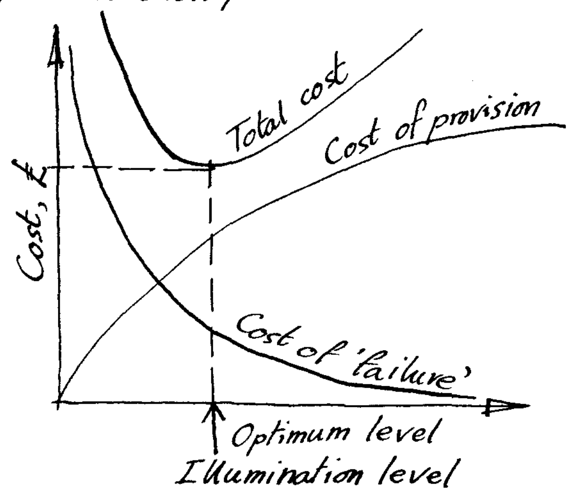


FIGURE 9.
Optimum illumination level

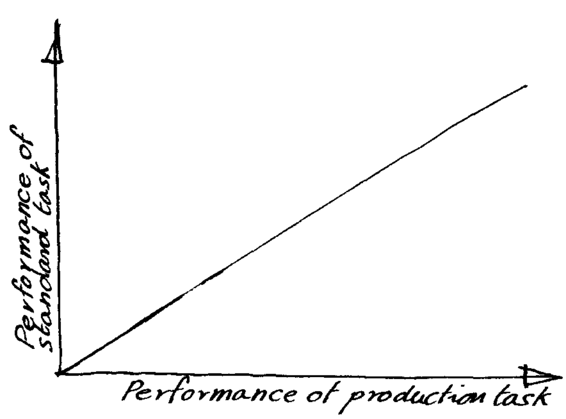


FIGURE 7.
Performance of standard and production tasks

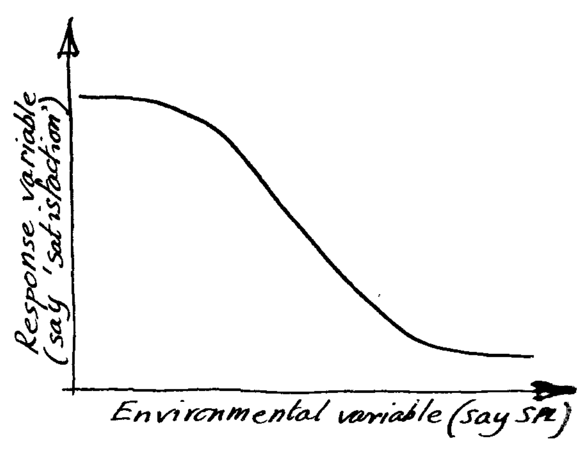


FIGURE 10.
Relationship between environment and response

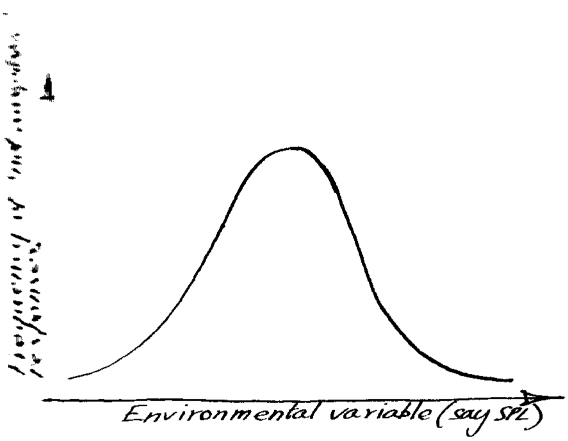


FIGURE 11.
Frequency distribution of responses with environmental variable

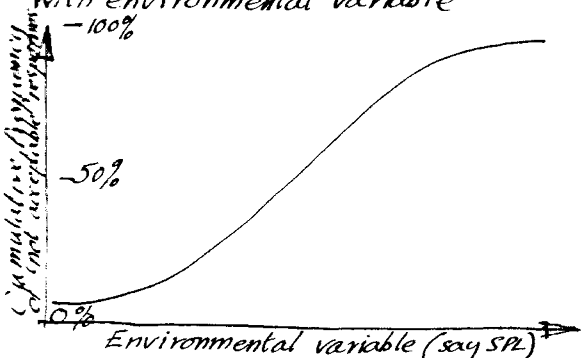


FIGURE 12.
Cumulative frequency responses with environmental variable

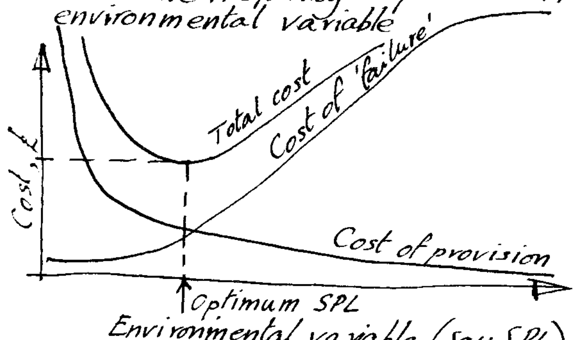


FIGURE 13.
Optimum SPL in an office

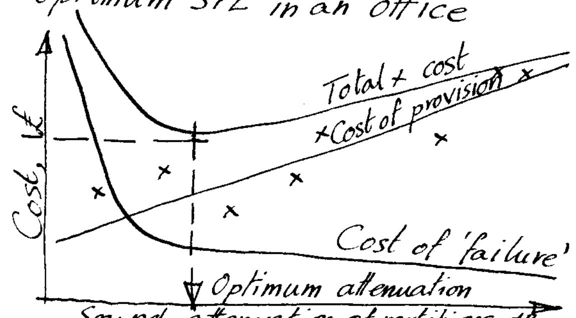


FIGURE 14. Optimum partition attenuation

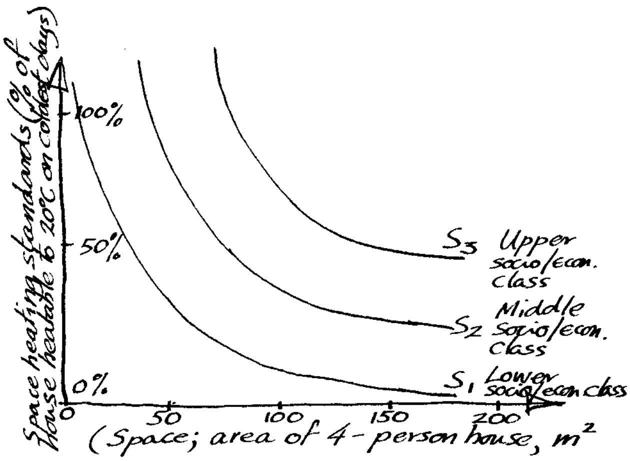


FIGURE 15A.
Indifference map (3 curves) for space heating and space for 3 socio/economic classes

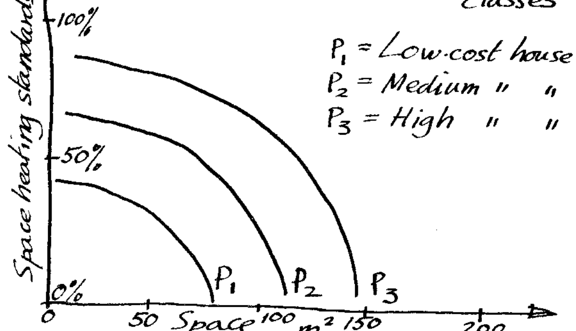


FIGURE 15B.
Production functions for 3 house types

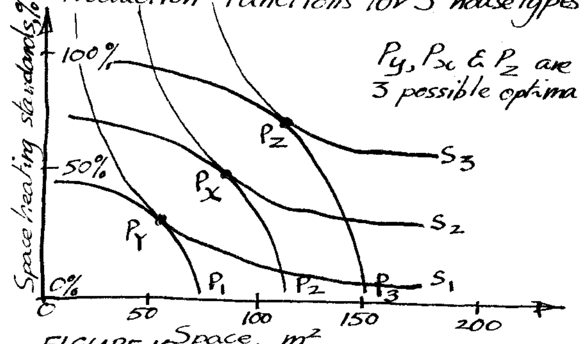


FIGURE 16.
Optimising space heating and space

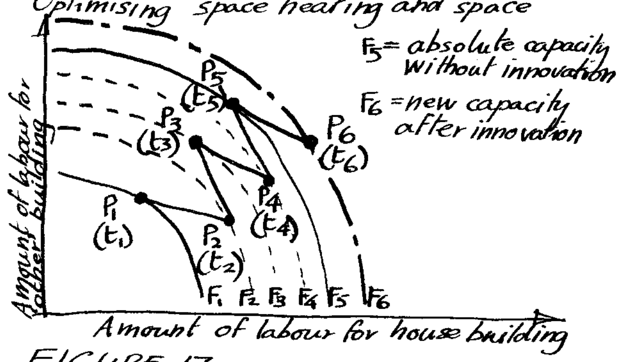


FIGURE 17.
Conflict, progressive improvement & innov'n.

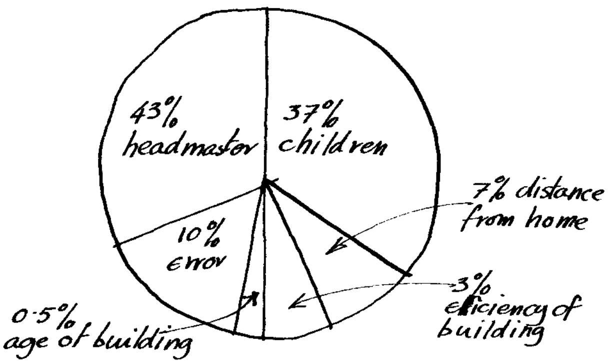


FIGURE 18.
Analysis of variance for teachers' responses

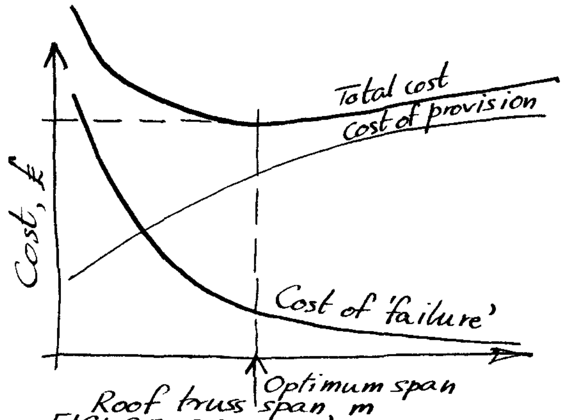


FIGURE 20.
Optimisation of factory roof-truss span

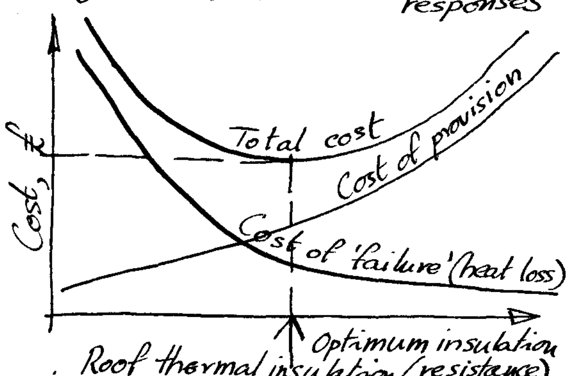


FIGURE 21.
Optimisation of factory roof insulation

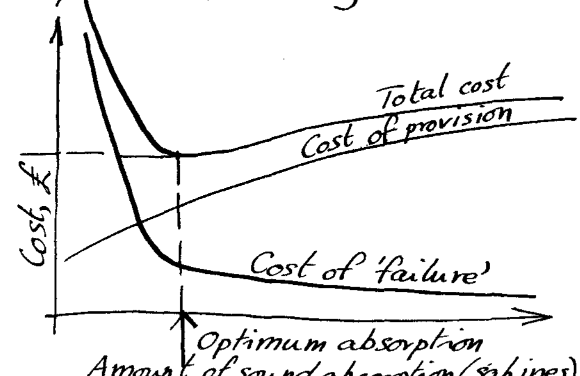


FIGURE 22
Optimisation of factory sound absorption

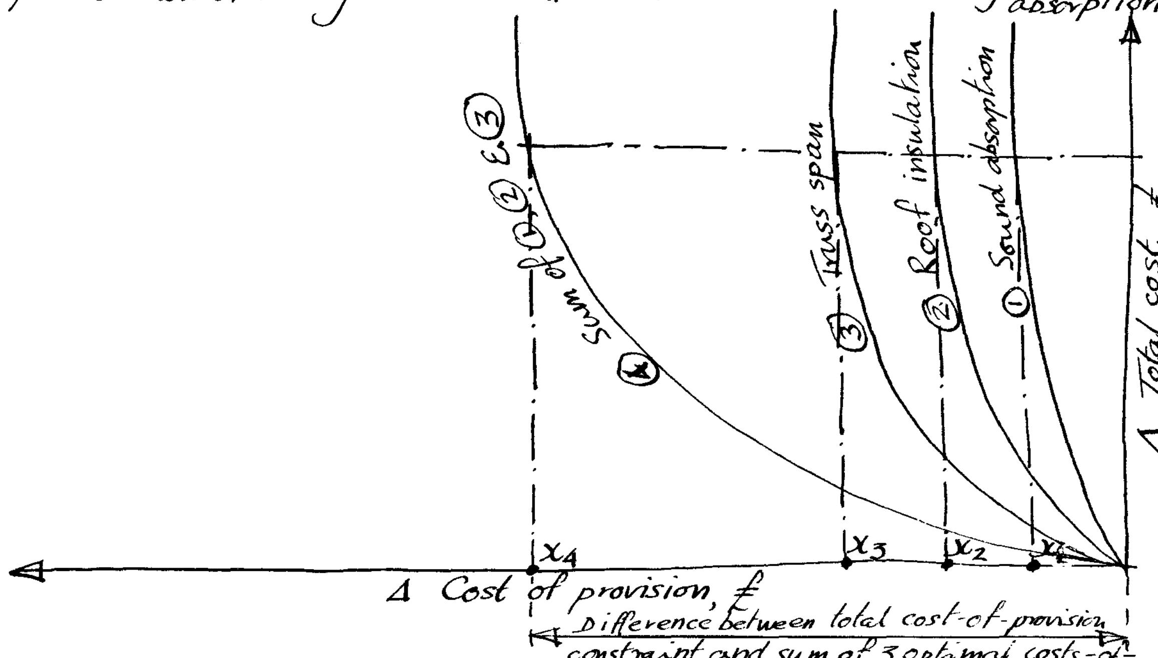


FIGURE 23.
Optimisation of sub-systems when overall cost-of-provision constraint prevents achievement of true optima for sub-systems.

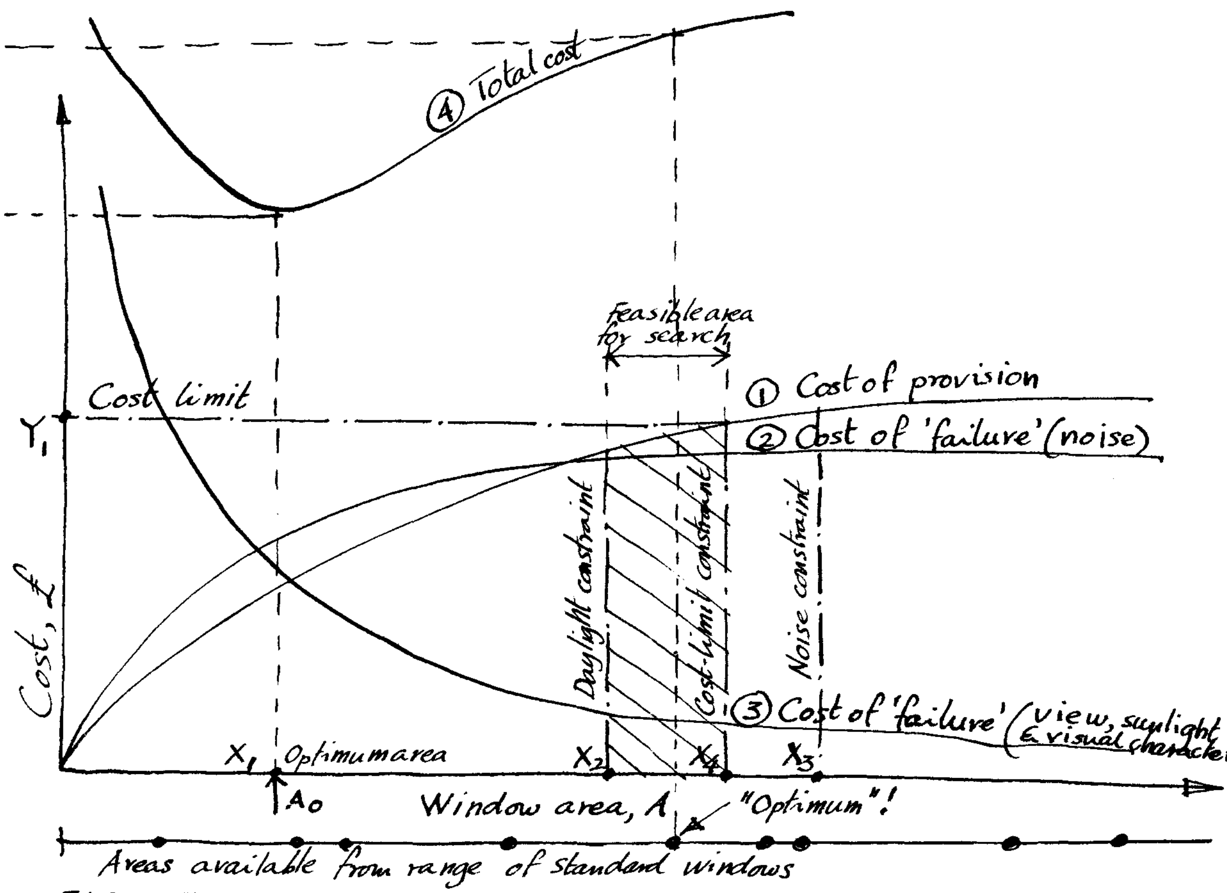


FIGURE 19.
Effects of functional, cost & production constraints on the design of a window

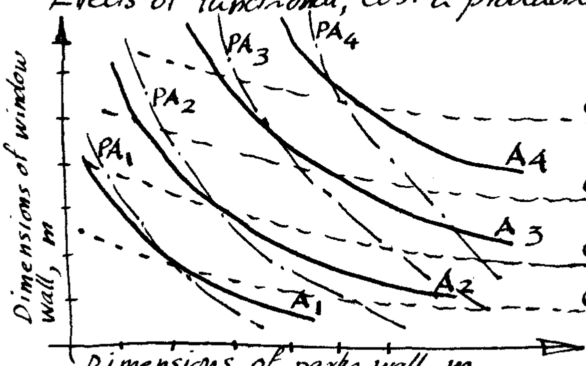


FIGURE 24. Room shape and cost

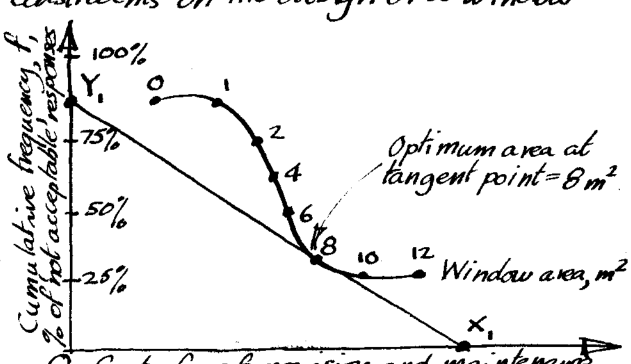


FIGURE 25. Optimum window size

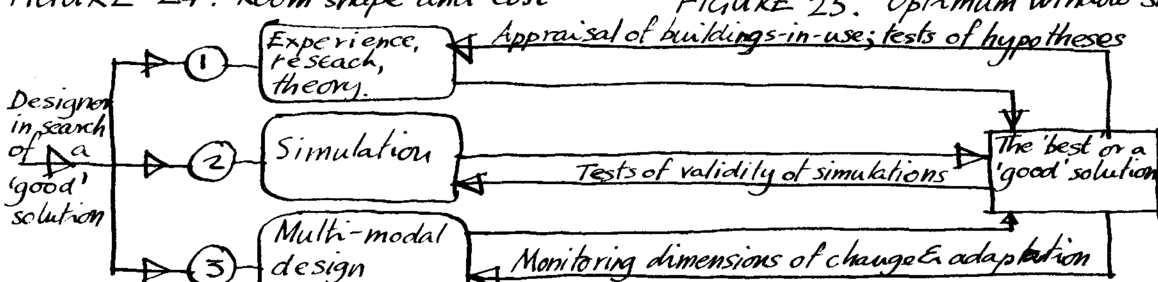


FIGURE 26. Three routes to the 'best' or a 'good' solution