2 Fundamentals

To develop an approach to design that can serve as a strategy for the development of solutions, we must first examine the fundamentals of technical systems and procedures along with the prerequisites for computer support. Only when that has been done is it possible to make detailed recommendations for design work.

2.1 Fundamentals of Technical Systems

2.1.1 Systems, Plant, Equipment, Machines, Assemblies and Components

Technical tasks are performed with the help of *technical artefacts* that include plant, equipment, machines, assemblies and components, listed here in approximate order of their complexity. These terms may not have identical uses in different fields. Thus, a piece of equipment (reactor, evaporator) is sometimes considered to be more complex than a plant, and artefacts described as "plant" in certain fields may be described as "machines" in others. A machine consists of assemblies and components. Control equipment is used in plant and machines alike and may also be made up of assemblies and components, and perhaps even of small machines. The variation in use of these terms reflect historical developments and application areas. There are attempts to define standards in which energy-transforming technical artefacts are referred to as machines, material-transforming artefacts as apparatus and signal-transforming artefacts as devices. It is evident that a clear division on the basis of these characteristics is not always possible and that the current terminology is not ideal.

There is much to be said for Hubka's suggestion [2.22–2.24] that technical artefacts should be treated as *systems* connected to the environment by means of *inputs* and *outputs*. A system can be divided into subsystems. What belongs to a particular system is determined by the *system boundary*. The inputs and outputs cross the system boundary (see Section 1.2.3). With this approach, it is possible to define appropriate systems at every stage of abstraction, analysis or classification. As a rule, such systems are parts of larger, superior systems.

A concrete example is the combined coupling shown in Figure 2.1. It can be considered as a system "coupling" which, within a machine, or when joining two

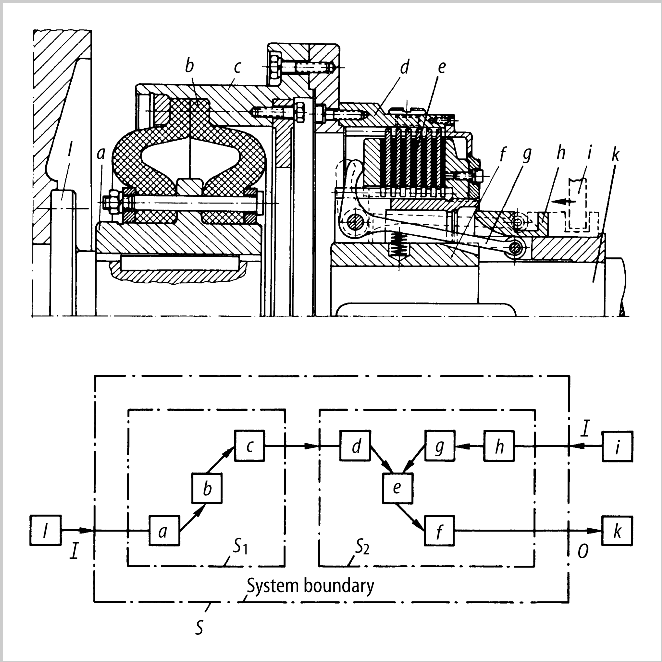


Figure 2.1. System "coupling": $a \ldots h$ system elements; $i \ldots l$ connecting elements, S overall system; S_1 subsystem "flexible coupling"; S_2 subsystem "clutch"; I inputs; O outputs

machines, can be considered to be an assembly. This coupling assembly can be treated as two *subsystems*—a "flexible coupling" and a "clutch". Each subsystem can, in turn, be subdivided into system elements, in this case components.

The system depicted in Figure 2.1 is based on its mechanical construction, referred to as the *construction structure*, see Figure 2.13. It is, however, equally possible to consider it in terms of its functions (see Section 2.1.3). In that case, the total system "coupling" can be split up into the subsystems "damping" and "clutching"; the second subsystem into the further subsystems "changing clutch operating force into normal force" and "transferring torque".

For example, the system element g could be treated as a subsystem whose *function* is to convert the actuating force into a larger normal force acting on the friction surface, and through its flexibility provide some equalisation of the wear.

Which viewpoint is used to divide the system depends on the intended purpose of the division. Common viewpoints are:

- Function: used to identify or describe the functional relationships
- Assembly: used to plan assembly operations
- Production: used to facilitate production and production planning.

Depending on their use, any number of such subdivisions may be made. Designers have to establish particular systems for particular purposes, and must specify their various inputs and outputs and fix their boundaries. In doing this, they may use what terminology they prefer or is customary in their particular field.

2.1.2 Conversion of Energy, Material and Signals

One encounters matter in many shapes and forms. Its natural form, or the form imposed upon it, provides information about its possible uses. Matter without form is inconceivable—form is a primary source of information about the state of matter. With the development of physics, the concept of force became essential. Force was conceived as being the means by which the motion of matter was changed. Ultimately this process was explained in terms of energy. The theory of relativity postulated the equivalence of energy and matter. Weizsacker [2.61] lists energy, matter and information as basic concepts. If change or flow is involved, time must be introduced as a fundamental quantity. Only by reference to time does the physical event in question become comprehensible, and can the interplay of energy, matter and information be adequately described.

In the technical sphere the previous terminology is usually linked to concrete physical or technical representations. *Energy* is often specified by its manifest form. We speak of, say, mechanical, electrical or optical energy. For matter, it is usual to substitute *material* with such properties as weight, colour, condition, etc. The general concept of information is generally given more concrete expression by means of the term *signal*—that is, the physical form in which the information is conveyed. Information exchanged between people is often called a message [2.20].

The analysis of technical systems—plant, equipment, machine, device, assembly or component—makes it clear that all of them involve technical processes in which energy, material and signals are channelled and converted. Such conversions of energy, material and signals have been analysed by Rodenacker [2.46].

Energy can be converted in a variety of ways. An electric motor converts electrical into mechanical and thermal energy, a combustion engine converts chemical into mechanical and thermal energy, a nuclear power station converts nuclear into thermal energy, and so on.

Materials too can be converted in a variety of ways. They can be mixed, separated, dyed, coated, packed, transported, reshaped and have their state changed. Raw materials are turned into part-finished and finished products. Mechanical parts are given particular shapes and surface finishes and some are destroyed for testing purposes.

Every plant must process information in the form of *signals*. Signals are received, prepared, compared and combined with others, transmitted, displayed, recorded, and so on.

In technical processes, one type of conversion (of energy, material or signals) may prevail over the others, depending on the problem or the type of solution. It is useful to consider these conversions as flows, and the prevailing one as the *main flow*. It is usually accompanied by a second type of flow, and quite frequently all three come into play. There can, for example, be no flow of material or signals without an accompanying flow of energy, however small. The provision and conversion of energy in such cases may not dominate, but it remains necessary to allow for them. Energy flow also involves the transfer of forces, torques, currents, etc., which are then referred to as force flow, torque flow and current flow.

The conversion of energy to produce electrical power, for example, is associated with a material conversion, even though no continuous material flow is visible in a nuclear power station compared to a coal-fired one. The associated flow of signals constitutes an important subsidiary flow for the control and regulation of the entire process.

However, numerous measuring instruments receive, transform and display signals without any flow of material. In many cases energy has to be specially provided for this purpose; in other cases latent energy can be drawn upon directly. Every flow of signals is associated with a flow of energy, though not necessarily with a flow of material.

In what follows, we shall be dealing with:

- Energy: mechanical, thermal, electrical, chemical, optical, nuclear ..., also force, current, heat ...
- Material: gas, liquid, solid, dust ..., also raw material, test sample, workpiece ..., end-product, component ...
- Signals: magnitude, display, control impulse, data, information ...

In this book technical systems whose main flow is energy-based are referred to as machines, those whose main flow is material-based as apparatus, and those whose main flow is signal-based as devices, unless these terms are not in line with established terminology.

In every type of proposed conversion, *quantity* and *quality* must be taken into consideration if rigorous criteria for the definition of the task, for the choice of solutions and for evaluation are to be established. No statement is fully defined unless its quantitative as well as its qualitative aspects are taken into account. Thus, the statement "100 kg/s of steam at 80 bar and 500 °C" is not a sufficient definition of the input of a steam turbine unless there is the further specification that these figures refer to a nominal quantity of steam and not, for instance, to the maximum flow capacity of the turbine, and the admissible fluctuations in the state of the steam are fixed at, say, 80 bar ± 5 bar and 500 °C ± 10 °C, that is, extended by a qualitative aspect.

In many applications, it is also essential to stipulate the *cost* or value of the inputs and the maximum permissible cost of the outputs (see [2.46], Categories: Quantity–Quality–Cost).

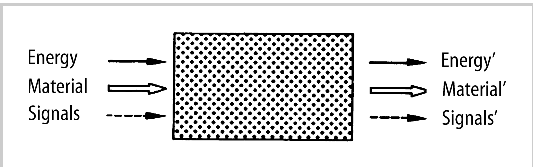


Figure 2.2. The conversion of energy, material and signals. Solution not yet known; task or function described on the basis of inputs and outputs

All technical systems, therefore, involve the conversion of energy, material and signals, which must be defined in quantitative, qualitative and economic terms (see Figure 2.2).

2.1.3 Functional Interrelationship

1. Task-Specific Description

In order to solve a technical problem, we need a system with a clear and easily reproduced relationship between inputs and outputs. In the case of material conversions, for instance, we require identical outputs for identical inputs. Also, between the beginning and the end of a process, for instance filling a tank, there must be a clear and reproducible relationship. Such relationships must always be planned—that is, designed to meet a specification. For the purpose of describing and solving design problems, it is useful to apply the term *function* to the intended input/output relationship of a system whose purpose is to perform a task.

For static processes it is enough to determine the inputs and outputs; for processes that change with time (dynamic processes), the task must be defined further by a description of the initial and final magnitudes. At this stage there is no need to stipulate what solution will satisfy this kind of function. The function thus becomes an abstract formulation of the task, independent of any particular solution. If the overall task has been adequately defined—that is, if the inputs and outputs of all the quantities involved and their actual or required properties are known—then it is possible to specify the *overall function*.

An overall function can often be divided directly into identifiable *subfunctions* corresponding to subtasks. The relationship between subfunctions and the overall function is very often governed by certain constraints, inasmuch as some subfunctions have to be satisfied before others.

On the other hand, it is usually possible to link subfunctions in various ways and hence to create variants. In all such cases, the links must be compatible.

The meaningful and compatible combination of subfunctions into an overall function produces a so-called *function structure*, which may be varied to satisfy the overall function. To that end it is useful to make a block diagram in which the processes and subsystems inside a given block (black box) are initially ignored, as shown in Figure 2.3 (see also Figure 2.2). The symbols used to represent subfunctions in a function structure are summarised in Figure 2.4.

Functions are usually defined by statements consisting of a verb and a noun, for example "increase pressure", "transfer torque" and "reduce speed". They are derived for each task from the conversions of energy, material and signals discussed in Section 2.1.2. So far as is possible, all of these data should be accompanied by specifications of the physical quantities. In most mechanical engineering applications, a combination of all three types of conversion is usually involved, with the conversion either of material or of energy influencing the function structure decisively. An analysis of all the functions involved is always useful (see also [2.59]).

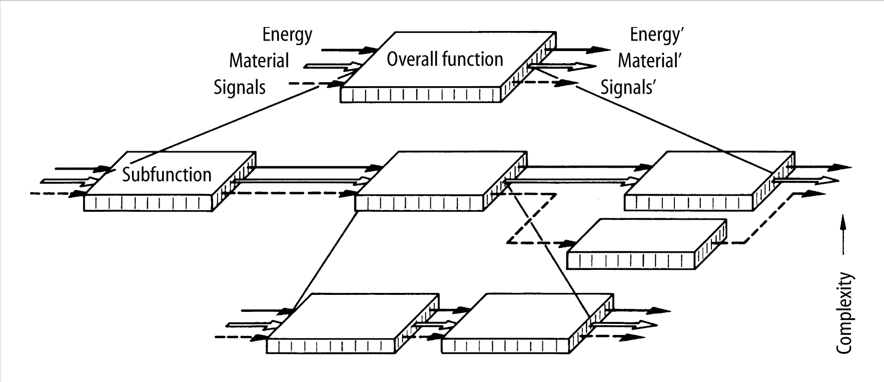


Figure 2.3. Establishing a function structure by breaking down an overall function into subfunctions

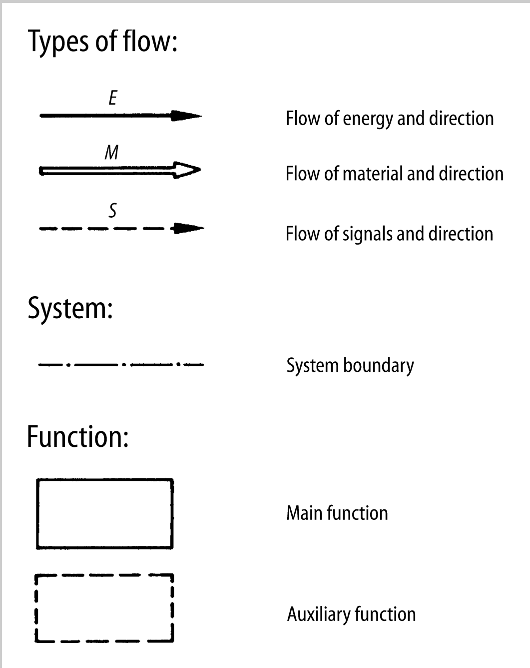


Figure 2.4. Symbols used to represent subfunctions in a function structure

It is useful to distinguish between main and auxiliary functions. While *main functions* are those subfunctions that serve the overall function directly, *auxiliary functions* are those that contribute to it indirectly. They have a supportive or complementary character and are often determined by the nature of the solutions for the main functions. These definitions are derived from Value Analysis [2.7,2.58, 2.60]. Although it may not always be possible to make a clear distinction between main and auxiliary functions, the terms are nevertheless useful. The division between them should be managed in a flexible manner. For example, a change in the system boundary resulting from a change of focus can transform an auxiliary function into a main function and vice versa.

It is also necessary to examine the relationship between the various subfunctions, and to pay particular attention to their logical sequence or required arrangement.

As an example, consider the packing of carpet tiles stamped out of a length of carpet. The first task is to introduce a method of control so that perfect tiles can be selected, counted and packed in specified lots. The main flow here is that of material, as shown in the form of a block diagram in Figure 2.5, which, in this case, is the only possible sequence. On closer examination we discover that this chain of subfunctions requires the introduction of auxiliary functions because:

- the stamping-out process creates offcuts that must be removed
- rejects must be removed separately and reprocessed
- packing material must be brought in.

The result is the function structure shown in Figure 2.6. It will be seen that the subfunction "count tiles" can also give the signal to pack the tiles into lots of a specified size, so it seems useful to introduce a signal flow with the subfunction "send signal to combine n tiles into one lot" into the function structure. The functions in this case are *task-specific functions*, whose definitions are derived from the terminology appropriate for the task being considered.

Outside the design domain, the term function is sometimes used in a broader sense, and sometimes in a narrower sense, depending on the context.

Brockhaus [2.40] has defined functions in general as activities, effects, goals and constraints. In mathematics, a function is the association of a magnitude y with a magnitude x such that a unique value (single-valued function) or more than one value (multi-valued function) of y is assigned for every value of x. According to the value analysis definition given in [2.7], functions define the behaviour of artefacts (tasks, activities, characteristics).

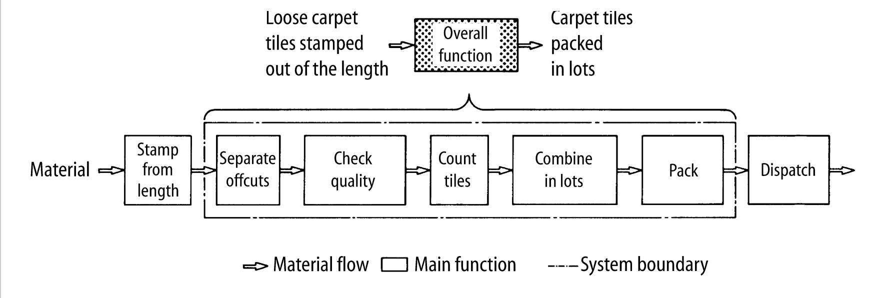


Figure 2.5. Function structure for the packing of carpet tiles

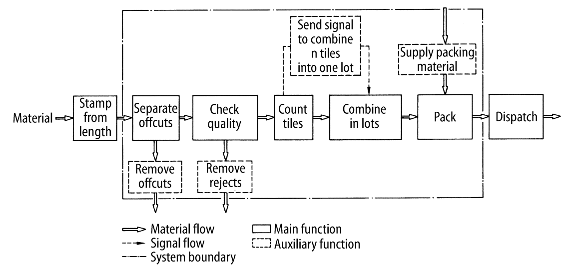


Figure 2.6. Function structure for the packing of carpet tiles as shown in Figure 2.5 but with auxiliary functions added

2. Generally Valid Description

Various design methodologists (see Section 1.2.3) have put forward wider or stricter definitions of *generally valid functions*. In theory, it is possible to classify functions so that the lowest level of the function structure consists exclusively of functions that cannot be subdivided further while remaining generally applicable. They therefore represent a high level of abstraction.

Rodenacker [2.46] has defined generally valid functions in terms of binary logic, Roth [2.47, 2.49] in terms of their general applicability, and Koller [2.28, 2.29] in terms of the required physical effects. Krumhauer [2.31] has examined general functions in the light of possible computer applications during the conceptual design phase, paying special attention to the relationship between inputs and outputs after changes in type, magnitude, number, place and time. By and large, he arrives at the same functions as Roth, except that by "change" he refers exclusively to changes in the type of input and output, while by "increase or decrease" he refers exclusively to changes in magnitude.

In the context of the design methodology presented here, the generally valid functions of Krumhauer will be used (see Figure 2.7).

The function chain shown in Figure 2.5 can be represented using generally valid functions, as shown in Figure 2.8.

A comparison between the functional representations in Figures 2.5 and 2.8 shows that the description that uses generally valid functions has a higher level of abstraction. For this reason, it leaves open all possible solutions and makes a systematic approach easier. However, using generally valid functions can represent a problem because such an abstract level can sometimes hinder the direct search for solutions. For more about the application of task-specific and generally valid functions, along with further examples, see Section 6.3.

3. Logical Description

The logical analysis of functional relationships starts with the search for the essential ones that must necessarily appear in a system if the overall problem is to

Characteristic Input (I)/Output (O)	Generally valid functions	Symbols	Explanations
Type	Change		Type and outward form of I and O differ
Magnitude	Vary		I < O I > O
Number	Connect		Number of I > O Number of I < O
Place	Channel		Place of I ≠ O Place of I = O
Time	Store		Time of I ≠ O

Figure 2.7. Generally valid functions derived from the characteristics type, magnitude, number, place and time for the conversion of energy, materials and signals

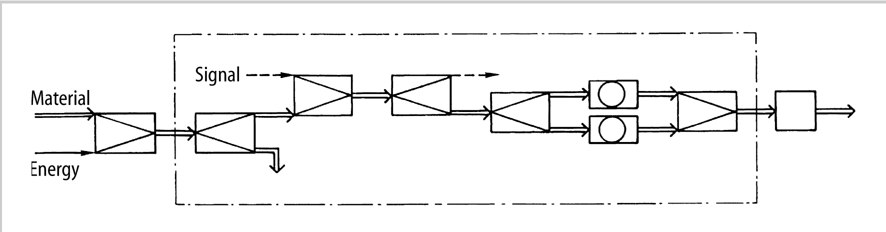


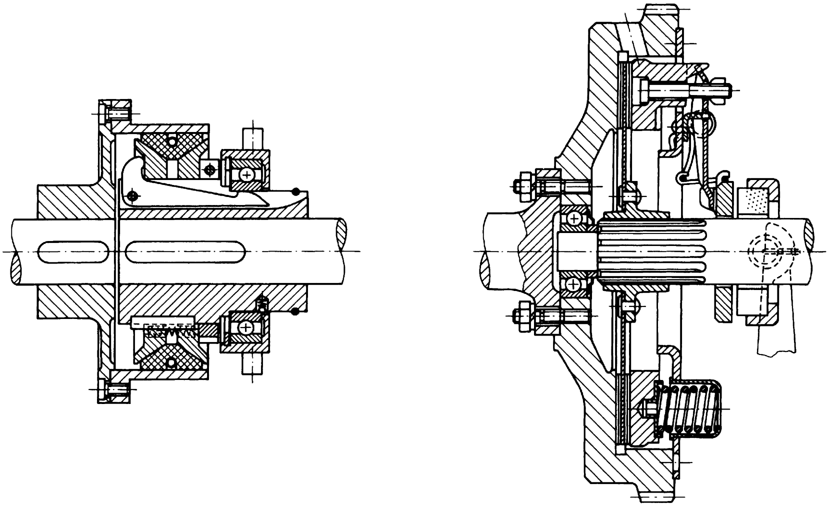
Figure 2.8. Same function structure as shown in Figure 2.5 but represented using generally valid functions, as defined in Figure 2.7

be solved. It may equally well be the relationships between subfunctions as those between inputs and outputs of particular subfunctions.

Let us first of all look at the relationships between subfunctions. As we have pointed out, certain subfunctions must be satisfied before another subfunction can be meaningfully introduced. The so-called "if–then" relationship helps to clarify this point: if subfunction *A* is present, then subfunction *B* can come into effect, and so on. Often several subfunctions must all be satisfied simultaneously before another subfunction can be put into effect. The arrangement of subfunctions thus determines the structure of the energy, material and signal conversions under consideration. Thus, during a test of tensile strength, the first subfunction—"load specimen"—must be satisfied before the other subfunctions—"measure force" and "measure deformation"—can be deployed. The last two subfunctions, moreover, must be satisfied simultaneously. Attention must be paid to consistency and order within the flow under consideration, and this is done by the unambiguous combination of the subfunctions.

Designation	AND-function (Conjunction)	OR-function (Disjunction)	NOT-function (Negation)
Symbol	X_1, X_2 — [&] — Y	X_1, X_2 — [≥1] — Y	X — [1]o— Y
Truth table	X_1 | 0 | 1 | 0 | 1 X_2 | 0 | 0 | 1 | 1 Y | 0 | 0 | 0 | 1	X_1 | 0 | 1 | 0 | 1 X_2 | 0 | 0 | 1 | 1 Y | 0 | 1 | 1 | 1	X | 0 | 1 Y | 1 | 0
Boolean algebra (Function)	$Y = X_2 \wedge X_1$	$Y = X_1 \vee X_2$	$Y = \overline{X}$

Figure 2.9. Logical functions. X independent statement (signal); Y dependent statement; "0", "1" value of statement, e.g. "off", "on"



AND

X_1, X_2 — [&] — $Y = X_1 \wedge X_2$

X_1	(Signal supplied)	0	1	0	1
X_2	(Clutch engaged)	0	0	1	1
Y	(Torque transmitted)	0	0	0	1

INHIBITION

X_1 o—, X_2 — [&] — $Y = \overline{X_1} \wedge X_2$

X_1	0	1	0	1
X_2	1	0	0	1
Y	1	0	0	0

Figure 2.10. Logical function of two clutches

Logical relationships, moreover, must also be established between the inputs and outputs of a particular subfunction. In most cases there are several inputs and outputs whose relationships can be treated like propositions in binary logic. *Elementary logical links* of the input and output magnitudes exist for this purpose. In binary logic these are statements such as true/false, yes/no, in/out, fulfilled/unfulfilled, present/not present, which can be computed using Boolean algebra.

We distinguish between AND functions, OR functions and NOT functions, and also between their combination into more complex NOR functions (OR with NOT), NAND functions (AND with NOT) and storage functions with the help of flip-flops [2.4, 2.45, 2.46]. Grouped together, these are called *logical functions*.

In the case of AND functions, all signals on the input side must have the same validity if a valid signal is to appear on the output side.

In the case of OR functions, only one signal needs to be valid on the input side if a valid signal is to appear on the output side.

In the case of NOT functions, the signal on the input side is negated so that the negated signal appears on the output side.

All of these logical functions can be expressed by standard symbols, which can be found in [2.4]. The logical validity of any signal can be read from the truth table shown in Figure 2.9, in which all of the inputs are combined systematically to yield the relevant outputs. The Boolean equations have been added for the sake of completeness. Using logical functions it is possible to construct complex switchess and thus to increase the safety and reliability of control and communication systems.

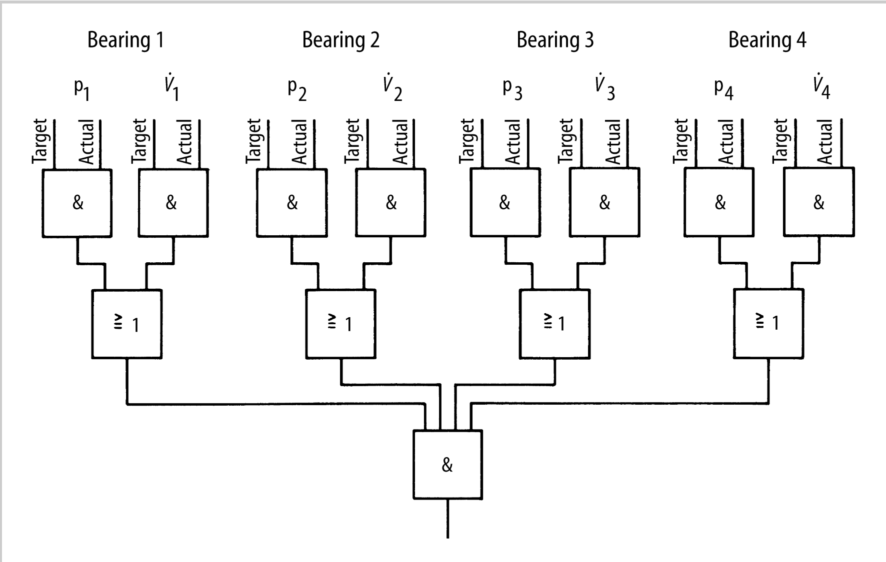


Figure 2.11. Logical functions for monitoring a bearing lubrication system. A positive signal for every bearing (oil present) permits operation. Monitor pressure p; monitor oil flow $\dot{V}$

Figure 2.10 shows two mechanical clutches with their characteristic logical functions. The workings of the clutch on the left can be represented by a simple AND function (the signal must be sent and the clutch engaged before the torque can be transmitted). The clutch on the right has been constructed such that, when the operating signal is given, the clutch is disengaged, meaning that X_1 must be negative if the torque is to be transmitted. In other words, only X_2 must be present or positive if the desired effect is to be produced.

Figure 2.11 shows a logical system for monitoring the bearing lubrication system of a multi-bearing machine shaft involving AND and OR functions. Every bearing position is monitored for oil pressure and oil flow by comparing a specified or target value with the actual value. However, only one positive value for each bearing position is needed to allow the system to operate.

2.1.4 Working Interrelationship

Establishing a function structure facilitates the discovery of solutions because it simplifies the general search for them, and also because solutions to subfunctions can be elaborated separately.

Individual subfunctions, originally represented by "black boxes", must now be replaced with more concrete statements. Subfunctions are usually fulfilled by physical, chemical or biological processes—mechanical engineering solutions are based mainly on physical processes whereas process engineering solutions are based mainly on chemical and biological processes. If, in what follows, we refer to *physical processes*, we tacitly include the effects of possible chemical and biological processes.

A physical process realised by the selected *physical effects* and the determined *geometric* and *material characteristics* results in a *working interrelationship* that ensures the function is fulfilled in accordance with the task. Hence a working interrelationship comes into existence through physical effects in combination with the chosen geometric and material characteristics.

1. Physical Effects

Physical effects can be described quantitatively by means of the physical laws governing the physical quantities involved. Thus, the friction effect is described by Coulomb's law, $F_F = \mu F_N$; the lever effect by the lever law $F_A \cdot a = F_B \cdot b$; and the expansion effect by the expansion law $\Delta l = \alpha \cdot l \cdot \Delta\vartheta$ (see Figure 2.12). Rodenacker [2.46] and Koller [2.28], in particular, have collated such effects.

Several physical effects may have to be combined in order to fulfil a subfunction. Thus the operation of a bimetallic strip is the result of a combination of two effects, namely thermal expansion and elasticity.

A subfunction can often be fulfilled by one of a number of physical effects. Thus a force can be amplified by the lever effect, the wedge effect, the electromagnetic effect, the hydraulic effect, etc. The physical effect chosen for a particular sub-function must, however, be compatible with the physical effects of other related

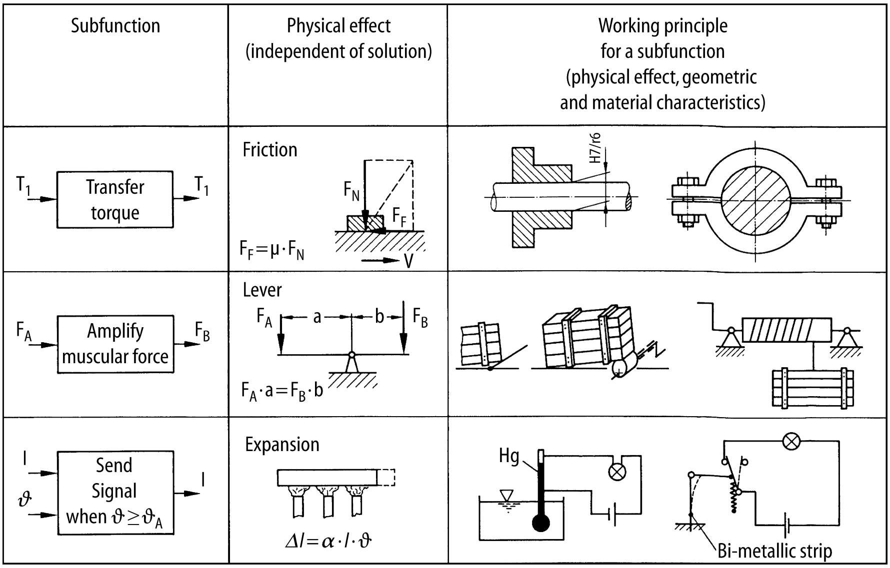


Figure 2.12. Fulfilling subfunctions by working principles built up from physical effects and geometric and material characteristics

subfunctions. A hydraulic amplifier, for instance, cannot be powered directly by an electric battery. Moreover, a particular physical effect can only fulfil a subfunction optimally under certain conditions. Thus a pneumatic control system will be superior to a mechanical or electrical control system only in particular circumstances.

As a rule, compatibility and optimal fulfilment can only be realistically assessed in relation to the overall function once the geometric and material characteristics have been established more concretely.

2. Geometric and Material Characteristics

The place where the physical process actually takes effect is the *working location*, i.e. the specific active location that is the focus of interest at the time. A function is fulfilled by the physical effect, which is realised by the *working geometry*, i.e. the arrangement of *working surfaces* (or working spaces), and by the choice of *working motions* [2.33].

The working surfaces are varied with respect to and determined by:

- Type
- Shape
- Position
- Size
- Number [2.46].

Similarly, the required working motions are determined by:

- Type: translation–rotation
- Nature: regular–irregular
- Direction: in x-, y-, z-directions and/or about x-, y-, z-axes
- Magnitude: velocity, etc.
- Number: one, several, etc.

In addition, we need a general idea of the *type of material* with which the working surfaces are to be produced, for example, whether it is solid, liquid or gaseous; rigid or flexible; elastic or plastic; stiff, hard or tough; or corrosion-resistant. A general idea of the final embodiment is often insufficient; the *main material properties* must be specified before a working interrelationship can be formulated adequately (see Figure 3.18).

Only the combination of the physical effect with the geometric and material characteristics (working surfaces, working motions and materials) allows the principle of the solution to emerge. This interrelationship is called the *working principle* (Hansen [2.19] refers to this as the working means), and it is the first concrete step in the implementation of the solution.

Figure 2.12 shows some examples:

- Transferring the torque through friction against a cylindrical *working surface* in accordance with Coulomb's law will, depending on the way in which the normal force is applied, lead to the selection of a shrink fit or a clamp connection as the working principle.
- Amplifying muscular force with the help of a lever in accordance with the lever law after determining the pivot and force application points (*working geometry*) and considering the necessary *working motion* will lead to a description of the working principle (lever solution, eccentric solution, etc.).
- Making electrical contact by bridging a gap using the expansion effect, applied in accordance with the linear expansion law, only leads to an overall working principle after determination of the sizes (e.g. the diameter and length) and the positions of the *working surfaces* needed for the *working motion* of the expanding medium: a *material*. For example, either mercury expanding by a fixed amount or a bimetallic strip serving as a switch.

To satisfy the overall function, the working principles of the various subfunctions have to be combined (see Section 3.2.4). There are obviously several ways in which this can be done. Guideline VDI 2222 [2.55] calls each combination a *combination of principles*.

The combination of several working principles results in the *working structure* of a solution. It is through this combination of working principles that the solution principle for fulfilling the overall task can be recognised. The working structure derived from the function structure thus represents how the solution will work at the fundamental principle level. Hubka refers to the working structure as the organ structure [2.22–2.24].

For known elements, a circuit diagram or a flow chart is sufficient as a means of representing a working structure. Mechanical artefacts can be effectively represented using engineering drawings, though new or uncommon elements may require additional explanatory sketches (see Figures 2.12 and 2.13).

Often the working structure alone will not be concrete enough to evaluate the solution principle. It may need to be quantified, for example by preliminary calculations and rough scale drawings, before the solution principle can be fixed. The result is called a *principle solution*.

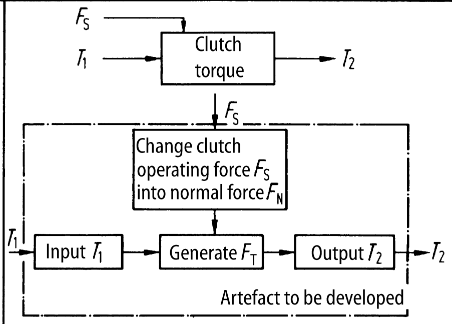
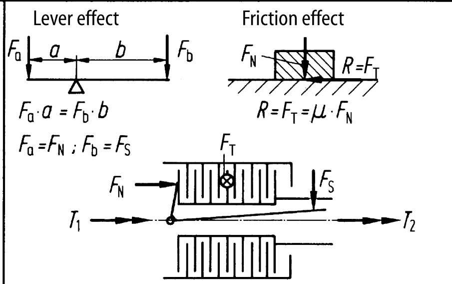
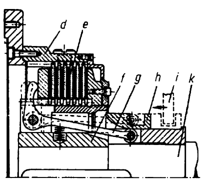
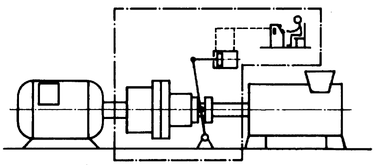
Interrelationships	Elements	Structures	Examples
Functional interrelationship	Functions	Functions structure	
Working interrelationship	Physical effects and geometric and material characteristics ↓ Working principles	Working structure	
Constructional interrelationship	Components Joints Assemblies	Construction structure	
System interrelationship	Artefacts Human beings Environment	System structure	

Figure 2.13. Interrelationships in technical systems

2.1.5 Constructional Interrelationship

The working interrelationship established in the working structure is the starting point for further concretisation leading to the *construction structure*. This interrelationship represents the concrete technical artefact or system by defining the components, assemblies and machines and their interconnections. The construction structure takes into account the needs of production, assembly, transport, etc. Figure 2.13 shows the fundamental interrelationships for the clutch shown in Figure 2.1. The increasing levels of concretisation can be seen clearly.

The concrete elements of a construction structure must satisfy the requirements of the selected working structure plus any other requirements necessary for the technical system to operate as intended. To identify these requirements fully, it is usually necessary to consider the system interrelationship.

2.1.6 System Interrelationship

Technical artefacts and systems do not operate in isolation and are, in general, part of a larger system. To fulfil its overall function, such a system often involves human beings who influence it through *input effects* (operating, controlling, etc.). The system returns *feedback effects* or signals that lead to further actions (see Figure 2.14). In this way, human beings support or enable the *intended effect* of the technical system.

Apart from desired inputs, undesired ones from the environment and from neighbouring systems can affect a technical system. Such *disturbing effects* (e.g. excess temperatures) can cause undesired *side-effects* (e.g. deviations from shape or shifts in position). Also, it is possible that in addition to the desired working interrelationship (intended effects), unwanted phenomena can occur (e.g. vibrations) as side-effects from individual components within the system or from the overall system itself. These side-effects can have an adverse effect on humans or the environment.

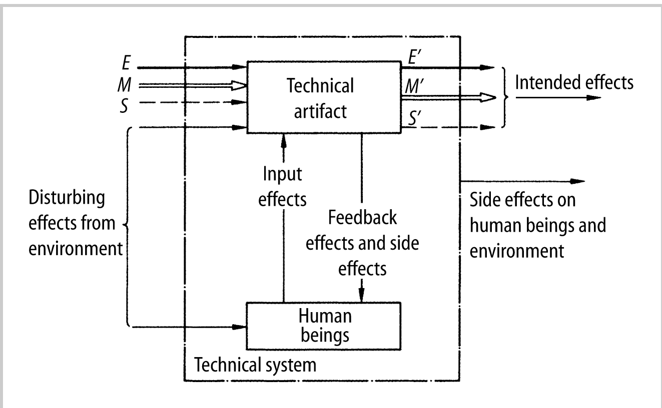


Figure 2.14. Interrelationships in technical systems including human beings

In accordance with Figure 2.14 it is useful to make the following distinctions (after [2.56]):

Intended effect:	Functionally desired effect in the sense of system operation.
Input effect:	Functional relationship due to human action on a technical system.
Feedback effect:	Functional relationship due to the action of a technical system on a human or another technical system.
Disturbing effect:	Functionally undesired influence from outside on a technical system or human that makes it difficult for a system to fulfil its function.
Side effect:	Functionally undesired and unintended effect of a technical system on a human or on the environment.

The overall interrelationship of all these effects must be carefully considered during the development of technical systems. To help recognise them in time, so that desired effects can be used and undesired ones avoided, it is helpful to follow a systematic guideline that adheres to the general objectives and constraints in Section 2.1.7.

2.1.7 Systematic Guideline

The solution of technical tasks is determined by the general objectives and constraints. The *fulfilment of the technical function*, the *attainment of economic feasibility* and the *observance of safety requirements* for humans and the environment can be considered as general objectives. The fulfilment of the technical function alone does not complete the task of designers; it would simply be an end unto itself. Economic feasibility is another essential requirement, and concern with human and environmental safety must impose itself for ethical reasons. Every one of these objectives has direct repercussions on the rest.

In addition, the solution of technical tasks imposes certain constraints or requirements resulting from ergonomics, production methods, transport facilities, the intended operation, etc., no matter whether these constraints are the result of the particular task or the general state of technology. In the first case we speak of task-specific constraints, in the second of general constraints that, although often not specified explicitly, must nevertheless be taken into account.

Hubka [2.22–2.24] separates the properties affected by the constraints into categories based variously on industrial, ergonomic, aesthetic, distribution, delivery, planning, design, production and economic factors.

Besides satisfying the functional and working interrelationships, a solution must also satisfy certain general or task-specific constraints. These can be classified under the following headings:

- Safety — also in the wider sense of reliability and availability
- Ergonomics — human–machine context, also aesthetics
- Production — production facilities and type of production

- Quality control throughout the design and production process
- Assembly during and after the production of components
- Transport inside and outside of the factory
- Operation intended use, handling
- Maintenance upkeep, inspection and repair
- Expenditure costs and schedules
- Recycling reuse, reconstitution, disposal, final storage.

The characteristics that can be derived from these constraints, which are generally formulated as requirements (see Section 5.2), affect the function, working and construction structures, and also influence one another. Hence they should be treated as *guidelines* throughout the design process, and adapted to each level of embodiment (see Figs. 2.15 and 12.3).

In addition there are influences from the designer, the development team and the suppliers as well as the customer, the specific context and the environment.

It is advisable to consider these guidelines even during the *conceptual phase*, at least in essence. During the *embodiment phase*, when the layout and form design of the more or less qualitatively elaborated working structure is first quantified, both the objectives of the task and also the general and task-specific constraints must be considered in concrete detail. This involves several steps—the collection of further information, layout and form design, and the elimination of weak spots, together with a fresh, if limited, search for solutions for a variety of subtasks, until finally,

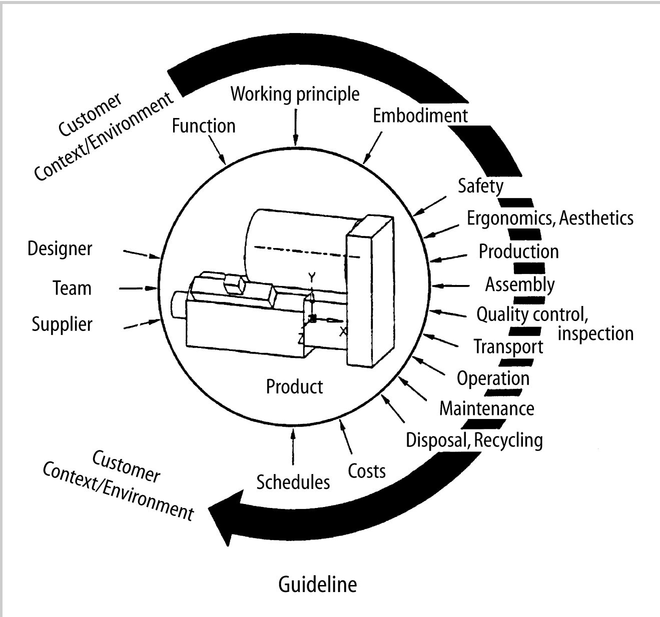


Figure 2.15. Influences and constraints during design and development. These can provide a guideline for quality control

in the *detail phase*, the elaboration of detailed production instructions brings the design process to a conclusion (see Chapters 5 to 7).

2.2 Fundamentals of the Systematic Approach

Before we deal with the specific steps and rules of systematic design, we must first discuss cognitive psychological relationships and general methodical principles. These help to structure the proposed procedures and individual methods so that they can be applied to the solution of design tasks in a purposeful way. The ideas come from a host of different disciplines, mainly non-technical ones, and are usually built on interdisciplinary fundamentals. Work science, psychology and philosophy are among the main inspirations, which is not surprising when we consider that methods designed to improve working procedures impinge on the qualities, capacities and limitations of human thought [2.41].

2.2.1 Problem Solving Process

Designers are often confronted with tasks containing problems they cannot solve immediately. Problem solving in different areas of application and at different levels of concretisation is a characteristic of their work. Researching the essence of human thinking is the focus of cognitive psychology. The results of this research must be taken into account in engineering design. The following sections are based largely on the work of Dörner [2.8, 2.10].

A *problem* has three components:

- an undesirable initial state, i.e. the existence of an unsatisfactory situation
- a desirable goal state, i.e. the realisation of a satisfactory situation
- obstacles that prevent a transformation from the undesirable initial state to the desirable goal state at a particular point in time.

An *obstacle* that prevents a transformation can arise from the following:

- The means to overcome the obstacle are unknown and have to be found (synthesis or operator problem).
- The means are known, but they are so numerous or involve so many combinations that a systematic investigation is impossible (interpolation problem, combination and selection problem).
- The goals are only known vaguely or are not formulated clearly. Finding a solution involves continuous deliberation and the removal of conflicts until a satisfactory situation is reached (dialectic problem, search and application problem).

A *problem* has the following typical characteristics:

- *Complexity:* many components are involved and these components, through links of different strength, influence each other.

- *Uncertainty:* not all requirements are known; not all criteria are established; the effect of a partial solution on the overall solution or on other partial solutions is not fully understood, or only emerges gradually. The difficulties become more pronounced if the characteristics of the problem area change with time.

A *task* is distinct from a problem because:

- A *task* imposes mental requirements for which various means and methods are available to assist. An example is the design of a shaft with given loads, connecting dimensions and production methods.

Tasks and problems occur in design in a number of ways, often combined and not clearly separable initially. A specific design task can, for example, turn out to be a problem when looked at more closely. Many large tasks can be divided into subtasks, some of which can reveal difficult subproblems. On the other hand, it is sometimes possible for a problem to be solved by fulfilling several subtasks in a previously unknown combination.

Thinking processes take place in the brain and involve changes in memory content. When thinking, the contents of the memory, and the way in which they are linked, play an important role.

In simple terms, one can say that in order to start solving a problem humans need a certain level of *factual knowledge* about the domain of the problem. In cognitive psychology, when this knowledge has been transferred into memory it represents the *epistemic structure*.

Humans also need certain *procedures* (methods) to find solutions and to find these effectively. This aspect involves the *heuristic structure* of human thought.

It is possible to distinguish between short-term and long-term memory. Short-term memory is a kind of working storage. It has limited capacity and can only retain about seven arguments or facts at the same time. Long-term memory probably has unlimited capacity and contains factual and heuristic knowledge that appears to be stored in a structured way.

In this way, humans are able to recognise specific relationships in many possible ways, to use these relationships and to create new ones. Such relationships are very important in the technical domain, for example:

- concrete—abstract relationship

  e.g. angular contact bearing—ball bearing—rolling element bearing—bearing—guide—transfer force and locate component.

- whole—part relationship (hierarchy)

  e.g. plant—machine—assembly—component.

- space and time relationships

  e.g. arrangement: front—back, below—above,

  e.g. sequence: this first—that next.

The memory can be thought of as a semantic network with nodes (knowledge) and connections (relationships) which can be modified and extended. Figure 2.16 shows a possible, though not necessarily complete, semantic network related to

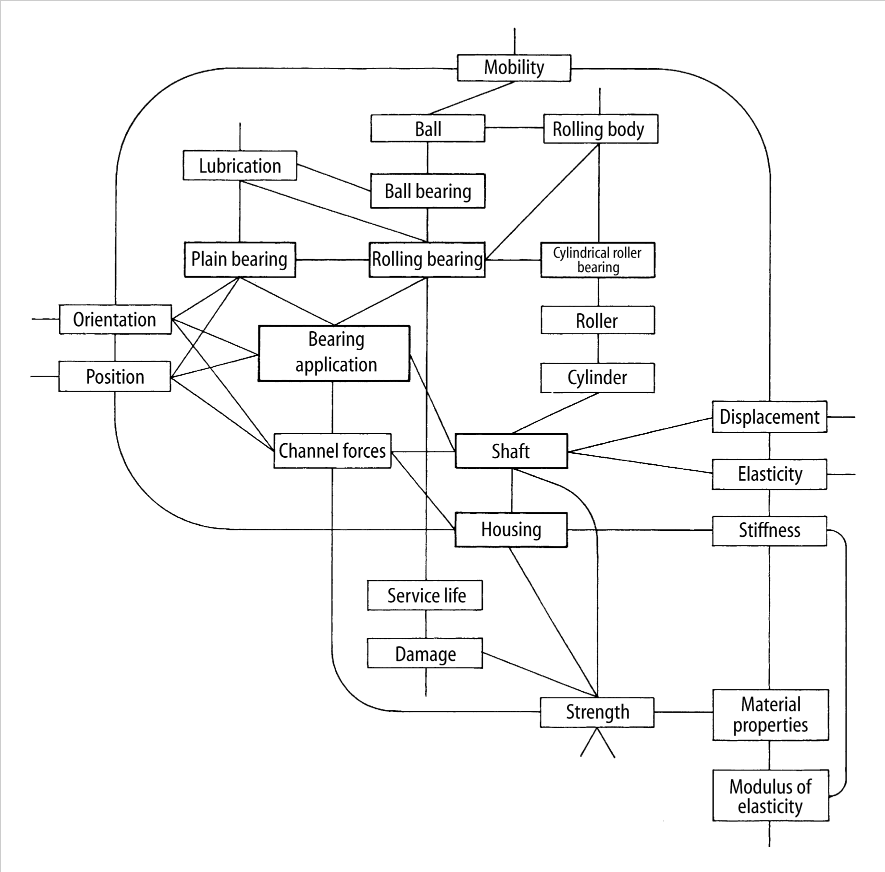


Figure 2.16. Extract of a semantic network related to bearings

the term "bearing". In this network it is possible to recognise the relationships mentioned above as well as others, such as property relationships and ones indicating opposites (polar relationships). Thinking involves building and restructuring such semantic networks, and the thinking process itself can proceed intuitively or discursively.

Intuitive thinking is strongly associated with flashes of inspiration. The actual thinking process takes place to a large extent unconsciously. Insights appear in the conscious mind suddenly, caused by some trigger or association. This is referred to as primary creativity [2.2, 2.30] and involves processing quite complex relations. In this context, Müller [2.36] refers to "silent knowledge", which includes common and background knowledge. This is also the knowledge that is available when one deals with episodic memories, vague concepts and imprecise definitions. It is activated by both conscious and unconscious thinking activities.

Generally time is needed for undisturbed and unconscious "thinking" before sudden insights appear. The length of this incubation period cannot be predetermined. Insights can be triggered, for example, by producing freehand sketches or engineering drawings of solution ideas. According to [2.14], these manual activities focus concentration on the subject, but still leave space in the mind that can by used by unconscious thinking processes, which can also be stimulated by such activities.

Discursive thinking is a conscious process that can be communicated and influenced. Facts and relationships are consciously analysed, varied, combined in new ways, checked, rejected, and considered further. In [2.2,2.30] this is referred to as secondary creativity. This type of thinking involves checking exact and scientific knowledge and building this into a knowledge structure. In contrast to intuitive thinking, this process is slow and involves many small conscious steps.

In the memory structure, explicit and consciously acquired knowledge cannot be separated precisely from the vaguer common or background knowledge. Besides, the two types of knowledge influence each other. For knowledge to be easily retrieved and combined, it is thought that an ordered and logical structure of factual knowledge in the mind of the problem solver (epistemic structure) is decisive, and that this is true whether the thinking process is intuitive or discursive.

The *heuristic structure* includes explicit knowledge (i.e. knowledge that can be explained) as well as implicit knowledge. This is necessary in order to organise the sequence of thinking operations, including modifying operations (searching and finding) and testing operations (checking and assessing). It appears that problem solvers often start without a fixed plan in the hope of immediately finding a solution from their knowledge bases without much effort. Only when this approach fails, or when contradictions emerge, do they adopt a more clearly planned or systematic sequence of thinking operations.

The so-called TOTE model [2.33] represents an important fundamental sequence for thinking processes (see Figure 2.17). It consists of two processes: a modification process and a testing process. The TOTE model shows that before an operation of change takes place, an operation of testing (*T*est) is invoked to analyse the initial state. Only then is the chosen operation of change (*O*peration) executed. This is followed by another operation of testing (*T*est), during which the resulting state is checked. If the result is satisfactory, the process is exited (*E*xit); if not, the operation is adapted and repeated.

In more complex thinking processes, the TOTE sequences are linked in a chain or several modification processes are executed before a testing process takes place. Thus, when linking mental processes, many combinations and sequences are possible, but all of them can be mapped onto the basic TOTE model.

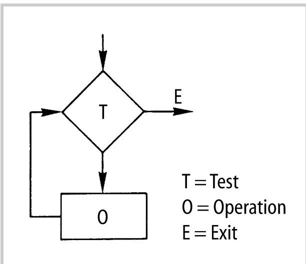


Figure 2.17. Basic TOTE model for thinking processes [2.8, 2.33]

2.2.2 Characteristics of Good Problem Solvers

The following statements are the result of the work of Dörner [2.9] and of research which has been undertaken with him by Ehrlenspiel and Pahl. The results of the research led by Ehrlenspiel and Pahl can be found in the publications of Rutz [2.50], Dylla [2.11, 2.12] and Fricke [2.15, 2.16]. This section provides a summary of their findings [2.42].

1. Intelligence and Creativity

In general, *intelligence* is thought to involve a certain cleverness, combined with the ability to understand and judge. Analytical approaches are often emphasised.

Creativity is an inspirational force that generates new ideas or produces novel combinations of existing ideas, leading to further solutions or deeper understanding. Creativity is often associated with an intuitive, synthesising approach.

Intelligence and creativity are personal characteristics. Up until now it has not been possible to come up with precise scientific definitions of or a clear distinction between intelligence and creativity. Attempts have been made to measure the level of intelligence of individuals using intelligence tests. The resulting Intelligence Quotients provide measures compared to the average of a large sample. Because of the different forms in which intelligence appears, various tests are needed to get a complete picture and draw tentative conclusions. The same is true for creativity tests.

For problem solving, a minimum level of intelligence is required and it appears that people with high Intelligence Quotients are more likely to be good problem solvers. However, according to [2.8, 2.9], intelligence tests on their own do not give much insight into which combination of factors makes a particular individual a good problem solver. The reason, according to Dörner [2.8], is that intelligence tests use tasks or problems that only require a few thinking steps to find a solution, so the sequence of steps seldom becomes conscious. Few intelligence tests require a large number of steps to be organised into a specific problem solving procedure. Such organisation requires switching between the different levels and possibilities of a general problem solving procedure, and is essential for the execution of long-term thinking activities.

Creativity tests too are often at such a low level that they do not address complex problem solving which involves planning and guiding one's own approach. Furthermore, in engineering design, creativity is always focused on a specific goal. Purely unfocused generation of ideas and variants can in fact hinder the problem solving process [2.2] or at best support a specific phase of the process.

2. Decision Making Behaviour

Apart from having well-structured factual knowledge, applying a systematic approach, and using focused creativity, designers have to master decision making processes. For decision making, the following mental activities and skills are essential:

- *Recognising Dependencies*

  In complex systems the dependencies between the individual elements can vary in strength. Recognising the types and strengths of such dependencies is an essential prerequisite for dividing the problem into more manageable, less complex subproblems or subgoals so that these can be addressed separately. However, those working on each separate subproblem must check to see how the short- and long-term effects of their own decisions influence the overall design.

- *Estimating Importance and Urgency*

  Good problem solvers know how to recognise *importance* (factual significance) and *urgency* (temporal significance), and how to use this information to modify their approach to problems.

  They try to resolve the most important things first and then tackle the dependent subproblems. They have the courage to be satisfied with suboptimal solutions for less significant problems if they have good or acceptable solutions for the most significant ones. By doing this they avoid immersing themselves in less relevant issues and thereby losing valuable time.

  The same is true when estimating the urgency. Good problem solvers estimate the time they need accurately. They prepare a demanding—but not impossible—time plan. Janis and Mann [2.25] have concluded that mild (i.e. bearable) stress is important for creativity. Therefore, realistic time planning has a positive effect on thinking processes, and new developments should take place under reasonable time pressure. But, of course, individuals react differently to time pressure.

- *Continuity and Flexibility*

  Continuity means an appropriate and continuous focus on achieving the goals, but there is a danger that excessive focus leads to a rigid approach. Flexibility means a ready ability to adapt to changing requirements. However, this should not lead to purposeless jumping from one approach to another.

  Good problem solvers find a suitable balance between continuity and flexibility. They demonstrate continuous and consistent, but at the same time flexible, behaviour. They stick to the given goals despite any hold-ups and difficulties they encounter. On the other hand, they adapt their approach immediately when the situation changes and when new problems occur.

  They consider heuristics, procedures and instructions first of all as guidelines and not as rigid prescriptions. Dörner states [2.8]: "Heuristics or heuristic plans should not degenerate into automatic procedures. Individuals should learn to develop what they have learnt. Heuristics should not be misinterpreted as prescriptions, but should be treated as guidelines that can, and often should, be developed."

- *Failures Cannot be Avoided*

  In complex systems with strong internal dependencies, at least partial failures are difficult to avoid because it is not possible to recognise all the potential effects simultaneously. When recognising such failures, the most important thing is the way one reacts. Being flexible is crucial, supported by the ability to analyse one's approach and the ability to make decisions that lead to corrective actions.

The results of cognitive psychology research are summarised below.

Good problem solvers:

- have a sound and structured technical knowledge, i.e. they have a well-structured model in their minds
- find an appropriate balance between concreteness and abstraction, depending on the situation
- can deal with uncertainty and fuzzy data
- continuously focus on the goals while adopting a flexible decision making behaviour.

Such heuristic competence depends largely on personal characteristics, but can be developed considerably through training on different types of problem.

The research mentioned earlier reveals that good designers demonstrate the following behaviour [2.42]:

- They thoroughly analyse the goals at the beginning of a task and continue to do so throughout the design process when formulating partial goals, in particular when the original problem formulation is vague.
- They first generate or identify the most suitable solution principles in a conceptual phase before developing concrete embodiments.
- They initially adopt a diverging search without generating too many variants and then quickly converge onto a small number of solutions; they choose the appropriate level of concretisation and switch easily between perspectives, e.g. abstract/concrete, overall problem/subproblem, working interrelationship/constructional interrelationship.
- They regularly assess their solutions using a comprehensive set of criteria, avoiding emphasising personal preferences.
- They continuously reflect on their approach and adapt it to the situation at hand.

These characteristics are in line with the aims and proposals for the design approach in this book.

2.2.3 Problem Solving as Information Processing

When we discussed the basic ideas of the systems approach (see Section 1.2.3), we found that problem solving demands a large and constant flow of information. Dörner [2.8] also views problem solving as information processing. The most important terms used in the theory of information processing are described in [2.5, 2.6]. Information is *received*, *processed* and *transmitted* (see Figure 2.18).

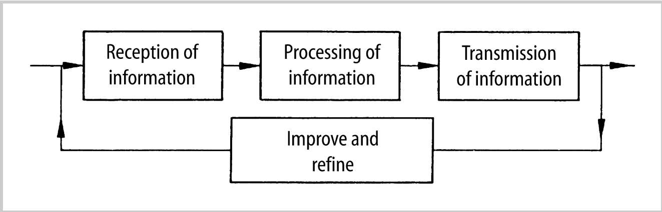


Figure 2.18. The conversion of information with iteration

Information is *received* from market analyses, trend studies, patents, technical journals, research results, licenses, inquiries from customers, concrete assignments, design catalogues, analyses of natural and artificial systems, calculations, experiments, analogies, general and in-house standards and regulations, stock sheets, delivery instructions, computer data, test reports, accident reports, and also by "asking questions". Data collection is an essential element of problem solving [2.3].

Information is *processed* by analysis and synthesis, the development of solution concepts, calculation, experiment, the elaboration of layout drawings and also the evaluation of solutions.

Information is *transmitted* by means of sketches, drawings, reports, tables, production documents, assembly manuals, user manuals, etc. These can be both in hard copy and electronic forms. Quite often provision must also be made for information to be *stored*.

In [2.32] some *criteria* for characterising information are given, and these can be used for formulating user information requirements. They include:

- Reliability: the probability of the information being available, trustworthy and correct.
- Sharpness: the precision and clarity of the information content.
- Volume and density: an indication of the number of words and pictures needed for the description of a system or process.
- Value: the importance of the information to the recipient.
- Actuality: an indication of the point in time when the information can be used.
- Form: the distinction between graphic and alphanumeric data.
- Originality: an indication of whether or not the original character of the information must be preserved.
- Complexity: the structure of, or connectivity between, information symbols and information elements, units or complexes.
- Degree of refinement: the quantity of detail in the information.

Information conversion is usually a very complicated process. Solving problems requires information of different types, content and range. In addition, to raise the level of information and improve it, it may be necessary to reiterate certain steps.

Iteration is the process by which a solution is approached step-by-step. In this process, one or more steps are repeated, each time at a higher level of information based on the results of the previous loop. Only in this way is it possible to obtain the information to refine a solution and ensure continuous improvement (see Figure 2.18). Such iterations occur frequently at all stages of the problem-solving process.

2.2.4 General Working Methodology

A general working methodology should be widely applicable, independent of discipline and should not require specific technical knowledge from the user. It should support a structured and effective thinking process. The following general ideas appear time and time again in specific approaches, either directly or slightly amended to adapt them to the special requirements of developing technical systems. The purpose of this section is to provide a general introduction to systematic procedures. The following procedures are based not only on our own professional experience and on the findings of cognitive psychology mentioned in Section 2.2.1, but also on the work of Holliger [2.20, 2.21], Nadler [2.38, 2.39], Müller [2.35, 2.36] and Schmidt [2.51]. They are also known as "heuristic principles" (a *heuristic* is a method for generating ideas and finding solutions) or "creativity techniques".

The following conditions must be satisfied by anyone using a systematic approach:

- *Define goals* by formulating the overall goal, the individual subgoals and their importance. This ensures the motivation to solve the task and supports insight into the problem.
- *Clarify conditions* by defining the initial and boundary constraints.
- *Dispel prejudice* to ensure the most wide-ranging search for solutions possible and to avoid logical errors.
- *Search for variants* to find a number of possible solutions or combinations of solutions from which the best can be selected.
- *Evaluate* based on the goals and conditions.
- *Make decisions*. This is facilitated by objective evaluations. Without decisions and experiencing their consequences there can be no progress.

To make these general methods work, the following *thinking* and *acting operations* must be considered.

1. Purposeful Thinking

As described in Section 2.2.1, intuitive and discursive thinking are possible. The former tends to be more unconscious, the latter more conscious.

Intuition has led to a large number of good and even excellent solutions. The prerequisite is, however, always a very conscious and intensive involvement with the given problem. Nevertheless, a purely intuitive approach has the following disadvantages:

- the right idea rarely comes at the right moment, since it cannot be elicited and elaborated at will
- the result depends strongly on individual talent and experience
- there is a danger that solutions will be circumscribed by preconceived ideas based on one's special training and experience.

It is therefore advisable to use more deliberate procedures that tackle problems step-by-step, and such procedures are denoted *discursive*. Here the steps are chosen intentionally; they can be influenced and communicated. Usually individual ideas or solution attempts are consciously analysed, varied and combined. It is an important aspect of this procedure that a problem is rarely tackled as a whole, but is first divided into manageable parts and then analysed.

It must, however, be stressed that intuitive and discursive methods are not opposites. Experience has shown that intuition is stimulated by discursive thought. Thus, while complex assignments must always be tackled one step at a time, the subsidiary problems involved may, and often should, be solved in intuitive ways.

In addition, it should be realised that creativity can be inhibited or encouraged by different influences [2.2]. It is, for example, often necessary to encourage intuitive thinking by interrupting the activity to provide some periods of incubation (see Section 2.2.1). On the other hand, too many interruptions can be disturbing and thereby inhibit creativity. A systematic approach including discursive elements and adopting different viewpoints encourages creativity. Examples include using different solution methods; moving between abstract and concrete ideas; collecting information using solution catalogues; and dividing work between team members. Furthermore, according to [2.25], realistic planning encourages rather than inhibits motivation and creativity.

2. Individual Working Styles

Designers should be given some freedom of action in their work to enable them to realise their own optimised working style. They should be free to select their preferred methods, the sequence in which they undertake individual working steps, and the sources of information they wish to consult. They should therefore be allowed to make their own plans for their area of responsibility and for them to have control over these plans. Obviously the individual working plans have to be compatible with the overall approach and make a useful contribution.

In general it is necessary to consider several subfunctions (subproblems) when developing new products. These functions, or combinations of them, lead to partial solutions. In such situations designers can proceed in different ways. One possibility is to search for working principles (solution principles) for every subfunction (or group of subfunctions), to roughly check their compatibility, and then to combine them into an overall working structure (solution concept). Finally the components are embodied, making sure their overall combination is compatible.

From a methodical point of view, this approach is systematic, stepwise and process-oriented; that is, the designer develops the different functional areas in parallel, from abstract (idea generation) to concrete (final embodiment) (see Figure 2.19a).

Another possibility is to proceed from idea generation to final embodiment for every problem or functional area, one after the other, and finally combine and modify these to make them all fit together. From a methodical point of view, this approach is problem-oriented; that is, the designer develops the different functional areas in sequence (see Figure 2.19b).

The investigations of Dylla [2.11, 2.12] and Fricke [2.15, 2.16] show that novices educated in systematic design tend to follow the process-oriented approach, whereas experienced designers tend to follow the problem-oriented approach. Experienced designers apply their wealth of experience, know a wide range of possible subsolutions, and are able to represent these solutions quickly. Hence they arrive relatively quickly at a concrete result. Then, using a corrective approach, they bring this together into an overall solution. This type of approach is successful in those cases where the individual components do not influence each other strongly and their properties are apparent. If these conditions are not met, this approach can lead to a relatively late recognition of a possible lack of compatibility between the functional areas. This approach can also result in different subsolutions being selected for identical, or similar, subfunctions, which is often not economic. In such cases further iterations are required to find other solutions.

The process-oriented approach largely avoids the potential disadvantages of the problem-oriented approach. However, more time is required because of the wider, more systematic perspective. This carries the danger of generating an unnecessarily large solution space. The process-oriented approach therefore requires designers to achieve an appropriate balance between abstract and concrete; that is, to know when a sufficiently large, but not too large, number of solution ideas has been generated (divergence), and the time has come to combine these into a concrete concept (convergence).

In practice, these two approaches (process-oriented and problem-oriented) are often not found in their pure form. They usually appear in various combinations depending on the problem situation. However, individual designers naturally tend to adopt one approach in preference to the other. Process-oriented approaches are recommended when subproblems are strongly interrelated and when breaking new ground. A problem-oriented approach is useful when the connectivity between functional areas is low and when subsolutions are known to exist in the area of application.

Similarly individual differences in approach can be observed during the search for solutions. If designers develop and investigate different solution principles or embodiment variants in parallel while searching for solutions for the individual subfunctions, and then compare these with one another to find the most suitable, this approach is called a *generative search for solutions* (see Figure 2.20a). If, on the other hand, a particular idea or example is used as a starting point and is then improved and adapted in a stepwise approach until a satisfactory solution emerges, this is called a *corrective search for solutions* (see Figure 2.20b). Adopting

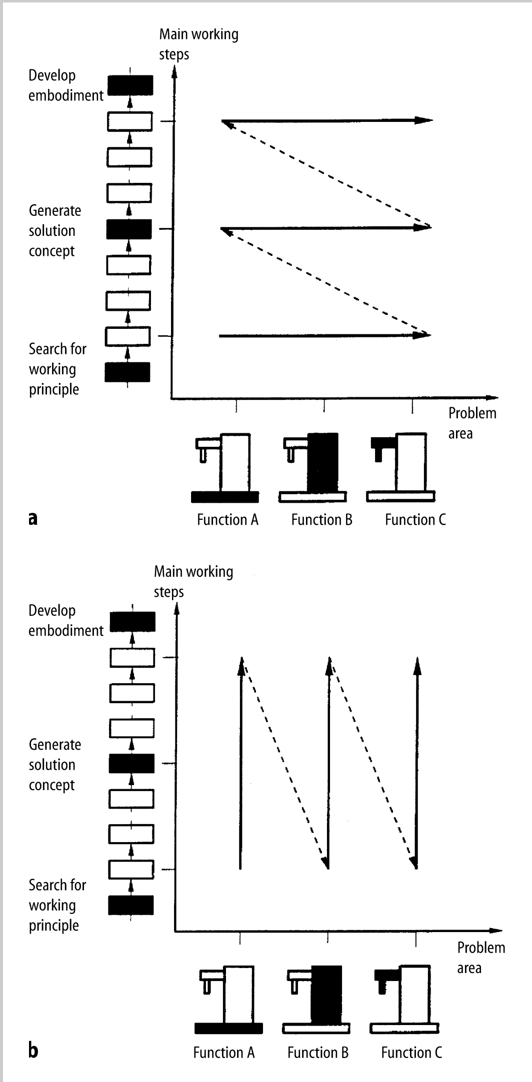


Figure 2.19. Different individual approaches during the development of solutions for a tea-making machine with several linked functional areas: baseplate/control (function A), water reservoir and heating element (function B), spout and closure (function C). **a** Systematic, stepwise, process-oriented, i.e. in every stage of development all functional areas are taken forward; **b** Problem-oriented, i.e. functional areas are developed in sequence before combining them (idealised process representation after Fricke [2.15, 2.16])

this latter approach will also result in a range of solution variants, if individual variants are not rejected.

A generative search for solutions increases the chances of finding new and unconventional ideas and considers many different principles, and thus may result in a larger solution space. The challenge, however, is a timely and goal-oriented

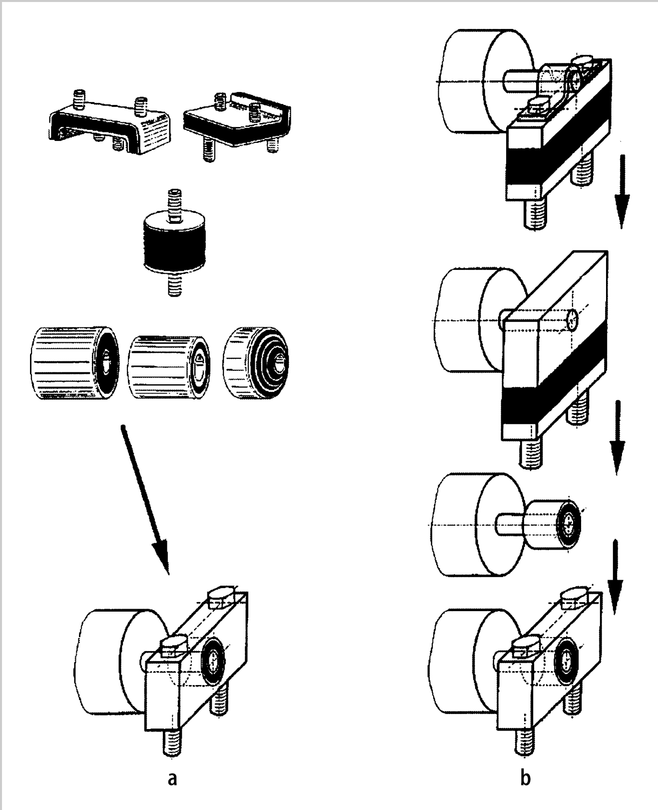


Figure 2.20. Different individual approaches during the search for solutions for an elastic support. **a** Generative, i.e. generation of various solutions and goal-oriented selection. **b** Corrective, i.e. search for solutions by improvement and adaptation of one idea

selection to avoid wasting time on unfeasible solutions. This type of search is typical for novices who have been taught systematic design and for designers who have adopted the systematic approach.

A corrective search for solutions is often used by inexperienced designers, in particular when they can think of a similar known solution in the application area. The advantage is that it is possible to concretise the solutions relatively quickly, even if these initial solutions are not really satisfactory. When adopting this type of search, designers tend to remain in their area of expertise and only expand this slowly. Possible dangers include fixating on solution ideas that are less suitable in principle and failing to recognise other better solution principles.

In practice, designers tend to adopt a mixture of search types with the main aim of minimising their work effort. However, designers clearly favour one or the other search type because of their individual talents and experience, usually without being aware of the advantages or dangers of their particular styles.

The consciously or unconsciously applied approaches depend on education and experience and can be influenced. Designers should not be forced into adopting a particular approach. On the contrary, it is better to make them aware of the advantages and dangers of the various approaches and leave the final decision up to them. It is, however, useful through training and further education, along with appropriate management during the project, to identify the most suitable overall approach and to agree on this.

2.2.5 Generally Applicable Methods

The following general methods provide further support for systematic work, and are widely used [2.21]. Often so-called "new" methods only involve repackaging one of the general methods described below.

1. Analysis

Analysis is the resolution of anything complex into its elements and the study of these elements and their interrelationships. It calls for identification, definition, structuring and arrangement. The acquired information is transformed into knowledge. If errors are to be minimised, then problems must be formulated clearly and unambiguously. To that end, they have to be analysed. *Problem analysis* means separating the essential from the nonessential and, in the case of complex problems, preparing a discursive solution by resolution into individual, more transparent, subproblems. If the search for the solution proves difficult, a new formulation of the problem may provide a better starting point. The reformulation of statements is often an effective means of finding new ideas and insights. Experience has shown that careful analysis and formulation of problems are among the most important steps of the systematic approach.

The solution of a problem can also be brought nearer by *structure analysis*, that is, the search for hierarchical structures or logical connections. In general, this type of analysis can be said to aim at the demonstration of similarities or repetitive features in different systems, for example by means of analogical reasoning (see Section 3.2.1).

Another helpful approach is *weak spot analysis*. It is based on the fact that every system has weaknesses caused by ignorance, mistaken ideas, external disturbances, physical limitations and production errors. During the development of a system it is therefore important to analyse the design concept or design embodiment for the express purpose of discovering possible weak spots and prescribing remedies. To that end, special selection and evaluation procedures (see Section 3.3) and weak spot identification methods (see Section 10.2) have been developed. Experience has shown that this type of analysis may not only lead to specific improvements of the chosen solution principle, but may also trigger off new solution principles.

2. Abstraction

Through abstraction it is possible to find a higher level interrelationship, that is, one which is more generic and comprehensive. Such a procedure reduces complexity and emphasises the essential characteristics of the problem and thereby provides an opportunity to search for and find other solutions containing the identified characteristics. At the same time new structures emerge in the minds of designers and these assist with the organisation and retrieval of the many ideas and representations. So abstraction supports both creativity and systematic thinking. It makes possible the definition of a problem in such a way that a coincidental solution path is avoided and a more generic solution is found (see example in Section 6.2).

3. Synthesis

Synthesis is the fitting together of parts or elements to produce new effects and to demonstrate that these effects create an overall order. It involves search and discovery, and also composition and combination. An essential feature of all design work is the combination of individual findings or subsolutions into an overall working system—in other words, the association of components to form a whole. During the process of synthesis the information discovered by analyses is processed as well. In general, it is advisable to base synthesis on a holistic or systems approach; in other words, to bear in mind the general task or course of events while working on subtasks or individual steps. Unless this is done, there is the grave risk that, despite the optimisation of individual assemblies or steps, no suitable overall solution will be reached. Appreciation of this fact is the basis of the interdisciplinary method known as Value Analysis, which proceeds from the analysis of the problem and structure to a holistic systems approach involving the early collaboration of all departments concerned with product development. Such an approach is also needed in large-scale projects, especially when preparing schedules by such techniques as critical path analysis (see Section 4.2.2). The entire systems approach and its methods are strongly based on holistic thinking, which is particularly important in the selection of evaluation criteria, because the value of a particular solution can only be gauged after overall assessment of all of the expectations, requirements and constraints (see Section 3.3.2).

4. Method of Persistent Questions

When using systematic procedures it is often a good idea to keep asking questions of both oneself and of others as a stimulus to fresh thought and intuition. A standard list of questions also fosters the discursive method. In short, asking questions is one of the most important methodological tools. This explains why many authors have drawn up special checklists for various working steps to support this method.

5. Method of Negation

The method of deliberate negation starts from a known solution, splits it into individual parts or describes it by individual statements, and negates these statements one-by-one or in groups. This deliberate inversion often creates new solution possibilities. Thus, when considering a “rotating” machine element, one might also examine the “static” case. Moreover, the mere omission of an element can be tantamount to a negation. This method is also known as “systematic doubting” [2.21].

6. Method of Forward Steps

Starting from a first solution attempt, one follows as many paths as possible to produce further solutions. This method is also called the method of divergent thought. It is not necessarily systematic, but frequently starts with an unsystematic divergence of ideas. The method is illustrated in Figure 2.21 for the development of a shaft–hub connection. The arrows indicate the direction of the thinking process.

Such a thinking process can be improved by using classifying criteria (see Figure 3.18) to support the systematic variation of the characteristics (see Figure 3.21). Where variation is done without conscious thought, even with well-structured representations, the identified characteristics are not used to their full potential.

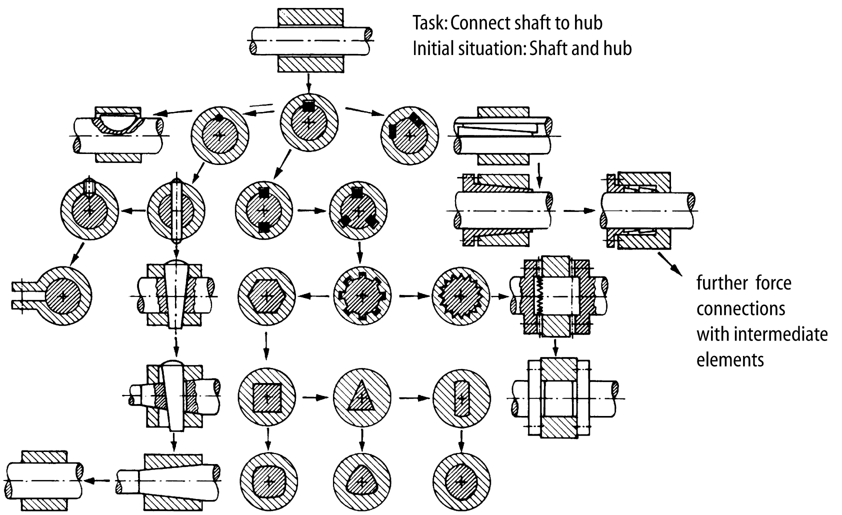


Figure 2.21. Development of shaft–hub connections in accordance with the method of forward steps

7. Method of Backward Steps

The starting point for this method is the goal rather than the initial problem. Beginning with the final objectives of the development, one retraces all of the possible paths that may have led up to it. This method is also called the method of convergent thought, because only ideas that converge on the ultimate goal are developed.

The method is particularly useful for drawing up production plans and developing systems for the production of components.

It is similar to the method of Nadler [2.38], who has proposed the construction of an ideal system that will satisfy all demands. This system is not developed in practice but formulated in the mind. It demands optimum conditions, such as an ideal environment which causes no external disturbances. Having formulated such a system, this is followed by a step-by-step investigation of what concessions must be made to turn this purely theoretical and ideal system into a technologically feasible one, and then finally into one that meets all the concrete requirements. Unfortunately, it is rarely possible to specify the ideal system in advance, because the ideal state of all functions, system elements and modules is difficult to specify, especially if they are linked together in a complex system.

8. Method of Factorisation

Factorisation involves breaking down a complex interrelationship or system into manageable, less complex and more easily definable individual elements (factors). The overall problem or task is divided into separate subproblems or subtasks that are, to a certain degree, independent (see Figure 2.3). Each of these subproblems or subtasks can initially be solved on its own, though the links between them in the overall structure must be kept in mind. Factorisation not only creates more manageable subtasks but it also clarifies their importance and influence in the overall structure, allowing priorities to be set. This approach is used in systematic design to divide an overall function into subfunctions and to develop function structures (see Sections 2.1.3 and 6.3), to search for working principles for subfunctions (see Section 6.4), and to plan the working steps during conceptual and embodiment design (see Section 4.2).

9. Method of Systematic Variation

Once the required characteristics of the solution are known, it is possible, by systematic variation, to develop a more or less complete solution field. This involves the construction of a generalised classification, that is, a schematic representation of the various characteristics and possible solutions (see Section 3.2.3). From the viewpoint of work science, too, it is obvious that the discovery of solutions is assisted by the construction and use of classification schemes. Nearly all authors consider systematic variation to be one of the most important methods.

10. Division of Labour and Collaboration

An essential finding of work science is that the implementation of large and complex tasks calls for the division of labour; more so as specialisation increases. This is also demanded by the increasingly tight schedules of modern industry. Now, division of labour implies interdisciplinary collaboration which, in turn, involves special organisational and staff arrangements along with appropriate staff attitudes, including receptiveness to the ideas of others. It must, however, be stressed that interdisciplinary collaboration and teamwork also demand a rigorous allocation of responsibility. Thus, the product manager should be in sole charge of the development of a particular product, regardless of departmental boundaries (see Section 4.3).

Systematic design, in combination with methods that make use of group dynamics, such as brainstorming, gallery method (see Section 3.2.3) and group evaluation (see Section 3.3), can overcome any lack of information exchange caused by the division of work, and can also help the search for solutions by stimulating ideas between team members.

2.2.6 Role of Computer Support

The systematic approach to design presented in this book can, in principle, be applied without the use of computers. However, the approach provides a sound basis for computer support of the design and development process that goes far beyond the use of complex analytical tools such as FEA and CFD, and the production of complex 3-D models. Computer support can be provided continuously throughout the process, for example, through the use of CAD, CAE, CAM, CIM, PDM and PLM software suites. The general use of IT also supports product improvement and reduces design and production effort.

It is not the purpose of this book, nor is there space, to describe the fundamental support that computers provide throughout the design process in detail. This topic is comprehensively covered in other texts, such as [2.1, 2.13, 2.17, 2.18, 2.27, 2.34, 2.37, 2.43, 2.44, 2.48, 2.52, 2.54, 2.57].

http://www.springer.com/978-1-84628-318-5

Engineering Design
A Systematic Approach
Pahl, G.; Beitz, W.; Feldhusen, J.; Grote, K.-H.
2007, XXI, 617 p., Hardcover
ISBN: 978-1-84628-318-5