Organizations: Social Systems Conducting Experiments

Contents

Chapter 1
Introducing Organizations as Social Systems Conducting Experiments

1.1 Introduction

In this book, it is our aim to describe organizations as social systems conducting experiments with their survival. More in particular, we want to *explain what we mean* by this description, and based on this explanation, we want to formulate *principles for the design* of organizations, enabling them to survive, i.e., enabling them to continue to conduct these experiments.

Organizations as social systems conducting experiments: "What kind of description is that?" "Can it deepen our understanding of organizations?" "Can it help to improve the conditions for their survival by providing principles underpinning organizational design?" and if so, "What are these design principles?" These are all relevant and "natural" questions that might come up when reading the aim of this book. We do think it deepens our understanding of organizations and allows for finding principles improving their design. However, it may take the rest of the book to argue why. In this introduction, we cannot exhaustively answer these questions, so we have to content ourselves with a tentative and hopefully sufficiently persuasive description of the main topic of the book: organizations as social systems conducting experiments and finding principles to improve their design.

1.2 Making Sense of Organizations: From "Phenomena" to "Key Features"

To appreciate our thesis that organizations are social systems conducting experiments, it may be useful to show that it concurs with our everyday experience with them. We do this by presenting two everyday organizational "situations" by pointing at their key features, and by relating these key features into a "theory" that can be used to make sense of these and other everyday organizational situations.

This "procedure" of starting with everyday experiences to end up with a theory allowing to understand them has a large tradition, for it was already a part of

J. Achterbergh and D. Vriens, *Organizations,*
DOI: 10.1007/978-3-642-00110-9_1, © Springer-Verlag Berlin Heidelberg 2010

Aristotle's method of study (cf. Aristotle's Posterior Analytics 1984b; Ackrill 1981). Aristotle's method starts with "phenomena," things as they present themselves in our experience. Given some phenomenon of interest, he collected commonly held opinions about it (called the "endoxa"). He analyzed these opinions to arrive at key features of the phenomenon: its "archai." Finally, he related these "archai" into a theory allowing for an understanding of the phenomenon. By reasoning from the phenomenon, the thing as it presents itself in our experience, to its (related) archai, Aristotle wanted to both remain as true as possible to the phenomenon and penetrate to its "intelligible core"; to that which is characteristic of it, and cannot be negated without negating the phenomenon altogether.

Although Aristotle proposed his method to study all kinds of phenomena, it seems to be particularly useful if the phenomenon under consideration is something we are already thoroughly involved in, yet is hard to pin down exactly; a phenomenon like "organization."

To explain our perspective on organizations, we take as a point of departure that we are already entangled with the phenomenon we call "organization." This seems a reasonable thing to do, for organizations pervade almost every aspect of our lives. We are born into a society relying on them. We are schooled "in" them to work "for" them. And as we work "for" them, even our "time-off" is conditioned, if not "invaded," by them. We purchase goods produced by them. We rely on their services. We depend on them for our health and protection. And at the same time, we fear them for their destructive potential; for polluting our environment, and endangering our safety, knowing well that to deal with the problems they create, we desperately need them. Although, we usually do not formulate explicit theories about them, we definitely "know" them because they are such an important part of our world.

To start explaining our perspective on organizations, then, let us "follow" Aristotle's method, and try to derive some key features from typical descriptions of what is going on in organizations.

1.2.1 Situation 1: Strategy Formulation

Imagine a group of managers selected to formulate a new strategy, specifying the organization's medium-term goals. To this purpose, they have convened a number of meetings. The objective of the current meeting is to assess the organization's product portfolio in the light of recent market research. As the meeting proceeds, three scenarios projecting different market developments are discussed. For each of these scenarios, a different portfolio of old and new product/market combinations appears as the most promising option. In the discussion, one of the managers in charge of a currently dominant business unit resists bold plans to penetrate new markets by arguing that the organization is weary of yet another transformation. However, during the present meeting she is unable to convince her colleagues. A lot of lobbying before the next formal meeting will be required to nip the plans for innovation in the bud.

Strategy formulation is a task of particular difficulty – choosing the wrong goals obviously has an effect on the organization's chances for survival. At the same time, there is no way to determine with certainty the right strategy – it depends on many uncertain factors. Hence, a large body of literature on strategy formulation and strategic decision making exists, as well as a vast amount of tools supporting it.

Relevant activities in strategy-formulation, such as analyzing the environment, judging the current product portfolio and organizational competencies, generating new ones, or creating commitment are also "social" activities: they are carried out by means of communication. That is, in their effort to formulate a strategy, managers discuss strategic issues, and in this discussion they express their ideas and viewpoints; try to convince each other; negotiate; make compromises; use all kinds of rhetorical tricks; take into account their own hidden agenda's as well as their ideas about the agenda's of others; exercise power; display, react to, (mis)use their emotions as well as the emotions of others; etc. In all, strategic issues are discussed and brought about in "social interaction" – in communication.

Not only are strategic goals defined by means of communication, the discussion about them also depends on and builds upon ideas, viewpoints and choices that are the result of *previous* communication. In social interaction prior to the meeting all participants have acquired ideas about strategic issues, about what others think about such issues and about what others think about how they think about these issues. Such ideas result from communication and create expectations for the current meeting. Put differently, communication about strategic issues depends on earlier communication about them. The same holds for choices resulting from previous social interaction. Choices that have already been made with respect to strategic issues prior to the meeting (e.g., the choice for the current strategy; the choice to evaluate it; or the choice to discuss it in a particular way – using some tool or hiring a consultant) structure the strategic discussion, for they determine, to a degree, the object of discussion and the way it is discussed.

In sum, with reference to the first situation, two relevant aspects can be highlighted.

1. It describes a stage in a more encompassing activity (strategy formulation) in which an organization under conditions of uncertainty commits itself to goals that affect its survival.

Formulating a strategy is an activity aimed at the selection of goals. Dependent on the selected goals and their realization, the organization's chances of survival in its environment will be affected. However, at the time of selection, it is uncertain how the selected goal will affect the organization's survival. This makes the activity of selecting goals both *difficult and risky*.

2. It describes a social activity consisting of "communication referring to communication."
   - By means of social interaction or communication, specific organizational members have been selected to participate in the strategy formulation process.
   - The strategy formulation process itself consists of communication referring to communication. Members of the strategy group express their ideas and

viewpoints, reacting to and trying to convince each other. They negotiate and compromise to safeguard their own position and/or what they perceive as the organization's chances of survival.

- The communication of the managers refers to prior communications and the selections communicated by them. For instance, they refer to goals that were selected earlier or to the three scenarios communicated by the marketing department.
- The resulting strategy "communicates" the selection of particular goals that function as a point of departure for future communication, e.g., discussions between managers about how to realize or how far to deviate from the selected strategic goals.

1.2.2 Situation 2: A Worker in Trouble

Imagine an individual worker, operating a machine as part of producing some product. Every day he receives material and instructions about what and how much he has to produce. However, he has been ill lately, and a backlog has formed. His boss has already been complaining about his not delivering products in time. And, on top of that, the machine has failed him a couple of times this week – and those guys from maintenance haven't done anything about it. How to get out of this mess? Could it be possible to make an appeal to his boss's leniency? Could he persuade his colleagues to help him? Could he perhaps alter the schedule of the maintenance department? What about doing overtime?

For every task in an organization – at any organizational level; at a strategic level, or as in this case, at an operational level – it is relevant to deal with disturbances in order to realize its goal. To deal with disturbances, we regulate. However, regulation is not easy, for most of the time it is impossible to predict when a disturbance will occur and which disturbance will occur. One reason for this is that tasks in organizations depend on the output of many other tasks (for an operational task, the output of other tasks includes: material, work-instructions, maintenance, tools, HR-guidelines, logistics, etc.), and each of these outputs may contain errors. Even if the input for a task is ok, many things can go wrong when performing it.

Given the unpredictability of disturbances it is hard to determine *a priori* which regulatory actions are needed. Therefore, some authors have suggested that, instead of building a set of fixed regulatory actions into a task, it often seems to be a better idea to equip workers with the *potential* to generate regulatory actions, given the disturbing circumstances. But this also introduces uncertainty: since no a priori certain regulatory activity can be given, regulatory actions have to be implemented *per hypothesis*, that is, one has to make a judgment about what will work in particular circumstances and try it out, even though one cannot be sure about its effect. In the example: Is it a good idea to count on leniency? Or will talking to the boss make him angry? Will going to the maintenance department have an effect?

Performing and regulating tasks (be it strategic or operational) is also always socially embedded. It always involves communication about, for instance, output, material, instructions, problems, or regulatory possibilities. And, just as we discussed earlier, this communication includes negotiation; making compromises; exercising power.

Performing tasks is also tied to communication in two other ways. First, everything we do in the context of our job can be seen as conveying some message. If the worker from the example operates the machine slowly, this can be seen as an act of subversion; as an act demonstrating his "illness," or as an act demonstrating his need for assistance. And, if interpreted as conveying some message, it may trigger further interaction. Second, almost every aspect of our job – the products we make, the machines we operate, the department we work for, or the boss we have – results from organizational decisions; i.e., results from communicative processes.

So, one could say that the tasks we perform are socially embedded in at least three ways: they depend on (the outcomes of organizational) communication, their performance and regulation involves communication, and, our (task-related) actions can always be interpreted as communicative acts.

With respect to this second situation, the following aspects can be distinguished.

1. It describes a worker selecting a regulatory activity under conditions of uncertainty.
   - Dealing with the disturbances is a case of operational regulation.
   - It is unclear what effects the selected act of regulation will have, both for the worker (e.g., angering the boss) and the organization's chances of survival (e.g., angry clients withdrawing orders because of late delivery).
2. It describes regulation in organizations as a socially embedded activity.
   - The worker's "task" is specified by means of prior communication.
   - The disturbance he faces can appear as such, because of prior communicatively established "norms."
   - The worker is inclined to regulate because he is "socially expected" to do so. Deviating from this expectation may be socially construed as "negligence," triggering "a conversation with the boss."
   - The activity of regulation itself is a communicative process, e.g., calling maintenance, asking colleagues for help.
   - The selection of regulatory actions co-depends on expectations about future communications, e.g., expectations about the reactions of his colleagues or his boss.

Now that we have described two more or less "typical" organizational situations and highlighted a number of relevant aspects, we may try to "extract" what Aristotle calls "archai" or key features underlying the phenomenon under consideration, in this case: organizations. In our view, two key features can be discerned: (1) their "experimental" and (2) their "social systemic" character. Below, we examine these key features in some more detail.

1.2.3 The First "Arche": The Experimental Character of Organizations

As a first approximation of the experimental character of organizations, it can be said that in organizations, we are continuously forced to make decisions about goals, about how to perform transformation processes, or about regulatory activities, without being certain about their effects.

As every manager knows, there are no rules to determine a priori, with certainty, the "best" goal, process realization, or regulatory activity. If there were – making a decision would be mere "calculation," and our organizational tasks would boil down to machine-like operations. That is, we would just have to follow these rules.

Yet, even though such rules are not available, organizational decision-situations call for action: goals *have* to be set; disturbances *have* to be dealt with – decisions *have* to be made, in spite of their uncertainty. Making such decisions is like conducting an experiment, in which the selection of a particular option can be regarded as a *hypothesis*. That is, in organizations every goal, every process, every regulatory activity, can be said to serve as a hypothesis in an "organizational experiment." By proposing it, we hope that some desired effect will be attained, but, due to the unpredictability of organizational decision situations, we cannot be sure about it. To be more precise, we cannot be sure that the desired effect will actually be realized, that it is as desirable as it was thought to be, or that undesirable side effects will not occur. In this sense, we are involved in a process of constantly devising and revising hypotheses.

Conducting experiments in this way is the first "arche" – capturing the intuition about organizations having to do with "trying to attain goals in an uncertain world."

1.2.4 The Second "Arche": The Social Systemic Character of Organizations

The social systemic characteristic is, in our view, close to many commonly accepted opinions about organizations. Often, organizations are described in terms of social interactions or communication (like we did in the descriptions above). For instance, a definition of organization often encountered in literature is: "a group of people in pursuit of a common goal" (cf. Morgan 1986). The "group of people"-part in this definition does not refer to a mere aggregate. It does not refer to a set of individuals standing isolated from each other. Instead, it requires dynamic, meaningful interaction – communication – between them. Without this communication, the "aggregate" would not become a "group," and the "individuals" would not become the "group's members."

The situations described above indicate that everything we do in the context of organizational tasks, like formulating a strategy or operational regulation, involves communication (e.g., dialogue, discussion, negotiation, coercion, etc).

Moreover, they indicate that this communication evolves against a background of both outcomes of prior communication (e.g., on norms, tasks, procedures, etc. established earlier) and expected future communications (e.g., on expectations about the bosses' reaction). Finally, they suggest that what we do when we perform our tasks can itself be seen as communication.

In this sense, one could say that organizations have a social systemic character. That is, social interaction or communication is central to them and more in particular, these communications are connected. They refer to and depend on each other, forming what might be called a "system of communications."

Thus far, we based our "extraction" of the two "archai" on just two examples. It may be objected that coming up with them, based on these two examples does not really prove anything. Or worse, "discovering" key features in examples one provides oneself, is, to say the least, scientifically somewhat suspicious...

We readily agree. The examples are not meant as a "proof" of the archai's existence. Nor do we want to deny the existence of other possible key features.

What we do hope to achieve by presenting and discussing the two situations is (1) to provide a first indication of the two key features *we* want to discuss in *our* book, and (2) to show that they reflect "common sense" ideas about and experiences with organizations (at least, as provided by the examples).

However, we do think that it is possible to generalize these two key features beyond the situations described above and hold that they are indeed key features of *all* organizations. That is, we believe that it is impossible to try to understand organizations and leave these two features out. Would anyone call "something" without either "arche," without communication referring to communication or without the experimental aspect an organization? We do not think so, and we hope to substantiate this in the rest of this book.

1.2.5 The Relation Between the Social Systemic and the Experimental Character of Organizations

Now, let us go back to Aristotle's method for a moment, and see what we have. To study a phenomenon he proposes to analyze commonly held ideas and opinions about it in order to arrive at its key features. Next, one should take these features and construct a "theory" doing justice to the phenomenon.

In line with this method, we "analyzed" "commonly held ideas about organizations" (as represented by the two descriptions of what is going on in organizations) and came up with two archai: the "experimental" and "social systemic" character of organizations. We can now use these features and relate them into a preliminary description of organizations as social systems conducting experiments with their survival.

Above, we said that the social systemic character of organizations has to do with "communication referring to communication." In organizations, communication has to be about something – it has some topic. In organizations, communication

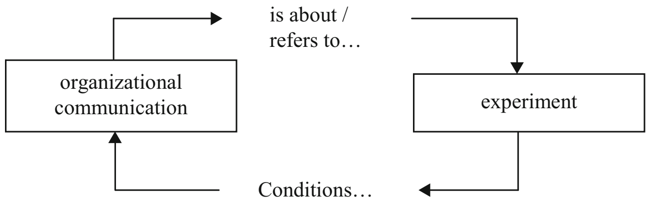


Fig. 1.1 Relation between the "social systemic" and "experimental" character of organizations

may be about all kinds of topics; yesterday's soccer match, a weekend out, the weather, our job, etc. But we believe that the topics of specifically *organizational* communication concern "conducting organizational experiments."

In these experiments goals, processes, and regulatory actions are continuously assessed, selected, implemented, monitored, evaluated, and reselected. To say that the experiment is the topic of organizational communication, then, means that such communication is about valuating, selecting, implementing, monitoring, assessing, or reselecting goals, processes, or regulatory actions and everything related to these activities (e.g., relevant developments in "the environment," methods for selecting goals, techniques for monitoring, etc., can all be a topic of organizational communication).

At the same time, communication about goals, processes, or regulatory actions should result in the selection of specific goals, processes, or regulatory actions, which, in turn, condition further organizational communication. This reasoning leads to a first version of an organizational model relating organizational communication and the experiment (see Fig. 1.1).

In Fig. 1.1 we can see that organizational communication is *about* "objects" involved in the experiment. It is about, for instance, organizational goals, regulatory actions, performing tasks, selecting personnel, etc. Moreover, current communication refers to prior communication about these objects, and in this sense it is *conditioned* by it. Finally, future communication refers to current communication, and in this sense it is conditioned by it. "Organizations as social systems conducting social experiments," now means that in a system consisting of communication referring to communication, decisions are made under conditions of uncertainty about objects that can figure in the experiment, such as goals, processes, means to goals, etc.

1.3 Organizations as Social Systems Conducting Experiments

Thus far, we introduced the two features of organizations we want to discuss in this book and the way in which we think they are related. However, this was just a preliminary introduction. To understand the book's main thesis and how each chapter contributes to it, we need to discuss the two features in some more detail: the topic of this section.

1.3.1 Conducting Experiments

In the book, we consider "conducting experiments" as a key feature of organizations. To elaborate this, we first need to ask ourselves what is at stake in these experiments. Below, we argue this is the organization's "meaningful survival." Secondly, we need to establish what is required to perform experiments; what are the "objects" appearing in them and what does the process of experimenting look like? We argue that the objects and the processes in question have to do with "adapting" and "realizing" goals affecting organizational survival. Finally, we deal with the question what is "experimental" about these adaptation and realization processes in organizations.

1.3.1.1 What is at Stake in the Experiment: Maintaining a Separate and Meaningful Existence

We stated that the experiment consists of selecting goals, regulatory actions, and transformation processes under conditions of uncertainty. Now, it can be asked, what, in the end, is the point of making these selections. What is at stake in the experiment?

To answer this question, we go back to the situations described above. In the first situation a team of managers is in the process of selecting strategic goals. With reference to these goals, it is possible to make a distinction between two types: (1) goals that express a relation between the organization and its environment and (2) goals that can be subsumed under the first type of goals. An example of the first type of goal may be "producing a specific product for some target group." By means of this goal, a specific relation between the organization and its environment is expressed. This goal, then, serves as a starting point for selecting goals of the second type, like production targets, quality norms, or norms for machine maintenance.

With respect to goals expressing a relation between organization and environment, one may ask whether there is something like an overall goal that serves as a desired effect for *all* organizational experiments. In principle, many goals can serve as such "overall goals" (e.g., making a sound profit, satisfying customers, providing employment), as long as they express a relation between the organization and the environment. However, common managerial logic seems to propose "survival" as a promising candidate. What seems to be at stake in the experiment is the "survival" of the organization.

To understand what is meant by "survival" we need to distinguish at least two of its possible meanings. The first, most "empty" or "abstract" sense of survival is that an organization "maintains a separate existence in its environment." Under this interpretation, what is at stake in the experiment is maintaining the separate existence of the organization. Goals, regulatory actions, and transformation processes are selected to realize this particular purpose. The problem with taking this sense of survival as the overall goal of organizational experiments is that not all organizations are geared to survival taken in this way. For instance, an organization may be devoted to realizing a particular goal, e.g., eradicating malaria. Once this goal has been realized – malaria has been wiped out – the organization becomes superfluous and may abolish itself. In a sense, this type of organizations conducts experiments not to survive per se, but to realize a meaningful goal. So, although many organizations do aim at maintaining a separate existence, it cannot be the *overall* goal of the experiment.

The second sense of survival is "maintaining a separate *and* meaningful existence in its environment." Above, we said that mere survival taken in isolation cannot be what is at stake in all organizational experiments. Maintaining a separate existence can only serve as an overall goal if there is a *point* to it: if it can be seen as serving purposes that are considered meaningful. In the case of organizations, their primary "purpose" consists in realizing goals that define their transformation processes (or *primary processes* as they are also called). These primary processes can be viewed as an organization's "*raison d'etre*." If the organization's survival is to be considered meaningful, realizing the goals of these primary processes must be considered meaningful.

There is a relation between organizational survival (as maintaining a separate existence) and particular goals defining an organization's primary activities. In order to select these goals, organizations must survive: survival is presupposed to select goals considered meaningful. In turn, survival is meaningful to the extent that goals are selected and realized that are considered meaningful (see Fig. 1.2).

So, what is at stake in the experiment is not merely "maintaining a separate existence." What seems to be needed is that the organization maintains a meaningful *and* separate existence. If we keep this in mind and discuss the example of the organization that abolishes itself, we can see that as long as the goal of its primary processes, eradicating malaria, is meaningful, this organization aims at survival. Once this goal has been realized, it either abolishes itself or replaces the now meaningless goal by a new meaningful one.[1] So, for the time being, we take this sense of survival as what is at stake in the experiment. This means that goals, regulatory actions, and transformation processes are selected in order to maintain a

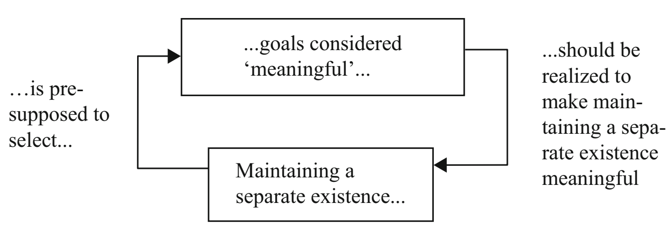


Fig. 1.2 Survival and organizational goals

[1] If no new meaningful goals are selected, 'abolishing the organization' becomes the meaningful goal, requiring the survival of the organization to realize it

separate and meaningful existence: to survive in a particular and meaningful way in its environment.

Given this explanation, it may be asked what goals can be considered as "meaningful." Many answers can be given to this question. Some will say that "making a sound profit" is a meaningful goal, or at least a *sign* of having selected a meaningful goal. Others might say that it is not up to us to say what a meaningful goal is, but to our (potential) clients. In this way, all kinds of organizational goals can be selected as "meaningful" for all kinds of reasons.

However, it is also possible to consider the question what meaningful goals are at a more fundamental level by looking at the organization's contribution to society. The question then becomes, what makes an organization's contribution a meaningful contribution to the larger society it is a part of. To answer this question, we take Aristotle's ethical studies as a source of inspiration. (Aristotle's Ethica Nicomachea 1984a).

In a nutshell, Aristotle said, in his ethics, that the highest purpose for us as human beings is developing our characteristically human capacities to their fullest extent. He called this highest purpose "*eudaimonia*" which can be translated as "living a fulfilled life." And, directly tied to this goal, he argued that it is the goal of "politics" to enable the members of a "polis" to develop their characteristic human capacities to live a fulfilled life.

In modern society, things do not seem to be very different. Living a fulfilled human life still seems to revolve around developing our characteristically human capacities. And it is still a function of modern society to enable its members to develop their characteristic capacities. In fact, following Aristotle's ethical and political reasoning, one might even say that this is the highest purpose of modern society.

Now, organizations play a very important role in modern society: modern societies are dependent upon their contributions (e.g., in terms of providing their specific products or services, employment, etc.), and, at the same time, organizations cause many societal problems (like pollution, inequality, unemployment). Because of this relation, it could be said that organizations always have an effect (in a positive or negative way) on the conditions needed for the development of the members of society. Taking it one step further, one could say that a *contribution* of organizations to modern society is to co-realize the *societal* function of enabling members of society to develop and grow. Although not all organizations may explicitly strive to deliver this contribution, its relevance is gaining ground in both managerial practice and literature (see, for instance, the large number of publications on corporate social responsibility and business ethics).

So what we have now is two senses of survival. The first is the "abstract" or "empty" sense. In this sense, survival means "maintaining a separate existence," irrespective the particular goals that are the organization's *raison d'etre*. As argued, this sense of survival cannot be what is at stake in organizational experiments.

In a second sense, survival means maintaining a separate and meaningful existence. In our view, it is this sense of survival that is at stake when organizations select goals, regulatory actions, and transformation processes. Given this second sense of survival, we distinguish two "modalities."

The first modality can be called a "*poor*" sense of survival. In this poor sense survival means: maintaining a separate meaningful existence by selecting and realizing in *whatever* way, *whatever* goals, considered meaningful for *whatever* reason (e.g., because there is good money in it, because our clients want it, or because we can make our clients to want it). This sense of survival is widely accepted in management literature and management practices.

The second modality is called the "*rich*" sense of survival. In this rich sense, survival means: maintaining a separate and meaningful existence by selecting and realizing goals to contribute to the creation of societal conditions enabling human beings to develop and realize their humanity because this is considered to be the meaningful thing to do. Please note that rich survival does not preclude making profit or other goals that do not aim (directly) at contributing to society. This sense of survival can be found in literature on socially responsible organizations.

In parts I and II of the book and in the rest of this introduction, meaningful survival is at issue (we do not yet distinguish between its poor and rich sense). In these parts we discuss respectively what it means that organizations are social experiments and what the principles enabling experiments with meaningful survival are. In Part III of the book, we discuss the "poor" and "rich" sense of meaningful survival. Moreover, we explore principles underpinning the design of infrastructures supporting rich survival.

1.3.1.2 Conducting the Experiment: Adaptation and Realization for Meaningful Survival

To address the second question concerning the "objects" and "processes" involved in the experiment we have to take a closer look at what is required for meaningful survival.

To stand a chance of surviving in a constantly changing environment, organizations need to do at least two things. First, they have to select and reselect, i.e., *adapt* their goals. Second, they need to *realize* selected goals. By adapting their goals, organizations can stay "in tune" with changes in their environment, increasing their chances of survival. Of course, goal adaptation is not sufficient. Once goals are selected they also need to be realized. By realizing (selected) goals, the organization tries to maintain its existence in its environment, putting the selected goals to the test.

To explain the experiment and the type of "objects" figuring in it, we need to take a closer look at adaptation and realization processes in organizations. To this purpose, we start with an explanation of the transformation processes needed to realize organizational goals.

A "transformation process" is a process turning some input into some output. *Realizing* a transformation process means producing its output. In every organization transformation processes are realized at many different "organizational levels." For instance, at an operational level a transformation process may turn raw materials into finished goods. Or, at a strategic level, a transformation process may turn "observations and reflections about the environment, as well as ideas about the organization's competencies" into a strategy. Transformation processes can also be described at different levels of "organizational detail." For instance, a description of a process at a rather low level of detail might be one in which an organization itself is seen as transforming some environmental input into a product or service. Describing what happens at the level of a business unit, department, or even at the level of individual jobs, involves higher levels of organizational detail.

So, realizing transformation processes means producing their output. If these processes are to deliver an organizational contribution, this cannot be *any* output. The output of a transformation process (i.e., its desired effect or goal) has to be specified. For every transformation process, then, goals have to be set so as to ensure their organizational contribution. Setting goals for a transformation process can be called "*strategic regulation*" of that process.

Realizing transformation processes can be, and often is affected by many different disturbances (either internal or external to the transformation process). That is, many disturbances negatively affect realizing the goal of the process. To deal with these disturbances, common managerial logic has it that one should (1) monitor the transformation processes, and, based on the monitoring results (2) take measures to do something about the disturbances (either reactively or proactively). Dealing with disturbances in this way is often called "*operational regulation*." To give an example, if a machine breaks down (disturbance) a (reactive) regulatory measure might be to repair it. A proactive measure might be to check machines twice a month, so as to prevent problems. Or, to give a less "technical" example, suppose two senior consultants participating in the same project have a different opinion about planning and keep on arguing about it (disturbance). A reactive measure might be to assign one of the consultants to another project. Preventing this kind of trouble may be accomplished by appointing a project-leader.

To realize transformation processes and to regulate them operationally (monitoring them and intervening in them, if necessary), certain "conditions" have to be installed. For instance, one has to recruit and develop skillful, knowledgeable and motivated personnel – "human resources." One also has to make sure that the tasks and responsibilities, needed to realize, monitor and intervene in transformation processes are properly defined and distributed. Moreover, relevant technological means should be made available for the human resources to carry out their allotted tasks. In this book, the term "infrastructural conditions" will be used to refer to this kind of conditions. These conditions are divided into three classes:

1. conditions with respect to the division of work (the organizational structure); these conditions refer to defining and allocating tasks and responsibilities
2. conditions with respect to human resources – referring to recruiting and developing skillful, knowledgeable, motivated personnel
3. conditions pertaining to "technological means" required for realizing transformation processes and regulating them operationally – e.g., machines, or ICT.

Taking care of infrastructural conditions for a transformation process now means: selecting and implementing measures with respect to the three classes of

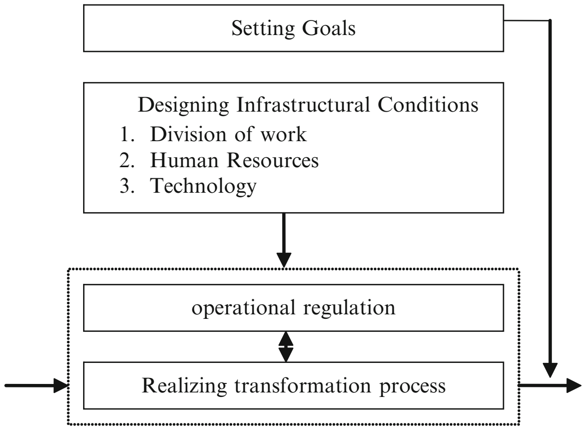


Fig. 1.3 Realizing and regulating transformations in organizations

conditions in such a way that the required division of work, human resources and technological means are available for realizing the transformation process and its operational regulation. Installing these infrastructural conditions is called "regulation by design."

To summarize: organizations perform transformation processes realizing certain goals (by delivering their output), these processes have to be monitored and intervened in when necessary (they have to regulated operationally), and, to realize and regulate transformation processes in organizations, infrastructural conditions from three "classes" are required.

This description entails that four different activities are performed in organizations: (1) realizing transformation processes, (2) regulating transformation processes operationally, (3) setting goals for transformation processes (regulating them "strategically"), and, (4) designing infrastructural conditions for transformation processes and their operational regulation (regulating them "by design") – see also Fig. 1.3.

So, with respect to some transformation process three *regulatory* processes are identified: strategic regulation; regulation by design and operational regulation.

This model is still incomplete. It indicates that an infrastructure should be designed for realizing transformation processes and their operational regulation. However, setting goals and designing infrastructural conditions are themselves processes for which infrastructural conditions are required. Without a division of work, containing tasks needed for goal-setting and design, and without the people or means to perform such tasks, these processes, clearly, can not be accomplished.

In an organization, then, the *process* of designing infrastructural conditions should lead to infrastructural conditions (a particular infrastructure containing conditions with respect to the division of work, human resources and technology) enabling (1) realizing transformation processes, as well as regulating them (2) operationally, (3) by design, and, (4) strategically. As a consequence, infrastructural conditions are needed to (re)design infrastructural conditions. Figure 1.4 shows this reasoning.

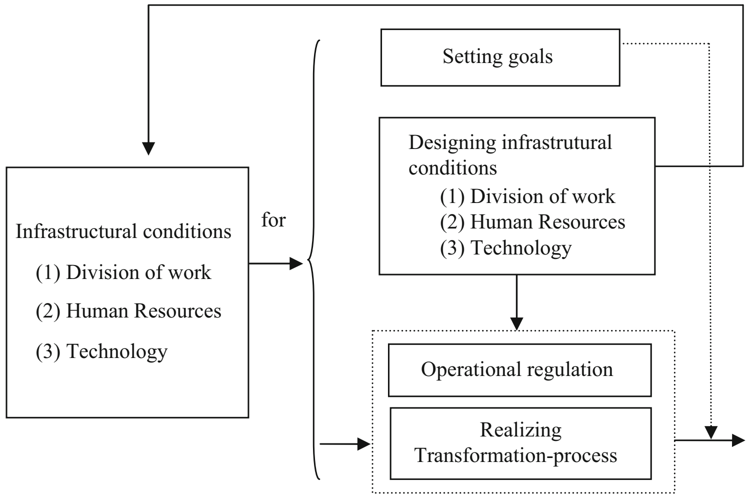


Fig. 1.4 Infrastructure as a condition for regulating and realizing transformations

This description of regulating and realizing organizational processes is quite basic and underlies many reflections and studies about organizations and their management. In particular, it specifies three different forms of regulation and relates them to realizing transformation processes.

The experimental character of organizations now refers to performing the transformation processes and the three types of regulation. Performing these processes and regulating them entails, making selections regarding "objects" such as goals, infrastructures, and regulatory activities (and issues related to them). In the experiment, these selections function as a kind of hypotheses: by selecting a particular goal, a particular infrastructural measure, or a particular regulatory activity, we expect to maintain the organization's meaningful survival.

1.3.1.3 The Experimental Character of Adaptation and Realization for Survival

According to the model developed thus far, we can say that if an organization aims at meaningful survival (either in the poor or the rich sense), it has to:

1. set the "right" goals
2. select and perform transformation processes to realize these goals
3. select and perform operational regulatory activities to deal with disturbances impinging on the transformation processes
4. select and implement infrastructural conditions (from the three classes) to carry out activities 1, 2, and 3.

However, each of these activities is problematic, because, for each of them, there are always many alternatives to choose from. To survive, many different organizational goals can be set. Similarly, to realize such goals, one can always choose from many different transformation processes. And, given a transformation process there is always a number of regulatory activities for dealing with disturbances operationally. Finally, there are always many different configurations of infrastructural conditions to support steps 1, 2 and 3.

In other words, every selection of a goal, process, regulatory action or infrastructural condition is *contingent,* that is – for every selection, there are always a number of alternatives to choose from, and the problem, then, is, that *choices* must be made. We can experience this in organizations every day.

Making a choice would not be problematic if one could somehow determine with certainty the "best" alternative. However, it is characteristic of choosing goals, processes, regulatory actions and infrastructural conditions, and, in fact of any *decision,* that one cannot. It is impossible to determine *with certainty* that a specific selection has the desired result. That is, as every manager knows, there is no a priori rule or procedure leading up to the best alternative. If there was, there would not be the need to decide – just the need to follow the rule. But, unfortunately, such rules do not exist.

At the same time, there are also no a posteriori rules – i.e., experience based rules – to make a selection with *certainty*. Of course, our experience guides us in making educated guesses about the adequacy of a selection, but it does not lead to certainty. After all, every situation in which a selection has to be made is a new one (although it may seem to be similar to the ones previously encountered), calling for new judgment and deliberation.

So, selections are contingent and it is impossible to determine with certainty that they are effective. At the same time, choices *must* be made. To survive, organizations are *forced* to select goals, transformation processes, regulatory actions and infrastructural configurations. It is this combination of contingency (a choice can be made), uncertainty (no certain way to success), and the imperative need to choose (a choice must be made) that makes "organizing" risky – i.e., every selection of a goal, process, regulatory action and infrastructural configuration necessarily contains risk.

In this respect, organizations can be thought of as systems conducting a particular kind of experiment with their survival. At any point in time, organizations are forced to suppose, *per hypothesis*, that a particular set of goals, processes, regulatory actions and infrastructural conditions will contribute to their "overall goals." Cast in the terminology of conducting experiments, one could say:

- *Dependent variable:* meaningful organizational survival
- *Independent variables:* goals, transformation processes, regulatory actions, infrastructural conditions
- *Hypothesis:* Given this set of goals, transformation processes, regulatory actions, and infrastructural conditions, there is reasonable chance to survive meaningfully.

In spite of this structural similarity, it should be clear that the organizational experiments we have in mind are not some kind of "controlled scientific experiment" conducted in, for instance, chemical or psychological laboratories.

What they have in common with this type of experiment is that certain variables are manipulated in order to find out whether an effect occurs, and that this manipulation is governed by conjectures about the relation between certain values of the variables and the (desired) outcome. In the organizational experiment, then, goals, transformation processes, regulatory actions and infrastructural conditions are "manipulated" to increase the chance to survive. In line with what we said before, this experiment is a risky affair, for one cannot be sure whether the chosen "manipulation" will lead to the desired effect (i.e., it is uncertain whether the hypothesis is correct).

Conducting the experiment is also an *ongoing* activity. Part of the experiment is to "test the hypothesis." That is, one has to find out whether the selected goals, processes etc., positively contribute to the organization's chances of survival. Problems with viability may be caused by a wrong selection of one or more independent variables. Such a selection could have been wrong from the beginning, or could turn out to be wrong because of changing circumstances. In either way, a new selection is needed, and the experiment starts again. This makes experimenting in organizations a continuous feedback process of (1) selection, (2) reflection (deliberation and judgment with respect to this selection), and, (3) reselection. It is only by means of such a continuous process that the organization maintains its ability to survive.

To summarize, the organizational experiment is a risky and ongoing activity. Although the effect of the selections is fundamentally uncertain, selections must be made. Organizations *have* to choose goals, transformation processes, regulatory actions and infrastructural conditions. In all, one might say that the experiment is an inescapable aspect of organizations: it can not be "organized away." It is an "arche" and belongs to the *form* of organizations.

1.3.2 Organizations as Social Systems

At the end of Sect. 1.2 we presented a first model relating the experimental character of organizations to their social systemic character. The model expresses the idea that organizational communication is about the "objects" figuring in the experiment. After our elaboration of the experiment, we can now say that organizational communication is about goals, infrastructural conditions, operational regulation and/or realizing transformation processes. These are "focal" objects of organizational communication. In the wake of these objects, other objects may follow. For instance, to establish goals, an organization needs to model its "environment," list relevant developments in this environment, and assess these developments. Both the "focal" objects and the issues related to them are "objects" of organizational communication. By means of organizational communication, these objects are considered, decided upon, monitored, reflected, reconsidered, etc.

In short, organizational communication is directed at experimenting with the adaptation and realization of organizational goals. This is captured in Fig. 1.5 (*top arrow*).

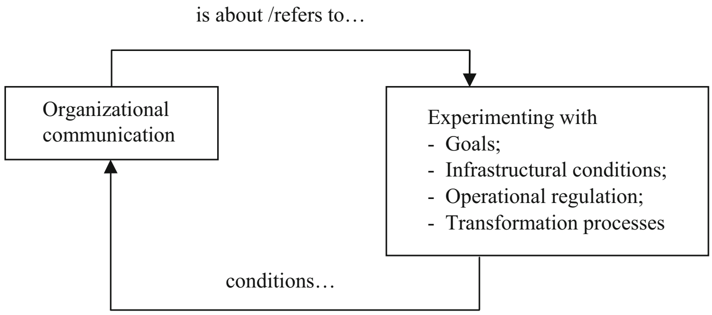


Fig. 1.5 Relation between organizational communication and the experiment

At the same time, all goals, infrastructural conditions, operational regulatory actions and transformation processes that have been selected serve as an "anchor" for further organizational communication. They do so in at least three ways.

First, they serve as an anchor because further communication always *refers* to past and current decisions with regard to objects involved in the experiment. It always starts off from them – either directly or indirectly. Directly, if in communication past and current objects are explicitly monitored, reflected upon, or reselected. The reference to experimental objects in communication can also be indirect, if they serve as a background against which other objects are considered. For instance, the choice for a particular division of work may trigger communication about recruiting human resources. In this case, the particular division of work acts as a background for communication about recruitment.

Second, selected objects serve as an anchor for organizational communication, because the particular infrastructure that has been selected *structures* it. For instance, dependent on the particular division of work we know with whom we are supposed to communicate about what topics. It also conditions, to a large degree, the content of organizational communication, for the type of job I have conditions what I can communicate about. So, obviously, it makes a difference (qua content of and partners in organizational communication) whether I work in a team and have a "rich" job involving full regulatory potential or whether I just perform a short-cycled operational task. A specific division of work can also make a difference for the way in which I communicate with others – e.g., the possibility of exercising power is to some degree conditioned by my hierarchical position in the organization. Similarly, the way I am monitored or rewarded can make a difference for my taking part and feeling involved in organizational communication. And, the available communication tools (e.g., ICT) can condition actual communication, because they condition the (un-) availability and (in-) accessibility of relevant information about the objects of organizational communication, and thus co-determine how they are used as a background for decision making.

There is also third way in which previously selected objects condition further communication. As we said in a previous section, every act can itself be considered as a communication. For instance, the fact that someone is operating a machine very slowly, can be seen as conveying the message that he is unsatisfied with his job, or that he is ill today, or something else. In fact, everything we do in the course of setting goals, designing infrastructural conditions, regulating operationally, or realizing transformation processes can be regarded as a communication, triggering further communication. Because the previously selected objects determine what counts as an organizational "act," they also condition the acts we can interpret as communications. In this sense, interpreting an act as a communication is determined by a particular selection of the objects, appearing in the experiment.

Organizational communication is conditioned by selected goals, transformation processes, infrastructural conditions and operational regulatory actions because (1) all organizational communication directly or indirectly refers to them, (2) the selected infrastructure structures actual communication, or (3) because something we do in the context of the experiment is itself regarded as a communication. This relation is captured by the bottom arrow of Fig. 1.5.

The above reasoning already gives some idea about the relation between the two "archai." It appears that organizational communication is directed at relevant experimental objects and that it is conditioned by particular experimental choices. What still remains unspecified, however, is what the *social systemic* character of organizations entails.

To specify this, we need to go back to the "anchors" discussed above. We argued that organizational communication is about or refers to objects figuring in the experiment. Moreover, we argued that organizational communication conditions organizational communication in three ways: by direct or indirect reference to it, by structuring it or by making it possible that some acts are interpreted as organizational communication.

Now, if we take the "objects" involved in the experiment as the basis of organizational communication, we can redraw Fig. 1.5 as Fig. 1.6.

Based on Fig. 1.6 it is easy to see that organizational communication referring to "objects" figuring in the experiment conditions organizational communication referring to these objects. This "conditioning" relation is important for understanding the social systemic character of organizations. For, by conditioning each other, organizational communications link up with organizational communications. In this way, they form a "system" of related communication. This "system" consists

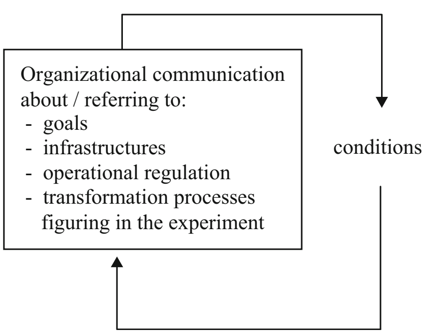


Fig. 1.6 Organizational communication conditioning organizational communication

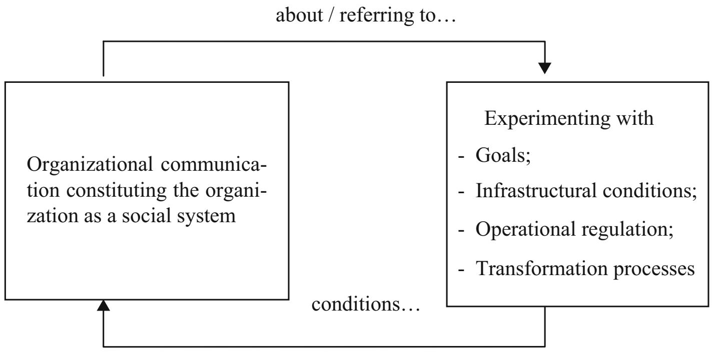


Fig. 1.7 Organizations as social systems conducting experiments with their survival

of linked organizational communications. And because these related organizational communications are characteristically *social* events, the system they constitute is not just any system (e.g., a mechanical, organic, or psychic system) but a "social system." Therefore, by referring to and structuring each other, organizational communications constitute the social system they are a part of.

In Chap. 4 we elaborate this social systemic view of organizations by discussing Luhmann's theory of social systems. For now, it suffices to see that we consider organizations as social systems consisting of interlinked organizational communications; the social systemic "arche" of organizations. The organizational communications constituting this social system are about "objects" central to what we have called experiments with organizational survival; the experimental "arche" of organizations (and about objects in the wake of these focal objects). The result of the elaboration of both "archai" and their relation is now presented in Fig. 1.7. This figure is a first model of "organizations as social systems conducting experiments." This model will be developed further in Part I of the book (Chaps. 2–5).

1.4 Principles Improving Organizational Design

If organizations are social systems conducting experiments with meaningful survival, they should create conditions to conduct these experiments. The question, then, is what these "experimental conditions" are and how they should be created.

To underline the importance of this question, we can ask ourselves what happens if in the ongoing experiment "wrong" selections are made or if selections appear as "wrong," apparently endangering the organization's chances of survival. We all know that such errors need not be catastrophic. As long as they are detected and corrected in time in the course of the experiment, the organization will probably survive. In fact, because of the fundamental uncertainty in the selection process, errors *cannot* be avoided – they are *part* of the experiment, just as their detection and correction.

What would be a real and fundamental threat to the organization's survival is that the experiment itself cannot be conducted – i.e., that the organization would no longer have the possibility to select, reflect upon, and reselect the "independent variables." If wrong goals cannot be adjusted, if a flawed process cannot be redesigned, if the lack of regulatory potential cannot be compensated for, if a detrimental infrastructure cannot be fixed, then, obviously, survival is really problematic.

This means that it is not so much the *specific* goals, processes, infrastructural configurations or regulatory potential that secures or endangers the organizational survival, it is rather the ability to keep on conducting the experiment in which these goals, etc, are generated and corrected that is the *real* driving force behind survival.

In line with our description of organizations presented earlier (see Fig. 1.4), this means that, in order to survive, it is vital to pay attention to the *infrastructural conditions* for conducting the experiment. For this reason, we consider formulating *principles* for designing adequate infrastructural conditions as an important topic in this book. Two classes of design principles are distinguished.

The first class consists of *functional design principles*. These principles state which "functions" should be served by the organization's infrastructure; i.e., they specify the desired effects the organization's infrastructure should be able to produce. Generally speaking, it is the infrastructure's main function to allow the organization to continue to perform its experiments with its own survival. This is what the infrastructure is for. In Part II of the book (Chap. 6), this main function is analyzed to specify the necessary and sufficient functions for conducting experiments.

The second class of design principles consists of *specific design principles*. These are principles for the design of the specific "parts" of the infrastructure. Given the functional design principles, these principles define rules and heuristics for the design of the division of work, human resources systems, and for the organization's technology.

1.4.1 Functional Design Principles

Functional design principles specify what the infrastructure should be able to do to contribute to conducting experiments (i.e., which "functions" it should realize). The organizational model in Sect. 1.3.1 introducing the adaptation and realization of transformation processes already provides such functions. It states, for instance, that the infrastructure should support goal setting and that it should enable the performance of transformation processes realizing the goals set for them.

However, this "functional model" is still rather sketchy. It is, for instance, possible to say that the function "setting goals" always requires another function: monitoring the environment in such a way that current and projected goals can be judged with respect to their fit to environmental demands and opportunities. This is only one example of a function required for "setting goals." One may ask which other functions could be identified for setting goals – and, of course, for the other parts of the functional model of "realizing and regulating organizational transformation processes." So, what is needed is a functional model specifying the necessary and sufficient functions enabling an organization to survive in its environment.

To specify how the infrastructure can contribute to conducting the experiment, a functional model should not only state the relevant functions, but it should also make apparent what their relations are. In particular, it should clarify what the relation is between "setting goals and realizing them" at different organizational levels. Goals may be set, for instance, at a corporate level, at the level of a business unit, at the level of departments, and even at the level of individual jobs. All these goals (and their realizations) should be related, and so, the (recursive) relation between functions at different levels should be dealt with in a functional model of organizations.

In this book, we present Stafford Beer's Viable System Model (Beer, 1979) as a functional model that deals with both problems (see Chap. 6). This model specifies the necessary and sufficient functions needed for "viability," where "viability" is defined as "being able to survive." Moreover, it specifies how these functions should be related at different organizational levels to enable survival. In this way the Viable System Model provides the principles serving as the functional background for designing organizational infrastructures.

1.4.2 Specific Design Principles

The second class of design principles pertains to the design of each of the three parts of the infrastructure. In particular, three types of specific design principles can be formulated. Their general form is:

1. Principles for designing the organizational structure (division of work): Conducting the experiment is enabled by an organizational structure having characteristics a,b,c,. . .
2. Principles for designing human resources management: Conducting the experiment is supported by human resources management having characteristics x,y,z,. . .
3. Principles for designing technology: Conducting the experiment is enabled by technological means having characteristics 1,2,3,. . .

In this book, we only treat principles for designing organizational structures, for in our view these structures function as a point of departure for thinking about the design of systems for human resources management and technology.

To summarize, if the experiment is to contribute to the survival of organizations (to their mere survival as well as to their meaningful contribution to society), an adequate infrastructure should be designed. In this book we will present two classes of principles for designing this infrastructure: (1) functional principles, stating the functions required for survival – i.e., stating what the infrastructure should be able to do to ensure survival, and (2) principles for designing organizational structures in such a way that they can realize the required functions adequately. In the course of their elaboration, we will show that these principles are general – i.e., that they hold for all organizations.

1.5 Conceptual Background

To describe organizations as social systems conducting experiments and to present principles for designing an infrastructure supporting the "social experiment," we use concepts from (organizational) cybernetics, social systems theory, and Aristotle's ethics. In this book, we hope to show that concepts from these traditions – as introduced by their relevant representatives – can be integrated into a framework supporting our perspective on organizations.

To this purpose, we introduce, in each of the following chapters, relevant concepts from an author "belonging" to one of these three traditions and show how these concepts contribute to describing organizations as social experiments (in Part I of the book), to formulating principles for the design of functions and organization structures supporting meaningful survival (Part II), and to formulating principles for the design of organization structures enabling the rich sense of meaningful survival (Part III).

Of course, the relevance of cybernetics, social systems theory and Aristotle's ethics can only be understood in full, after they have been treated in more detail – but based on what we said above, it may already be possible to see why these theories have been chosen as conceptual background.

In essence, *cybernetics* provides concepts for understanding the regulation of any kind of system, and, therefore, it can help to understand and make explicit the regulation of organizations. Understanding regulation is important for our quest for two reasons: (1) the objects, appearing in the experiment, refer to three types of regulation in organizations, and (2) the experiment itself must be regulated. As we discussed, the regulation of the experiment itself, requires infrastructural conditions, and both formulating and judging such conditions is grounded in an understanding of regulating the experiment. A theory presenting concepts to understand regulation, then, is central to both describing the experiment as well as formulating its infrastructural requirements.

Cybernetics plays a central role in our book: general cybernetic insights (based on Ashby's conceptual framework Ashby 1958) are used to describe organizations as experiments, and theories derived from general cybernetics are used to formulate (1) functional requirements for the organization's supportive infrastructure (based on Beer's Viable System Model) as well as (2) specific infrastructural requirements. With respect to these specific requirements: we rely on a cybernetically oriented theory to formulate principles for designing the structure of an organization supporting the experiment (i.c. de Sitter's so-called sociotechnical theory, de Sitter 1994). Because of the central role of cybernetics in understanding organizations, one could say that this book may be read as a specific "introduction to organizational cybernetics."

Based on our discussion of the two organizational "archai" it should not be too difficult to appreciate why *social systems theory* is included in our conceptual apparatus. As discussed, the organizational experiment is a continuous, thoroughly social activity: it is carried out *in* communication; it conditions organizational communication, and all organizational communication refers to "objects" appearing in the experiment. Therefore, when describing organizations as social systems conducting experiments, we need a theory explaining what the social character of organizing is. Social system theory as advanced by Niklas Luhmann does precisely this (Luhmann 1984).

Luhmann's theory is a cybernetically grounded sociological theory explaining what the elements of social systems and their relations are. In his theory, communication is viewed as the hallmark of any social system. Given this general understanding of social systems, his theory also specifies organizations and society as particular types of social systems, explaining how communication is central (in specific ways) to these types of social systems. This makes Luhmann's social systems theory relevant to our purposes in two ways.

Firstly, it provides the conceptual background for understanding the social "arche" and is therefore useful in Part I of the book, where we describe organizations as *social systems* conducting experiments. Based on Luhmann's theory we can understand how conducting the experiment both depends on and conditions communication in organizations.

Secondly, because Luhmann's social system theory discusses both organizations and society as specific types of social systems it provides the conceptual background to understand the relation between organizations and society, which is relevant for our discussion about contributions of organizations to society (as a particular kind of meaningful survival).

Understanding the nature of the contribution of organizations to society is the main reason to include *Aristotle's ethics*. As we briefly introduced above, a relevant "general" goal of organizations is to provide a meaningful contribution to its societal environment. But – one may ask – what is the nature of such a contribution? As we have already indicated, Aristotle's ethics can be a source of inspiration to formulate an answer to this question.

Based on his ethics one could say that such a contribution aims at supporting individual members of society to develop themselves as human beings, so that they can live a "fulfilled life" as Aristotle would say. In the case of organizations, this can be interpreted in two ways. First, organizations can contribute to society by creating conditions for the development of their own members. They can do this by designing infrastructures that allow for job-related development of their members fitting the requirements set by living a fulfilled live. Second, organizations can contribute to society by producing in a "socially responsible" way. From an Aristotelian perspective, this means that both their products and services and the way these products and services are produced contribute to the realization of societal values, increasing the chances for individual members of society to develop their own humanity, i.e., to live a fulfilled human life. Once again, by means of the design of their infrastructure, organizations can increase their potential for socially responsible behaviour. Aristotle's ethics, then, is a relevant source of concepts for formulating principles for designing infrastructural conditions with respect to organization structures improving the organization's capacity for providing a meaningful contribution to society.

1.6 Outline of the Book

The book consists of three parts. In Part I we *describe* organizations as social experiments. In Part II we *formulate principles for designing infrastructural conditions* for conducting experiments. In Part III we discuss poor and rich survival and explore principles underpinning the infrastructural design of rich survival. Individual chapters in each part cover concepts put forward by one author. In these chapters we introduce these concepts and show how they are relevant for either understanding or designing organizations as "social experiments."

At this point it should be noted that we treat these concepts only insofar as they are useful in the elaboration of our own perspective on organizations. This has two consequences. First, the chapters should not be read as an overview of the work the author. In most cases, the author's own work includes (far) more than the concepts we discuss. Second, to make the concepts fit our purposes we sometimes amend and adjust them. The implication, then, is that the concepts as presented in the chapters do not always correspond neatly to the ones of the authors. We will warn the reader when this kind of interpretation occurs.

Below, we present an overview of the different authors, whose concepts we describe, use and amend; their most relevant concepts; and what their relevance is for our perspective.

1.6.1 Part I: The Experimental and Social Arche of Organizations

The aim of Part I of the book is to unfold our perspective on organizations as social systems conducting experiments with their own survival. To describe what is experimental about organizations, and to describe the characteristics of these experiments, we rely on concepts from cybernetic theory, as put forward by Ross Ashby and Heinz von Foerster. To describe the social systemic character of organizations, we use Niklas Luhmann's general theory of social systems and specifications of this theory for organizations. The first part of the book consists of three chapters and an epilogue.

The purpose of Chap. 2 is to arrive at a description of the goals, processes and "objects" involved in the experiment. Above we stated that what is at stake in organizational experiments is "meaningful survival." In order to survive, organizations should perform specific activities, such as adapting and realizing organizational goals. In organizations, adaptation and realization processes, in turn, require the performance of transformation processes and particular regulatory activities, such as goal setting, design, and operational regulation.

To explain what these activities entail we go back to one of the founding fathers of cybernetics: Ross Ashby. In his "An Introduction to Cybernetics" (1958), Ashby conceptualizes precisely the processes and objects central to the experiment. Particularly, he explains what regulation is, what types of regulation can be distinguished, and how regulation, realization, adaptation, and survival can be linked. And, what is more, he specifies principles central to the design of systemic infrastructures. In this way, Ashby's ideas as unfolded in Chap. 2 serve as a foundation for the rest of the book.

In Chap. 3 we delve deeper into the characteristics of organizational experiments, by discussing Heinz von Foerster's elaborations of Ashby's cybernetics (von Foerster 1981). More in particular, we highlight their inescapable risky and ongoing nature. Above, we suggested that in organizations, the selection of goals, infrastructures, regulatory actions, and transformation processes involves contingency, i.e., there are always alternatives. Now, selection would not be a problem, if it were possible to predict with certainty which of the alternatives serves best the general goal of meaningful survival. However, we do not have a priori rules from which we can derive – with logical necessity – what the best selection in a particular situation is. Of course, we may acquire experiential knowledge about the behavior of complex systems enabling predictions of a kind. However, this type of knowledge does not provide certainty either. We argue that this inability to predict with certainty the effects of particular organizational selections, is not something that can be fundamentally "repaired" by introducing more "management science," better "information processing systems," or improved "modeling devices" into organizations. It rather is something about organizations that forbids us to make predictions about their behavior with certainty. To show what this is, we discuss the distinction between trivial machines and non-trivial machines introduced into cybernetics by Heinz von Foerster. Particularly the concept of non-trivial machines is important here. Loosely formulated, non-trivial machines are systems whose behavior is fundamentally unpredictable for external observers. We argue that in organizations we have to deal with non-trivial behavior. Because of this, it is impossible to make predictions with certainty, introducing "experiment" and "risk" into the heart of organizations.

Chapter 4 is devoted to the social systemic character of organizational experiments. To explore it, we introduce parts of the work of the German sociologist Niklas Luhmann. To explain what, in Luhmann's view, social systems are, we discuss central concepts from his general theory of social systems as expounded in his 1984 book "Soziale Systeme." More in particular, we discuss what "communication" is and how communications are a part of and form "systems of communications": social systems.

According to Luhmann, organizations are a particular type of social systems, consisting of a particular type of communications. To explain what this means, we discuss Luhmann's (later) writings on organizations. From this discussion, it appears that in line with the experimental "arche" of organizations, contingency, decision, and risk are central to the communications constituting organizations as a particular type of social systems.

In the Epilogue, we reflect on the findings of Part I. Organizations inescapably face the challenge of experimenting with their own survival. This experimental character of organizing is reflected in the type of social systems organizations are. Both the "experimental" and the "social systemic" character are "archai" of organizations. They are key features that cannot be negated without negating the phenomenon of organizing or organizations altogether. What organizations can do to meet this challenge is to provide conditions enabling their ongoing social experiments. They can do this by means of the design of their infrastructures. So, it becomes the question whether principles can be formulated for the design of infrastructures enabling organizational experiments with survival.

1.6.2 Part II: Designing Organizations as Social Systems Conducting Experiments

In Part II of the book the "hunt" for design principles is on. As indicated above, we distinguish two types of principles: functional and specific design principles. Part II of the book consists of two chapters and an epilogue.

In Chap. 6, functional design principles are at issue. To unfold them, we discuss Stafford Beer's Viable System Model (Beer, 1979). This model builds on Ashby's basic insights into the connection between regulation and survival. More in particular, the Viable System Model specifies the necessary and sufficient conditions systems have to meet in order to be able to survive, i.e., in order to be viable. By specifying these conditions, Beer formulates the functional principles for the design of the infrastructure of viable systems in general and viable organizations in particular. Whatever the organizational infrastructure looks like, if the organization is to be able to continue to conduct its experiments, it should meet the functional principles specified by Beer.

Chapter 7 is devoted to unfolding specific principles geared to the design of organizational structures. The term "organizational structure" refers the division of work in organizations. In Chap. 7, it will be argued that the division of work is of crucial importance for an organization's potential for survival: some designs are quite disruptive, while others are highly beneficial to it. Specific structural design principles allow designers to differentiate between "disruptive" and "beneficial" organizational structures, enabling the diagnosis of organizational structures.

Moreover, they allow designers to build organizational structures increasing an organization's potential for survival. To formulate these specific structural design principles, we discuss Ulbo de Sitter's work. In his book "Synergetisch Produceren" (1994), he builds on insights taken from cybernetics and organization sociology, deriving a set of principles for the design of organizational structures enabling organizational experiments.

In the epilogue to Part II we reflect on the status of the principles for the design of organizational infrastructures discussed in this Part. We will argue that these principles are not a contingent and empirical outcome of experiments with the organization's meaningful survival. They are not hypotheses like the goals, infrastructures or operational regulatory actions that may be rejected in the course of the experiment. They rather are necessary conditions of social experiments with meaningful survival that may be uncovered by a reflection on the conditions of the possibility of organizational experiments with meaningful survival.

1.6.3 Part III: Poor and Rich Survival

Part III of the book deals with poor and rich survival. The aim of this part is to analyze poor and rich survival and to find principles for the design of organizational structures supporting rich survival.

In Chap. 9 we discuss Foucault's description of the "disciplines" (Foucault 1975) – a powerful example of the poor sense of survival. According to Foucault, theses "disciplines" are a specific set of related methods that started to emerge in the eighteenth century, explicitly designed to "make human behavior controllable and useful." In this chapter, we arrive at two conclusions (1) the disciplines are to be regarded as a gloomy epitome of poor survival in organizations with a number of rather troublesome features attached to them, and (2) that they, as all other methods for managing human behavior in organizations are based on the same set of cybernetic and social systemic principles. Given these conclusions, the question becomes whether there are infrastructures based on cybernetic and social systemic principles that support organizational survival in the *rich* sense.

To answer this question, we take Aristotle's Ethics as a source of inspiration. To formulate an alternative for the disciplines, we start in Chap. 10 by making an inventory of desired, instead of disquieting, effects of organizational structures, i.e., effects contributing to rich survival. To find these desired effects, we analyze Aristotle's concept of "*eudaimonia*" which may be translated as "living a fulfilled life."

Based on the results of this analysis, we rethink the notion of organizational "survival" in Chap. 11. As indicated above, survival not only means, maintaining a separate existence in an environment. It also means contributing to this environment to create conditions allowing human beings to live a fulfilled life both inside and outside organizations. Based on this notion of survival, we show that an alternative exists for discipline-like organization structures, thereby finalizing our search for specific design principles supporting organizational experiments with survival.

Chapter 12 is the epilogue to part III and the book.

References

Ackrill, J. L. (1981). *Aristotle the philosopher*. Oxford: Oxford University Press.
Aristotle (1984a). Ethica Nicomachea. In Barnes J. (Ed.), *The complete works of Aristotle*. Princeton: Princeton university press.
Aristotle (1984b). Posterior analytics. In Barnes J. (Ed.), *The complete works of Aristotle*, Princeton: Princeton university press.
Ashby, W. R. (1958). *An introduction to cybernetics*. London: Chapman and Hall.
Beer, S. (1979). *The heart of enterprise*. New York: Wiley.
von Foerster, H. (1981). *Observing systems*. Seaside, CA: Intersystems publications.
Foucault, M. (1975). *Surveiller et Punir: Naissance de la prison*. Paris: Gallimard.
Luhmann, N. (1984). *Soziale Systeme*. Frankfurt am Main: Surkamp.
Morgan, G. (1986). *Images of organization*. London: Sage.
de Sitter, L. U. (1994). *Synergetisch produceren*. Assen: van Gorcum.

Chapter 2
The Experimental Arche: Ashby's Cybernetics

2.1 Introduction

In this chapter we introduce Ashby's cybernetics – a theory about the regulation of all kinds of systems. Ashby's cybernetic theory is fundamental to our perspective on organizations as "social systems conducting experiments" because it provides us with the conceptual tools to describe the "experimental arche" of organizations (see Chap. 1). In particular, Ashby's theory on regulation enables us to arrive at a first description of organizations conducting experiments, making apparent (1) that the objects organizations experiment with – goals, transformation processes, infrastructural parts or operational regulatory activities – are related to three types of (organizational) regulation, and (2) how conducting such experiments should be regulated itself. Moreover, because Ashby's notion of regulation is intimately tied to the survival of systems, his theory can be used to make explicit how conducting organizational experiments is linked to the survival of organizations.

In his work, Ashby formulates the regulatory principles and methods that underpin organizing as an experiment and many currently popular devices supporting this experiment. In his books "An Introduction to Cybernetics" (1958) and "Design for a Brain" (1960) and in many articles, Ashby systematically unfolds basic principles and methods of cybernetics that are still at the core of organizing and all kinds of fashionable management techniques advanced as "new" in contemporary textbooks and journals. Particularly in "An Introduction to Cybernetics" Ashby strives to develop these principles and methods in a way that is as insightful as possible. For this reason, we focus on this book in this chapter (some excursions granted).

In his introduction to cybernetics, Ashby provides a conceptual articulation – rather than an empirical description – of regulation. It is not his aim to empirically describe instances of regulation that are tied to a particular embodiment, a particular place or time. The principles and methods he lays down are truly general and certain. They provide a rigorous treatment of all instances of regulation whether it is regulation in mechanisms, organisms (to which most of his examples refer), organizations, or societies.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_2, © Springer-Verlag Berlin Heidelberg 2010

To explain the organization of this chapter, it is helpful to understand that regulating a system is, in essence, trying to influence its behavior. That is, every "concrete system" we encounter – be it a car, a dog, some transformation process, or an organization – shows particular behavior, and regulating it, in essence, means trying to influence it in such a way that it behaves "properly."

Based on this idea of regulation, at least two requirements can be given: (1) a description of the (proper) behavior of a system, and (2) based on this description, notions about how to influence the system's behavior. These two requirements fit Ashby's description of the aim of cybernetics. He writes that cybernetics hopes to provide "effective methods for the study, and control, of systems that are intrinsically extremely complex" (Ashby 1958, p. 6). The *study* of complex systems (such as organizations) entails arriving at a description of their behavior (in terms of what is labeled a "transformation") and *control* has to do with influencing their (own) behavior – it has to do with regulation (Fig. 2.1).

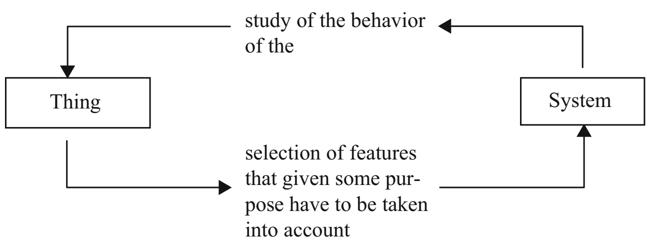


Fig. 2.1 Things and systems

Accordingly, we discuss, in this chapter Ashby's method for the *study* of complex systems (in Sect. 2.2) and effective methods to control (regulate) the behavior of complex systems (in Sect. 2.3). Finally, we show (in Sect. 2.4) the relevance of the presented cybernetic concepts for understanding organizations as "social systems conducting *experiments*."

2.2 Cybernetics: Effective Methods for the Study of Complex Systems

In this section we discuss Ashby's method for the study of complex systems. Above, we gave a first description of regulation as influencing the behavior of some "concrete entity." This entails that we should "model" the behavior of this entity in such a way that we can understand *how* it behaves in the first place, and how this behavior *reacts* to "influences." One could say that (at least) two kinds of influences on behavior (input) can be discerned: "disturbances" – causing the concrete entity to behave "improperly," and "regulatory actions" – causing "proper" behavior (by preventing or dealing with disturbances). A description of the behavior of a system should take such influences (input) into account.

Moreover, in order to regulate a system, a regulator needs to find regularities in the behavior of a system given disturbances and regulatory actions. If no regularities can be found, influencing the system with the hope of making it behave properly is impossible. Accordingly, a description of the behavior of some concrete entity needs to show these regularities.

In all, the description of the behavior of some concrete system we want to study (and regulate) should make apparent how a system behaves, how this behavior "reacts" to disturbances and regulatory actions, and it should show regularity.

In this section, we start with the conceptual devices, needed to describe the behavior of any concrete entity (Ashby's notions of "system" and "transformation"). Next, we move on to regularity and input. Finally we describe Ashby's famous "black box" method – a procedure to describe the behavior of some concrete entity and to determine whether this behavior meets the criteria for regularity.

2.2.1 Describing Behavior: "Systems" and "Transformations"

In its study of complex systems, cybernetics stresses their behavior. As Ashby puts it, "Cybernetics treats, not things but ways of behaving. It does not ask "what is this thing?" but "what does it do?" (Ashby 1958, p. 1).

To study behavior, observers need to model it. To this purpose, Ashby introduces the concept of *system*. He defines a system as a set of variables. "Variables" are features of things that, given some purpose, have to be taken into account (Ashby 1958, p. 40; 1960, p. 15).[1] The system is Ashby's basic conceptual device to study behavior.

For instance, a manager may be interested in the "growth behavior" of her organization. To study this behavior, she may select "net profit" and "number of employees" as features that have to be taken into account. Thus, a system with two variables is defined: "net profit of organization X" and "number of employees of organization X." Given this system, the value of its variables at different moments in time can be determined. Ashby calls the set of values of the variables at a given moment in time, the *state* of the system. So, for instance, the state ($10.000, – , 40) here means that, at a particular moment in time, the variable "net profit" has the value "$10.000" and the variable "number of employees" has the value "40."

By measuring the values of the variables (the states) at different moments in time, it becomes possible to study the behavior of a system. Behavior can be defined as the sequence of states of a system in the course of time.

[1] "Things" are the "concrete entities" or "processes" of which we select variables to describe their behavior. In the remainder of this chapter we will use the term "system" to refer to the set of variables. The "thing" to which these variables refer is indicated with several other terms (e.g. "concrete system," "concrete entities," "real system" or "real machine").

A basic element in the description of behavior is the "*transition*." A transition "is specified by two states and the indication of which changed to which" (1958, p. 10). Ashby calls the *operand* the state that changes to another. For instance, at time = t, an organization's inventory may contain 100 items: 60 items of product X and 40 of product Y. If we define the system to study the behavior of the organization's inventory as "number of items of product X" and "number of items of product Y" then the state of the system at time = t is (60,40). Let this be state A.

All kinds of influences, Ashby calls them "operators" (e.g., client behavior, deliveries of suppliers), act upon this state and may cause it to change. The state into which the operand changes, is called the *transform*. If, for instance, due to all kinds of influences, the inventory at the next moment in time (say, one day later)[2] changes to 50 items of product X and 40 of product Y – i.e., the new state (transform) is (50, 40). Let us call this state, state B.

The transition describes the change from operand to transform. In the example (see Table 2.1), a transition called T1 describes the change from state A at time = t to state B at time = t + 1. The arrow indicates which state changes to which.

Table 2.1 Notation of a transition

T1↓	A (operand, state at time = t)
	B (transform, state at time = t + 1)

In the inventory example, the (vector) notation would be as in Table 2.2

Table 2.2 Notation of transition using vectors

T1↓	(60,40)
	(50,40)

Given these definitions, we can start to study the *behavior* of the system. At time = t + 2 (two days later), the state of the inventory system may become (40, 40) – i.e., 40 items of both products. We call this state C. Now, transition T2 can be written as Table 2.3.

Table 2.3 Example of transition

T2↓	B
	C

At time = t + 3, the state may become (60, 40) – for instance because of the delivery of items of product X. This is exactly the state we called A above, so we can write Table 2.4.

Table 2.4 Transition example (continued)

T3↓	C
	A

[2]For reasons of simplicity and clarity, we abstract (like Ashby) from continuous change and assume Δt to be of constant value.

These results can also be written in the form of a *series* of transitions (see Table 2.5).

Table 2.5 A series of transitions

T1	T2	T3
A	B	C
B	C	A

Now, suppose that an observer studies the behavior of the system for a considerable period of time. After a while he stops his observations and writes down the following series of transitions (Table 2.6).

Table 2.6 A series of transitons (continued)

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
A	B	C	A	D	B	C	A	D	E
B	C	A	D	B	C	A	D	E	A

If the observer supposes that this series adequately describes the behavior of the system, he can abstract from the particular moments in time at which the transitions occurred and rewrite Table 2.6 in a *conditional* form (see Table 2.7).

Table 2.7 A transformation

T↓	A	B	C	D	E
	B or D	C	A	B or E	A

Table 2.7 states that *if* the system at any moment in time is in a particular operand state in the top row (e.g., state A), *then* it will be in a transform state indicated below this operand the next moment in time (e.g., state B or state D). So, A transforms into B or D, B transforms into C, etc. We refer to a description of a system's behavior in terms of a conditional set of transitions by the term *transformation* (cf. Ashby 1958, p. 10).

It appears from the table, that operands A and D have two transforms, B and D, and B and E respectively. Operands B, C and E have only one transform, C, A, and A respectively. From this description, the observer can, for instance, infer that if the system at any moment in time (say, t) starts in state B it will be in state C at the next moment in time (t + 1). If it starts in state C, it will move to state A, and starting from state A it will move to either state B or state D.

It should be noted that it is a decision of the observer or regulator to suppose that some series of transitions describes the behavior "adequately" and can be taken as a point of departure for deriving a conditional transformation. Such decisions are based on experience with or knowledge about the concrete system and always contain uncertainty. This means that the resulting transformation is always a *hypothesis* about how the concrete entity will behave.

To study the behavior of concrete entities, then, observers may define systems. Systems consist of selections of variables that, given the purpose of the observer, have to be taken into account. Using the system as a device, the observer can describe behavior as it occurs over time in terms of transitions. By rewriting these transitions in a (conditional) transformation, the observer can start looking for regular behavior. However, to be able to judge whether regular behavior actually occurs, criteria for regularity are needed.

2.2.2 *Regular Behavior and Input*

Using a set of variables (a system), an observer can describe behavior in terms of a transformation. However, describing behavior is not an end in itself. Behavior is described to enable control. For regulatory purposes, the observer needs to know whether the described behavior is regular or not. In Ashby's words, "Cybernetics deals with all forms of behavior in so far as they are regular, or determinate, or reproducible" (Ashby 1958, p. 1). So, the observer needs criteria to determine the regularity of the described behavior or, as Ashby calls it, whether the behavior is "machine-like" or not. Ashby derives these criteria from two distinctions regarding transformations: open versus closed and single-valued versus multi-valued transformations.

In an open transformation, there is at least one state in the transformation that cannot act as an operand because no transform has been specified for that state. In this case, at least one element in the set of transforms is not an element in the set of operands. Transformation "O" (Table 2.8) is an example of an open transformation.

Table 2.8 An open transformation

O↓	A	B	C	D
	C	D	E	A

In the case of transformation "O," transform "E" is not an element of the set of operands {A, B, C, D} and so, after the occurrence of transform "E" the next transform cannot be determined. Open transformations are unsuited for describing regular behavior, because for some known states, the next state cannot be determined.

A transformation is closed if it is not open. In this case, all its transforms also appear as operands in the transformation. All elements in the set of transforms are an element of the set of operands. Transformation "C" (Table 2.9) is an example of a closed transformation.

Table 2.9 A closed transformation

C↓	A	B	C	D
	A	C	B	A

Closed transformations can be used to determine the next state for all the states that (are known to) occur. According to Ashby, the behavior of a system is regular if it can be described as a closed transformation.

The distinction between single-valued and multi-valued transformations enables Ashby to specify *types* of regularity.

A transformation is single-valued if, "it converts each operand to only one transform" (Ashby 1958, p. 14). See Table 2.10 for an example of a single-valued transformation.

Table 2.10 A Single valued transformation

S↓	A	B	C	D
	A	B	D	A

A transformation is multi-valued if an operand does not always change to the same transform. In the example (see Table 2.11), D may change into A or into C.

Table 2.11 A multi-valued transformation

M↓	A	B	C	D
	A or B	C or D	A	A or C

If the behavior of some system can be described as a closed and single-valued transformation, it is labeled a *determinate machine*. Given the operands of a closed and single-valued transformation, its transforms can be predicted with certainty. Hence, the behavior of a determinate machine can be predicted with certainty.

If the behavior of a system can be described as a closed and multi-valued transformation, it can be modeled as a *Markovian machine*. The behavior of the Markovian machine is described by a "matrix of transition probabilities" (Ashby 1958, p. 225). For a Markovian machine, the occurrence of a transform can only be predicted with a probability. For instance: state A may change into B with a probability of 0.9 and into C with probability 0.1.

Until now, we only paid attention to changes in the values of the variables of the system, defined to describe the behavior of some thing. Using these variables, we can describe change in behavior, but we cannot yet account for the reasons of this change. One of these reasons is that a system is susceptible to "input." That is, there may be variables (that are *no* part of the system that is used to describe the behavior of some thing) influencing the system's behavior. Ashby calls these variables *parameters*. He defines input as the specific value of a parameter at a specific moment in time. Dependent on the input, i.e., the value of a parameter, a system behaves in a particular way.

To describe the influence of parameter-value changes, Ashby introduces the "determinate machine with input." A determinate machine with input is a collection of determinate machines (sharing a set of operands). The value of the parameter is the input. The input specifies which one of the determinate machines determines the behavior (Ashby 1958, p. 44). Table 2.12 is an example of a determinate machine with input.

Table 2.12 A determinate machine with input

I_p↓	80	70	60	50	40
P = 20	70	60	50	40	60
P = 40	70	60	50	40	80

In the example, the operands and transforms represent the number of products in stock at some moment in time. The parameter P represents the order quantity. It can have two states: 20 or 40. Dependent on the state of P, the behavior of the stock differs. If $P = 20$, then the transformation consisting of the first and second row of the table results. The behavior it represents can be interpreted as follows: starting with a stock of 80 items, every time period 10 are sold and when it reaches 40, 20 items are added to the stock. If $P = 40$, then the transformation consists of rows 1 and 3 – with a comparable interpretation.

Normally, a machine with input is given a name containing an index representing its parameter(s) (I_p in the example). In this example, the machine with input contained one parameter with only two values. It is, of course, possible to have many parameters with many different values as input.[3]

In sum, to be considered regular, the behavior of a system must be describable as a closed transformation either with or without parameters (input). An observer searching for regularity first determines whether the behavior of the system meets the criteria of the determinate machine or the machine with input. If the behavior of the system does not meet these criteria, one has to be content with the statistical predictability of the Markovian machine.

To study (and regulate) the behavior of complex concrete entities, a regulator should define a system (a set of variables) and should try to arrive at a transformation, describing its behavior. To account for input, a transformation should contain parameters. Moreover, two criteria have been given to determine whether the behavior, described by the transformation, is regular. Given a system, parameters and criteria, it is, in theory, possible to describe the behavior of systems in terms of a transformation and establish whether they show regularity. However, one may ask how an observer should go about to describe in practice the behavior of a concrete system. What is needed, then, is a *procedure* prescribing the steps an observer has to take to describe the behavior of a system, to assess whether this transformation is regular or not, and to determine the type of regularity.

2.2.3 A Procedure to Describe Behavior and Identify Regularity

At the heart of the procedure for the identification of regular behavior is Ashby's famous "Black Box." According to Ashby everything that can be the object of study and manipulation is a Black Box. The things we experience, manipulate, or study in everyday life, are all Black Boxes. We experience their behavior in terms of relations between input and output without reference to the "inner mechanism" producing their behavior.

[3] In addition to determinate machines with input, there are also Markovian machines with input. These are composed of a set of Markovian machines (sharing the same operands). The state(s) of the parameter(s) determines which one of the Markovian machines determines the behavior of the Markovian machine with input.

As we said before, whenever we encounter something we want to deal with, Ashby proposes that we "model" the behavior of this thing using "a system and a transformation." That is, we (implicitly) define variables to describe the behavior and we describe behavior in terms of changes in the values of these variables. Moreover, if we want to manipulate (or understand) its behavior we also have to define parameters that may influence it. In this way, we describe the "thing's" susceptibility to input (among which "disturbances" and our "regulatory actions"). We may proceed by studying the behavior, given various changes in the parameter values – a study that may lead to conjectures about regularities between input and output. This study results in a final description of both the behavior of some thing and the effect of input on this behavior: the transformation. Based on this transformation, we may try to actually manipulate the behavior by (re) setting the parameter values.

Ashby stresses that while we try to establish conjectures and try to derive the transformation of the thing (the black box) we cannot open it to inspect its inner "mechanism." The only things we can do are manipulating its input (parameter states), observing its behavior (the output – the change of the state(s) of variables of the system). Because the inner mechanism of the Box remains hidden (it stays in the dark, so to speak) it is called black.

To give an example, suppose an observer wants to study and manipulate (regulate) the behavior of company X (some black box). To this purpose, the observer defines a system, say, the annual economic result of company X, to describe its behavior (its output). He also defines as a parameter (input) the particular CEO in charge. After some observation, the observer writes down Table 2.13.

Table 2.13 Example of a machine with input (see text)

$R_p\downarrow$	Result > 0	result $\leq$ 0
P = CEO 1	Result > 0	result > 0
P = CEO 2	Result $\leq$ 0	result $\leq$ 0

Table 2.13 specifies the relation between the parameter (input – i.c. the CEO in charge) and the behavior (output) of the system (the result of the company). This relation meets the requirements of the determinate machine with input. If the observer is able to determine who is in charge of company X (the input), and if the observer has enough "authority" to change the parameter value, (s)he can manipulate its annual economic result (the organization's behavior – the output).

One may ask how the observer has arrived at this transformation, showing regularity in the relation between input and output of the black box. What is the procedure one has to follow? The procedure for finding transformations describing behavior can be summarized as an iteration of five steps.

Step 1. Select a Purpose

To study the behavior of Black Boxes, Ashby suggests that we work from some "main interest that is already given" (Ashby 1958, p. 40). In other words, what is needed is that we work from some purpose. In the example, for instance, the observer may want to understand and manipulate the organization's profitability.

Step 2. Define the System, the Parameters, and the Measurement Interval

Given the purpose, an observer has to select variables that define the system. The values of the variables of the system (its states) are the output of the black box. In the example, the observer chooses to select the annual economic result of company X as the variable describing its output. At the same time, given the purpose, the observer can select relevant parameters. The states of the parameters are the input to the black box. In the example, the observer believes that the CEO in charge influences the annual economic result – hence the parameter "CEO in charge" is selected. This parameter can be in either of two states "CEO 1 is in charge" or "CEO 2 is in charge." Finally, the observer needs to define the points in time at which the behavior is recorded. Suppose that in the example, the point in time at which measurement takes place is December 31 at 24.00 h of each year, beginning in 1980.

Given the specification of the variables to measure the output of the Black Box, the parameters to measure its input, and the points in time measurement takes place, it is possible to record, "at each of a sequence of times the states of the Box's various parts, input and output" (Ashby 1958, p. 88).

Step 3. Record the Behavior of the Black Box in Terms of Input and Output

Given the selection of the variables, parameters, and measurement points, the observer can "manipulate" the input and observe the output of the Black Box. These manipulations and observations are recorded in a protocol. In the protocol, the observer records the values of the input and the output at each of the selected points in time. Thus, the Black Box, "is investigated by the collection of a long protocol, drawn out in time, showing the sequence of input and output states" (Ashby 1958, p. 88).

To illustrate how recording behavior works, we code, in our example, the parameter values "CEO 1 is in charge" as 1 and "CEO 2 is in charge" as 2. The output states, company X's results, are coded in millions of dollars as follows:

A: outcome < -5
B: $-5 \leq$ outcome < 0
C: $0 \leq$ outcome < 5
D: outcome ≥ 5

On the basis of observation, the following protocol is recorded (see Table 2.14).

Table 2.14 Protocol of observed behavior

	'80	'81	'82	'83	'84	'85	'86	'87	'88	'89	'90	'91	'92	'93	'94
CEO	2	1	1	2	2	1	1	2	1	2	1	1	2	2	2
Output	B	C	D	B	A	C	C	B	C	B	D	D	B	B	A
	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09
CEO	1	1	1	2	2	1	1	2	1	2	1	1	2	2	2
Output	C	D	D	B	A	C	D	B	C	B	D	D	B	A	A

Table 2.15 List of states following a particular state (see text)

1,C	→	1,D; 1,C; 2,B; 2,B; 1,D; 1,D; 2,B
1,D	→	2,B,1,D; 2,B; 1,D
		2,B; 2,B; 1,D; 2,B
2,A	→	1,C; 1,C; 1,C; 2,A
2,B	→	1,C; 2,A; 1,C; 1,D
		2,B; 2,A; 2A; 1,C; 1,D; 2,A

Table 2.16 Matrix of transitions describing behavior

$R_p \downarrow$	A	B	C	D
P = CEO 1	C, C, C	C, C, D, C, D	D, C, D, D	D, D, D
P = CEO 2	A	A, B, A, A, A	B, B, B	B, B, B, B, B

Table 2.17 Matrix of transitions describing behavior (simplified)

$R_p \downarrow$	A	B	C	D
P = CEO 1	C	C or D	C or D	D
P = CEO 2	A	A or B	B	B

Step 4: Construct a Conditional Transformation

To construct a conditional transformation, the observer needs to re-code the protocol resulting from step 3, and, if possible, simplify it. To do so, the observer collects, for each occurring combination of input and output state, all the combinations that immediately follow it. This results in Table 2.15.

Using Table 2.15, the observer can arrive at a transition-matrix presented in Table 2.16 and its simplified form given in Table 2.17.

According to Ashby, "all knowledge obtainable from a Black Box (of given input and output) is such as can be obtained by re-coding the protocol; all that, and nothing more" (Ashby 1958, p. 89). So, the protocol, and "nothing more," is the point of departure for step 5.

Step 5. Establish Regularities in the Behavior of the Black Box (and Iterate)

The conditional transformation, resulting from step 4 can now be used to establish the regularities in the behavior, described by the transformation. To do this, the observer compares the resulting transformation to the criteria for regular behavior.

Table 2.18 Description of behavior from example (abstracted version)

R_p (abstracted)↓	result > 0	result ≤ 0
P = CEO 1	result > 0	result > 0
P = CEO 2	result ≤ 0	result ≤ 0

Table 2.19 Description of behavior from example using probabilities

R_p↓	A	B	C	D
P = CEO 1	C (p = 1)	C (p = 0,6) or D (p = 0,4)	C (p = 0,25) or D (p = 0,75)	D (p = 1)
P = CEO 2	A (p = 1)	A (p = 0,8) B (p = 0,2)	B (p = 1)	B (p = 1)

From Table 2.16, it appears that transformation R_p is a closed and multi-valued transformation.

Depending on the kind of regularity found, the observer can take further action. If the transformation is closed and single-valued, the observer can stop: based on this transformation the behavior is perfectly predictable. If the transformation does not meet these criteria, the observer may now follow four different strategies.

Strategy A. Given the protocol, the observer may redefine its results by means of abstraction. For instance, using the protocol, the observer abstracts from the particular state of the result and only takes into account whether it is greater than zero or smaller than or equal to zero (Table 2.18).

The abstracted representation of transformation R_p has the advantage that it meets the criterion of the determinate machine with input. The observer can predict that if CEO 1 is in charge, the result of company X will always become positive and if CEO 2 is in charge, it will always become negative. It has the disadvantage that specificity regarding the values of the output states is lost.

Strategy B. Given the protocol, the observer may also be content with statistical determinacy (i.c. a Markovian machine with input). In this case, the observer uses the protocol to determine the probabilities of the occurrence of a particular outcome (Table 2.19 shows these probabilities which are based on Table 2.16). Using Table 2.19, it is impossible to predict the output of the black box with certainty – but it can be predicted with a particular probability.

Strategy C. It may be that the observer is both unable to find determinacy after applying strategy A and discontent with the statistical determinacy resulting from applying strategy B. In this case, it is possible to go back to step 2 of the procedure. Given the main interest, the observer can add parameters to the selected ones, select new parameters, select new measurement intervals, and/or select new variables. Then the observer proceeds by drawing up a new protocol (step 3), deriving a new conditional transformation (step 4), and analyzing whether this meets the criteria for the determinate or the Markovian machine with input (step 5).

Strategy D. Finally, the observer may go back to step 1 of the procedure and select another "main interest." For instance, the observer shifts his main interest from the organization's performance to the satisfaction of the organization's employees. Now, the observer has to work his way down from step 1 to step 4.

Strategy C and D turn the procedure for finding regular behavior into a *cycle* that can be used to redirect the observer's search for regularity. This cycle allows observers to explore the world by looking for regularities that can be used for the purpose of the regulation of the behavior of black boxes.

In this section we have discussed how an observer can try to arrive at a description of the behavior of some "thing" or black box, capturing the regularities between input and output: the transformation. Before we continue with a discussion about how these transformations can be used to regulate complex systems in the next section, we want to comment on the use of them.

In our daily lives we encounter all kind of "things" we try to deal with. Yet, we seldom explicitly construct a conditional transformation according to Ashby's prescriptions, let alone that we explicitly check whether it passes the criteria for regularity. However, in dealing with the black boxes of our daily lives we always *implicitly* have or build some model of how they behave. Without such a model, interaction with them would simply be impossible. We use such models to predict and understand their behavior and think about ways to interact with them. However, we also know from experience that many black boxes behave unexpectedly – they behave in ways we cannot predict (in fact, we will argue that our models of most black boxes *cannot* be perfect – see Chap. 3). To deal with this we can change our model of a black box – and see if it allows for better prediction and interaction with it.

The use of Ashby's theory, then, is that it (1) describes the necessary elements of these models (variables, parameters, and their relation) – even though these models remain implicit and are in practice imperfect, (2) makes explicit which steps are needed to build and rebuild models of behavior, and (3) gives criteria for "optimal models." This knowledge may be used to diagnose problems in our interaction with systems and it may help us to anticipate problems and prevent them. In all, Ashby's theory provides normative guidelines for the description of behavior that can be used in dealing with problematic situations in our daily lives – for instance in regulating organizations.

To illustrate the importance of Ashby's notions about describing behavior for regulating organizations, imagine some organizational process that is to be managed. Managing a process implies knowledge about its desired result – often given in terms of output of the process (i.e., in terms of variables like "number of products," or "quality of products"). It also implies knowledge about what negatively affects the output (disturbances like unmotivated personnel; machines breaking down, etc.) and about how to prevent or deal with such disturbances (regulatory actions like installing a reward system; or intensifying machine maintenance). In fact, all these situations imply a model containing (1) variables describing the behavior of the process, (2) notions about desired behavior, and (3) parameters (disturbances and regulatory actions) affecting the behavior. In the course of managing organizational processes, then, every manager always constructs and uses (and reconstructs) models of their behavior – more or less explicitly. In other words, building and using models of behavior using variables and parameters is at the heart of management (Conant and Ashby 1970).

2.3 Cybernetics: Effective Methods for the Control of Complex Systems

Ashby's ideas about the *study* of complex systems are a first step to what is the central theme of cybernetics: the "control" of complex systems (Ashby 1958, p. 195). Given the main theme of cybernetics, the concept of "control" should be thoroughly explained. However, it appears that control can be seen in at least two different ways. The first is to define control as some kind of "dealing with" as we did above. In this case both setting targets and realizing them is part of control. This interpretation of control is more or less equal to most common-sense ideas about regulation.

Another way to define control is to equate it with "setting the targets" when dealing with complex systems. Ashby seems to do both. In introducing the topic of cybernetics (1958, pp. 1–4) his idea of control seems to refer to "dealing with" or "regulating" complex systems – while in Chap. 11 of his introduction he reserves control for setting targets. In this chapter he also introduces the term regulation – referring to *realizing* the targets set by control. He explains: Once a controller sets the target, the regulator has to take care of realizing it. A controller is in complete control if, whatever the target, the regulator is always able to realize it. So, the completeness of control depends on the effectiveness of the regulator. Regulation and control are, in this sense, intimately related: they are both necessary components of "dealing with complex systems."

To explain Ashby's ideas about methods for effective "control" (i.e., for "dealing with complex systems") we divide this section into three parts. In the first part we explain Ashby's views on regulation by defining it, presenting different types of regulation, and by discussing how the effectiveness of a regulator can be increased.

Given our discussion of regulation, we introduce and explain three instances of regulation – control, design, and operational regulation – in the second part of this section. These instances of regulation are Ashby's counterparts of "strategic regulation," "regulation by design" and "operational regulation," which we discussed in Chap. 1.

Finally, we discuss the notion of adaptive behavior – which can be seen as a form of "self-regulation."

2.3.1 *Ashby's Views on Regulation: Definition, Types of Regulation and Requisite Variety*

2.3.1.1 Regulation: Ashby's Definition

A good starting point for discussing regulation is Ashby's description of a (good) regulator: "an essential feature of the good regulator is that *it blocks the flow of variety from disturbances to essential variables*" (Ashby 1958, p. 201).

Table 2.20 The regulation table

		R	
		α	ß
D	A	a	b
	B	b	a

Before explaining this description, it is important to notice that Ashby provides a functional description, i.e., he specifies what a good regulator *should do*. The advantage of a functional description is that it abstains from the specific manifestation of a regulator. It states the function of a regulator, regardless of its physical, chemical, psychical, or social appearance.

To explain Ashby's functional description of the regulator, examine Table 2.20.

This table describes the behavior of a machine with input. The system, defined to describe the behavior, consists of one variable which can have two states: a and b. Moreover, two parameters can be discerned: D (with states A and B) and R (with states α and ß). The states of the system, given the values of both parameter D and R, are given in the cells of the table. So, the table indicates, for instance, that, whatever its previous value, if D = A and R = α then the next state of the system is a.[4]

In the table, a and b are possible states of a so-called essential variable. Ashby defines an *essential variable* as a variable that has to be "kept within assigned ("physiological") limits" if an organism is to survive in its environment (Ashby 1958, p. 197). The symbol for the essential variable(s) is: E. Although Ashby's definition stems from biology, it can be extended to other domains, for instance, to the domain of organizations. A more general definition of an essential variable, which also suits the organizational situation, would be that it is a variable that has to be kept within assigned limits to achieve a particular goal (e.g., survival).

From this definition we can see that essential variables always involve an element of selection that is related to a purpose. From the indefinitely many features of some concrete entity, some are selected as "essential" given a particular purpose. In some cases, the selection of essential variables is not subject to choice. For most biological systems, essential variables are fixed. For instance, for human beings both the variables "body temperature" and "blood pressure" are fixed. That is, we

[4] Using the previous transformation-notation, the notation for this system would be:

$T_{D,R}$		a	b
$D = A$	$R = \alpha$	a	a
$D = B$	$R = \alpha$	b	b
$D = A$	$R = \beta$	b	b
$D = B$	$R = \beta$	b	a

In the case of the particular system we use to explain regulation (i.e. the one appearing in Table 2.20, with two parameters, one essential variable, and system states being independent of earlier system states) the notation in the text is more convenient.

cannot choose to leave them out of the set of essential variables, for our survival depends on them. In other cases, the selected essential variables may be subject to deliberate choice. For instance, a firm specializing in information and communication technology may deliberately select the capacity to innovate as a variable essential for its survival.

In the table we also find a D, standing for disturbances. Ashby defines a disturbance as "that which displaces, that which moves a system from one state to another" (Ashby 1958, p. 77). In the example, this means, that D can move the value of the essential variable either from a to b or form b to a. D is, therefore, a parameter of the essential variable. In the table, its possible states are A and B.

The R in the table stands for regulator. The states (or moves) available to regulator R are α and ß. It is the task of the regulator to block the flow of variety from the disturbances to the essential variables by means of acts of *regulation*. R, too, is a parameter of the essential variable.

Now, in order to regulate the behavior of some "thing," a transformation similar to the one described in Table 2.20 has to be available. That is, needed for regulation is a transformation describing the behavior of a system of essential variables associated with the "thing," given the states of D and R.

Ashby uses the term *variety* to refer to the number of different states a variable can have. The symbol for variety is V. For instance, referring to the table, the variety of the essential variable is 2 (the essential variable has two different states, a and b). This can also be written as $V(E) = 2$. Similarly, in the table the varieties of the disturbances V(D) and of the regulator V(R) are 2.

Given these definitions we can explain what it means that a good regulator blocks the flow of variety from disturbances to essential variables. To this purpose, the table should be read as follows. If the disturbances take on a specific state, say A, the regulator selects a move, say α. The combination of A and α has as an outcome of regulation a state of the essential variable, *in casu* a. There are in total 4 combinations of D-R values that all have as an outcome a specific value of the essential variable. In general, a separate variable "outcome" (O) can be introduced to refer to the values of the essential variables that come about after the regulator has selected a move, given some value of D. The set of different states it can have is a subset of the set of possible states of the essential variables and, hence, it follows that $V(O) \leq V(E)$.

Now, let us see whether the regulator can keep the variety of the outcome (V(O)) as low as possible in the face of changing states of D.

First suppose that the regulator has only one move at its disposal, say α. Now, if D is in state A and R is in state α, the outcome of regulation will be that E is in state a. That is, $O = a$. And so, $V(O) = 1$ (we only have state a). If D changes to state B and R remains in state α, the result of regulation will be that E changes to state b. O will be b, and hence, the variety of the outcome increases from 1 to 2 (state a and state b).

In this case, the outcome is determined by the state of the disturbances. The variety of the essential variable after regulation – V(O) – is the same as the variety of the disturbances. The full variety of the disturbances "flows" to the essential variables. In other words: no variety is blocked. Thus, R, with only one move at its disposal, is not a good regulator. In fact, it is the worst possible regulator, for it blocks no variety flowing from the disturbances to the essential variables.

Now, let us suppose that the regulator has moves α *and* ß at its disposal. If D is in state A and R is in state α, the result of regulation will be that E is in state a. The variety of E after regulation – V(O) – is 1 (state a). If D changes to state B and if the regulator wants to keep the variety in the outcome as low as possible it should change the state of R into ß, producing outcome state a. The variety of E after regulation – V(O) –remains 1 (state a). By selecting its moves R can continuously produce outcome state a, whatever the state of D. In fact, whatever the state of D, R is always able to continuously produce either a or b as outcome. This means that the regulator can always keep the variety of the outcome at 1. Apparently, no variety is transmitted from the disturbances to the outcome, or, put differently, the value of the essential variable does not change due to a change in D. R is a good (even a perfect) regulator. R fully blocks the flow of variety from the disturbances to the essential variable.

Thus, we have found a general way to speak about regulators and their effectiveness. Regulators block variety flowing from disturbances to essential variables. The more variety they block, the more effective the regulator. The most effective regulator is the one that blocks all the variety flowing from disturbances to essential variables. In this case, the regulator can fully determine the state of the essential variable.

Given this description of a regulator, we can now also define regulation as the activity performed by the regulator. That is, regulation is "blocking the flow of variety from disturbances to essential variables." If the more general description of essential variables is used (i.e., those variables that must be kept within limits to achieve some goal) then the *purpose* of regulation is that it enables the realization of this goal. If the goal is the survival of some concrete system, then the purpose of regulation is trying to keep the values of its essential variables within the limits needed for survival, in spite of the values of disturbances.

Using the definitions and the table as a point of departure, we can now discuss different types of regulation. Moreover, we can discuss the law of Requisite Variety as the principle underpinning effective regulation.

2.3.1.2 Types of Regulation

Above, we defined regulation as "blocking the flow of variety from disturbances to essential variables." According to Ashby, different types of regulation exist. Ashby distinguishes two of them: passive and active regulation. If we take Table 2.20 as a point of departure and call that which selects regulatory actions in the face of disturbances a regulator, we can explain the difference between passive and active regulation.

In the case of passive regulation there exists a passive block between the disturbances and the essential variables (Ashby 1958, p. 201). This passive block, for instance the shell of a turtle, separates the essential variables from a variety of disturbances. It is characteristic of passive regulation that it does not involve selection. A passive block is part of the regulation table and "does its job" – the regulator does not select a regulatory move dependent on the occurrence of a possibly disturbing event, for the block is given independent of disturbances. Because no selection is involved, the passive "regulator" does not need information about changes in the state of the essential variable or about disturbances causing such changes to perform its regulatory activity.

In the case of active regulation, the regulator needs to select a regulatory move. Dependent on either the occurrence of a change of the state of the essential variable or of a disturbance, the regulator selects the regulatory move to block the flow of variety to the essential variables. Because it has to select a regulatory move, the active regulator either needs information about changes in the state of the essential variable or about the disturbances causing such changes in order to perform its regulatory function. Ashby uses these two sources of information that trigger the selection of regulatory moves to make an additional distinction within the class of active regulation. (Ashby 1958, p. 219; Conant and Ashby 1970, p. 92): error-controlled and cause-controlled regulation

To explain the difference between error and cause-controlled regulation, we refer back to Table 2.20 (repeated in Table 2.21).

This table was to be interpreted as "Given some value of D and R, the essential variable will have some specific value." Ashby calls the mechanism producing the value of the essential variable (given the values of the parameters D and R): T (which stands for Table). Ashby now acknowledges two different ways in which the outcome is determined: error-controlled and cause-controlled regulation. Error-controlled regulation is characterized by the following sequence of events:

At $t = 1$ R has some value
At $t = 2$ D takes on a new value
At $t = 3$ the value of D (at $t = 2$) and the value of R (at $t = 1$) are input for T
At $t = 4$ T determines the value of the essential variables (O) given the values of D and R. This is an undesired value – E is outside its specified norm-values
At $t = 5$ Based on the fact that the changed value is undesired, R selects a new regulatory move. That is, R takes on a new value
At $t = 6$ Based on the new value of R and the value of D, T determines a new value of E, that lies within the specified norm-values

In error-controlled regulation, regulatory moves are selected only *after* the value of an essential variable is changed. As Ashby puts it: this type of regulation "reacts

Table 2.21 The regulation table

		R	
		α	ß
D	A	a	b
	B	b	a

to disaster" (1958, p. 221). The selection of a regulatory move is triggered by a change in the value of the essential variable. Its main goal is to repair the damage.[5]

The sequence of events characterizing cause-controlled regulation is:

At t = 1 R has some value
At t = 2 D takes on a new value
At t = 3 the regulator recognizes the possible threat of D if R's state remains unchanged. R changes its state to one in which the new value of D cannot affect the essential variables
At t = 4 the value of D (at t = 2) and the value of R (at t = 3) are input for T
At t = 5 T determines the value of the essential variables (O) given the values of D and R. This is a desired value; E stays within its norm

In this sequence, R selects a regulatory move *before* the new value of D can change the value of the essential variable. Ashby: "R reacts to threat" (1958, p. 221). In the case of cause-controlled regulation, the selection of a regulatory move is triggered by information about disturbances threatening to displace the state of the essential variable. Before the state of the essential variable actually changes, the regulator selects a move that keeps the state of the essential variable constant. The main goal of cause-controlled regulation is to prevent disaster, given the occurrence of some potentially harmful event.

We now have two basic types of regulation: passive and active regulation. Furthermore, active regulation has been split up into error-controlled and cause-controlled regulation. These distinctions are used to draw regulation-Table 2.22. The table shows the three possible ways of dealing with disturbances discussed thus far: (1) by means of passive blocks, installed with the purpose of blocking one

Table 2.22 Passive and active regulation

		Regulatory potential								
		Passive regulation			Active regulation					
					Error-controlled			Cause controlled		
		Passive Block 1	…	Passive Block j	EC_1	…	EC_k	CC_1	…	CC_m
D	D_1									
	D_2									
	..									
	D_i									
	D_{i+1}				X					
	…							…		
	D_n								X	

[5]It should be noted that if the essential variable is moved beyond its limits, the concrete system to which the essential variables refer should (by definition of the essential variables) cease to exits. If not, the variables are not essential. To deal with this, Ashby suggests that "the states of the essential variables lie on a scale of undesirability" (p. 224) – creating a non-lethal buffer-zone in which regulatory action is required.

or more disturbances (independent of the specific occurrence of some possibly disturbing event) more or less permanently (shaded areas), and (2) by means of several error- and cause-controlled regulatory actions, made available to deal with "residual" disturbances (i.e., the disturbances not blocked by the passive blocks: D_{i+1} to D_n). These regulatory actions have to be selected in the face of actual damage or threat.

To illustrate these different modes of regulation, imagine a medieval knight on a battlefield. One of the essential variables might be "pain," with the norm value "none." In combat, the knight will encounter many opponents with different weapons all potentially threatening this essential variable. To deal with these disturbances, he might wear suitable armor: a passive block. If a sword hits him nevertheless (e.g., somewhere, not covered by the armor), he might withdraw from the fight, treat his wounds and try to recover: an error-controlled regulatory activity, directed at dealing with the pain. A cause-controlled regulatory activity might be to actively parry the attacks of an opponent, with the effect that these attacks cannot harm him.

It is important to note that, here, we only discuss passive and active regulation, as far as they appear in a table that has already been constructed and remains unaltered. *Given* the table, a regulatory activity is performed, and this activity is either selected by a regulator (active regulation) or not (passive regulation). It is, of course, also possible to deal with disturbances by *adding* regulatory potential, i.e., by adding passive blocks, error or cause-controlled regulatory actions to the table – thus altering it. For instance, if the knight picks up a shield, he adds a passive block to "the table." For now, the distinction between passive and active regulation as given suffices – we will discuss adding regulatory potential later in the text.

Cause-controlled and error-controlled regulation differ in three important respects: complexity, effectiveness, and logical dependence.

Firstly, relative to cause-controlled regulation, error-controlled regulation is simple. Unlike cause-controlled regulation, error-controlled regulation only needs the change of the state of the essential variables as a trigger for action. Cause-controlled regulation needs conjectures about more or less stable relations between the occurrence of particular disturbances and changes in the state of the essential variable. Given these conjectures and given a particular disturbance, the cause-controlled regulator can select a specific regulatory move.

Secondly, relative to error-controlled regulation, cause-controlled regulation can be perfect. In order to function, the error-controlled regulator is dependent on the actual occurrence of an error, a displacement from the state of the essential variable to another state. Unlike error-controlled regulation, cause-controlled regulation does not need a change of the state of the essential variable to function as an active block. It depends on a change of the state of the disturbances and is independent of the actual occurrence of an error.

Thirdly, cause-controlled regulation is logically dependent on the occurrence of disturbances and errors. In order to establish conjectures with respect to relations between the occurrence of disturbances and changes in the state of the essential variables, these relations somehow have to be observed. This means that error-controlled regulation can be used in the process of the design of cause-controlled regulation.

To sum up, Ashby distinguishes passive and active regulation. The need for selection of a regulatory move and the need for information to select this move are discriminatory for this distinction. Passive regulation does not require selection. No information is needed. Active regulation involves selection. It can be distinguished into cause and error-controlled regulation dependent on the source of information triggering the selection of the regulatory moves. This source of the information is either a change of the state of D (cause-controlled regulation) or a change of the state of E (error-controlled regulation).

2.3.1.3 Effective Regulation: The Law of Requisite Variety

Ashby's famous Law of Requisite Variety defines the effectiveness of regulators in terms of the relation between the varieties of the disturbances D, the regulator R and the essential variables E. In his discussions of the Law of Requisite Variety, Ashby focuses on active regulation. Because he is only interested in the relation between the varieties of D, R, and E, it does not matter whether these (active) regulators are cause or error-controlled. To get started, we refer back to the regulator of regulation table from the previous section (see Table 2.21).

We already stated that this regulator can completely block the flow of variety from the disturbances to the essential variables. R is a perfect regulator by Ashby's standards.

Of course, regulation is more complex than can be described by means of a simple dichotomy between "no regulation" and "perfect regulation." In the majority of cases, a regulator blocks only a *part* of the flow of variety from the disturbances to the essential variable. In these cases, the regulator is "more or less good." Given this situation, Ashby asks whether any "general statement can be made about R's modes of play and prospects of success" (Ashby 1958, p. 204).

To understand Ashby's answer to this question, it is helpful to examine Table 2.23.

Table 2.23 Regulation table

		R'	
		α	ß
	A	a	b
D	B	b	c
	C	c	a

In this table, we see an essential variable which, before regulation, can take on three values: a, b, and c. V(E) before regulation is 3. Moreover, we see a disturbance that has three possible states: A, B, and C. V(D) is 3. Finally, the regulator has two regulatory moves at its disposal, α and ß. V(R) is 2. In Table 2.23, it is easy to see that the minimal variety of E after regulation (V(O), as we labeled this variety) that can be achieved by R' is 2. This means that R' is a less than perfect regulator.

Given this performance of R', the question becomes how R' can be changed to improve its prospects of success. To answer this question, look at Table 2.24.

Table 2.24 Regulation table with improved regulatory potential

		R' improved		
		α	ß	γ
	A	a	b	c
D	B	b	c	a
	C	c	a	b

In this table, regulatory move γ has been added to the repertoire of R'. By this increase in the variety of the regulatory moves, R' is now able to force down the variety of the essential variable after regulation to 1, which is below the minimum that could be achieved by the old R'. By increasing the variety of the regulator, the variety of the essential variable after regulation (V(O)) is reduced from 2 to 1.

The example of R' is an instance of the Law of Requisite Variety. Ashby formulates this Law as follows.

> "Thus the variety in the outcomes, if minimal, can be decreased further only by a corresponding increase in that of R. [. . .] This is the law of Requisite Variety. To put it more picturesquely: only variety in R can force down the variety due to D; only variety can destroy variety" (Ashby 1958, p. 207).

A more precise formulation of the Law of Requisite Variety is:
IF,

V(D) is given and fixed, and

V(E) before regulation is given and fixed, and

V(E) after regulation (V(O)) is minimal, but greater than one

THEN,

V(E) after regulation (= V(O), the "variety in the outcomes") can only be decreased by increasing V(R).

In essence, this law means that in the given circumstances, only amplifying regulatory potential will help in dealing with disturbances. It specifies, in a general way, how to increase the effectiveness of regulators.

2.3.2 Control, Design and Operational Regulation

Above, we defined regulation as "blocking the flow of variety from disturbances to essential variables." Referring to essential variables as those variables that must be kept within limits to achieve some goal, the *purpose* of regulation is to realize this goal (survival can be seen as one of these goals). An even more general formulation is that the purpose of regulation is to ensure that some concrete system shows "desired behavior."

At the heart of the discussion about regulation was the regulation table. Regulation means using the regulation table in a specific situation. That is, given essential variables and their desired states; given possible occurrences of disturbances, and given the availability of several regulatory actions, a passive block deals with disturbances and/or a regulator selects an error or cause-controlled regulatory action with the purpose of ensuring that the essential variables stay at or reach their desired states. For regulation, then, the regulation table is required.

However, implicit in the discussion thus far, is the question how this table is "constructed." With regard to Ashby's theory, two additional activities can be introduced to deal with this question: *control* and *design*. In this section we discuss these activities and use them to arrive at a method for dealing with complex systems.

2.3.2.1 Control

The first activity is "control." If the purpose of regulation is to ensure that some concrete system reaches some target, then this target should be set. For instance, referring to Table 2.24, it may be that, to stay alive, states a and c are unwanted. If some organism remains in a or c too long, it eventually dies. To stay alive this organism has to maintain the value of the essential variable at b. This is the target set for survival. Ashby calls that which sets the target, the controller. Setting the target (control) entails selecting the essential variables and the desired values of the essential variables.

The controller is in *complete* control if it can select and realize whatever states of the essential variable as its target. To realize the selected targets, the controller needs the regulator. Dependent on the selection of the target by the controller, the regulator performs its task. It "obeys" the controller and attempts to keep the state of the essential variables within the margins of the target specified by the controller by selecting the "right" regulatory moves.

Now, if the controller wants to be in complete control over the outcomes of the essential variables, i.e., if it wants to be able to select and realize whatever of these states as its target, it needs a regulator that is able to produce the selected state irrespective the state of the disturbances. To this purpose, the regulator must be "perfectly" effective. To quote Ashby,

> "[. . .] the fact that R is a perfect regulator gives C complete control over the output in spite of the entrance of disturbing effects by way of D. Thus perfect **regulation** *of the outcome by R makes possible a complete* **control** *over the outcome by C*." (1958, p. 214).

Therefore, regulation and control are not only intimately related in the sense that the regulator "obeys" the controller. They are also related in the sense that the completeness of control over the outcome by the controller depends on the effectiveness of the regulator. The controller's potential to continually produce whatever state(s) of the essential variables as the target, depends on the regulator's potential for regulation. Perfect regulators allow for complete control, imperfect regulators admit only of incomplete control.

At this point it is useful to distinguish between two types of effectiveness of a regulator. A regulator can be said to be "generally effective" if it can realize whatever target the controller sets. A generally effective regulator is a perfect regulator. One can also evaluate the potential of the regulator to realize one or some of the goals a controller can set. The potential of a regulator to realize a specific goal can be labeled as its "specific effectiveness." Of course, general effectiveness entails specific effectiveness. In many (practical) situations, however, one can be content with the regulator's specific effectiveness.

What we have now, is a controller setting the target and a regulator attempting to produce the corresponding states of the essential variables. Dependent on the effectiveness of the regulator, control is more or less complete. If the completeness of control is at stake, it seems crucial to have a regulator that is as perfect as possible. If we are only interested in a specific target (or a set of targets) it seems crucial to have a regulator that is able to realize this (or these) targets. Therefore, the question becomes, How can we achieve this, how can we make a regulator as (generally or specifically) effective as possible? The answer to this question amounts to providing the regulator with the means to deal adequately with relevant disturbances, i.e., to providing the regulator with an effective "regulation table."

To this purpose we have to introduce another activity in addition to that of the controller and that of the regulator. We call this activity *design*. The purpose of design is to select out of the set of possible "mechanisms," a "mechanism" that maximizes the regulator's potential for blocking the flow of variety from the disturbances to the essential variables. That which performs the design activity is called the *designer*.

2.3.2.2 Design

To explain what design entails, we need to underline that the relation between D, R, and E depends on all kinds of physical, chemical, organic, or social processes in the real world. Dependent on these processes (we call them the "mechanism") a particular relation between D, R, and E emerges that can be expressed by a table. When designing for effective regulation, the designer strives for setting the optimal relation between D, R, and E (and between V(D), V(R), and V(E)) as a goal, and selects out of all the possible "mechanisms," the "mechanism" with the highest prospect of success in realizing this goal. A mechanism can be described by a regulation table expressing the relation between possible disturbances, values of essential variables and regulatory activities. And, because we wish to abstract from the actual processes realizing this mechanism, we will, below, refer to the table as the "object" of design. In this way, it can be said that the designer "constructs" or "reconstructs" the regulatory table.

In "An Introduction to Cybernetics," Ashby focuses on explaining in cybernetic terms the *process* of the design of a regulator. In our text, we do not want to go into Ashby's technical discussion of this process. We just want to concentrate on the question how we can enhance the regulator's prospects of success by means of design.

To answer this question, it should be kept in mind that a *regulator* deals with a given set of disturbances in specific actual situations (by means of passive or active regulation) and that the task of a *designer* is to provide a "mechanism"/table by means of which actual regulation *vis-à-vis* a set of disturbances becomes possible. In other words, given some goal (a set of essential variables and their norm values), the designer thinks of disturbances possibly affecting these goals (in general) and constructs a table, enabling a regulator to make sure that these goals are met when confronted with specific disturbances. To construct an effective table, the designer can do two things:

1. Decrease the variety of disturbances (attenuation) and
2. Increase the regulatory potential of a regulator (amplification).

By constructing a table with a decreased variety of disturbances (attenuation), a designer makes the regulator's task less difficult. By amplifying regulatory potential – adding passive blocks, error-controlled or cause-controlled regulatory actions – a designer also increases the prospect of success of a regulator: for the table now includes more ways to deal with disturbances (amplification directly follows from the law of requisite variety: only an increase in R makes it possible to decrease the variety in O).

We now have two related criteria for selecting a "mechanism" to increase the regulator's prospect of success. The selected "mechanism" in the *first* place should decrease as much as possible the variety of the disturbances the regulator must face (attenuation) and in the *second* place it should increase as much as needed the regulatory variety of the regulator (amplification).

At this point it is important to note that design refers to *every* new selection of a mechanism. Phrased in terms of tables: it refers to constructing the table (for the first time) and to all reconstructions of it (i.e., by means of adding new passive blocks or active regulatory activities or by means of removing disturbances).

If we want to design "mechanisms" for effective and efficient regulation, it is relevant to emphasize the logical priority between these criteria. First, we should construct the "mechanism" in such a way that the variety of the disturbances is decreased as much as possible. Otherwise, the regulator would face a larger than necessary variety of disturbances, which is inefficient. Second, we should construct the "mechanism" in such a way that the variety of the regulator is increased as much as needed to block the variety flowing from D to E.

The resulting "mechanism" (expressed by the regulation table) is "generally effective" if it allows for perfect regulation and complete control; it is specifically effective if it allows for realizing a specific target given all the relevant disturbances. Moreover, it is efficient if it adds no regulatory moves beyond the point of either perfect regulation or specific effective regulation. In this way, we are able to design "mechanisms" realizing the functional requirement of effective and efficient regulation and providing (complete) control over the outcomes of the essential variables.

To summarize: based on an overview and understanding of disturbances that may threaten the essential variables (set by control), the designer (1) removes disturbances, as much as possible, and (2) selects passive blocks and error and cause-controlled regulatory activities to deal with them. The designer constructs the regulatory table, to be "used" by a regulator.

Given this explanation of design, we can understand that activities, performed with the purpose of constructing or reconstructing the mechanism (table) are activities of a designer. However, it can also be the case that some cause-controlled regulatory action has as an effect that some disturbance is removed or, even, that regulatory potential is increased. Or, to put it more generally, it can be that an activity has a cause-controlled "aspect" as well as a design "aspect." For instance, the knight from a previous example may kill his opponent in his cause-controlled efforts to prevent damage. The effect is that this opponent is removed from the table as a disturbance. In this case, an activity can be said to have a cause-controlled aspect (it is meant to prevent damage possibly coming from a specific disturbance to the essential variables) and a design aspect (it changes the table, for a specific disturbance is removed from it).

Therefore, it is important to note that the same activity can have a cause-controlled as well as a design "aspect." The cause-controlled aspect entails dealing with an actual threat by preventing it to do real damage. The design aspect entails that the activity changes the table. Given a particular table, it is easy to determine these aspects. In order to have a cause-controlled aspect, the activity should be contained in the table, and it should be selected in the face of a particular, actual threat. Cause-controlled regulation always acts with respect to an actual *specific* event, possibly disturbing the essential variables: it *reacts* to this event by preventing damage. The design aspect manifests itself whenever the table is changed – by removing a disturbance or by adding regulatory potential.

2.3.2.3 A Method to Deal with Complex Systems

The design activity completes the listing of activities needed to formulate a method for the effective control of complex systems. This method consists of the three activities, control, design, and regulation.

The relation between these three activities is – in general – that without a target (essential variables and desired values) set by control, it is impossible to design a mechanism realizing these targets. In particular, it is impossible to (1) identify disturbances and reduce them and (2) device regulatory actions. Given a designed mechanism (expressed by the regulation table) a regulator selects and implements specific regulatory actions in specific circumstances to deal with specific disturbances. So, without the designed regulatory actions, a regulator cannot deal with disturbances. In other words: targets are needed for design, design is needed for regulation and actual regulatory actions are needed to deal with specific disturbances in specific circumstances to reach the targets set by control.

Using the specifics of and the relations between control, design, and regulation, we can now specify a method supporting the effective control of the behavior of complex systems. This method consists of the following three steps in order of priority.

Step 1: Control

Control sets the target, i.e., it specifies what the essential variables (E) are and which states of E are desired states. Control is a *sine qua non* for both design and regulation.

Step 2: Design

By means of design, the designer constructs a mechanism embodying a more or less effective and efficient relation between D, R and E that can be expressed in a regulation table. In constructing this mechanism, the designer should

1. First decrease the variety of disturbances as much as is possible (attenuation)
2. Second, increase the variety of the regulatory actions to the level needed to block the remaining variety of the disturbances (amplification – by installing passive blocks or providing active regulatory options).

Step 3: Regulation

In the case of passive regulation, disturbances are automatically blocked by the regulator. In the case of active regulation, the regulator *selects* the regulatory actions needed to block as much as possible the flow of variety from the disturbances to the essential variable(s) and *implements* them in order to keep the states of the essential variable within the limits specified by control. Given the design of the "mechanism," the regulator more or less effectively blocks the flow of variety from disturbances to essential variables.

Unfortunately, the term regulation has now become ambiguous – it can refer to "dealing with disturbances, given a regulation table" (this is the interpretation in step 3 of the method above), and it can refer to the more general notion of "influencing the behavior of something" – an interpretation of regulation we used earlier in Chap. 1 and in the introduction of this chapter. We resolve this ambiguity by sticking to the latter interpretation of regulation, and re-label the first as "operational regulation". So, step 3 should read "operational regulation."

The three-step method for regulating complex systems is at the heart of regulation in organizations. And, because organizational regulation and continuously conducting experiments are intimately related, the method also underpins the organizational experiment. In the sections below, we therefore want to explain how this method underlies regulation and "organizational experimentation."

However, to understand how it underlies regulation and experimentation, it may be useful to start with an illustration and elaboration of the method's three different steps. Below, we will first provide this illustration. In the sections that follow we discuss how the method underlies organizational regulation en experimentation.

2.3.2.4 An Illustration of Ashby's Method: Problem-Solving as a Series of "Control," "Design" and "Operational Regulation" – Activities

To illustrate and elaborate Ashby's three-step method it may be useful to think of the method as a general model for solving problems. The association between Ashby's method and problem solving is quite natural. A problematic situation may be considered as a situation in which relevant variables do not meet their associated norm-values. Solving problems, then, is, essentially, ensuring that relevant variables regain their norm-values.

To be a bit more precise, the following steps offer a "cybernetic" (i.e., based on Ashby's method) model for dealing with problems. Given some (more or less articulated) problematic situation one should:

1. Define a system whose behavior is associated with the problematic situation
2. Define norms for the variables of the system
3. Redefine the problematic situation in terms of systemic variables being outside their norm-values
4. Determine disturbances – events having a negative impact on the systemic variables (That is: determine parameters, whose values have a negative influence on the variables)
5. Try to eliminate disturbances (attenuation) if possible
6. If the problematic situation is not resolved; define and install, for the remaining disturbances, regulatory potential passive blocks, error- and cause controlled regulation (amplification)
7. If the problem is not resolved by a passive block, perform error- or cause controlled regulatory activities
8. Evaluate the effect of the regulation on the variables of the system, and given the result either decide that the problem is solved or go back to previous steps.

This sequence of steps incorporates the three different steps from Ashby's method. Steps 1 and 2 cover the control activity; steps 4, 5, and 6 are design activities (they construct the regulation table by attenuation of disturbances and by amplification of regulation). Step 7 covers operational regulation: the designed table is used to select and implement active regulatory activities.

Let us see how this sequence works in practice. Suppose that someone owns a lawn and wants to keep it as smooth as possible. That is, the lawn should be without holes or bumps, and, so, every mole or dog is considered as a threat and may call for action. In this example, an *actual* problem occurs when the variables are outside their assigned limits (i.e., there are holes or bumps). A *virtual* problem exists when no holes or bumps are currently present, but may be expected (due to the possible entry of moles and dogs). To deal with a problem (actual or virtual) a system should be defined (step 1). The system in this case should contain variables by means of which "smoothness of the lawn" can be expressed. For instance, the variables "number of holes" and "number of bumps" can be selected. To represent smoothness, the norm for both variables can be set to "0" (step 2). A problematic situation can now be defined (step 3) as the situation in which the value of one of these variables exceeds 0. Possible disturbances (step 4) may be (in this case) moles or dogs. Forcing all dog-owners in the neighborhood to move to another city (and take their dogs with them) or killing dogs are (rather far-fetched) examples of attenuation, (step 5) for they remove the dogs as disturbances. One may also define and implement (step 6) passive blocks that automatically block the influence of disturbances on the essential variables – e.g., building a fence around the lawn is an example of a passive block automatically taking care of the dogs as disturbances. To actively deal with the (possible) effect of moles on the lawn, error- or cause-controlled activities can be defined (step 6). A cause-controlled action may be to place ultrasonic anti-mole equipment every time moles are noticed in neighboring lawns. An error-controlled action is to flatten the surface after a molehill was produced. Trying to kill the mole (already present in the garden) by placing a trap is also a cause-controlled activity – for it is meant to prevent (further) damage. Actually selecting and implementing one or more of these actions, given specific circumstances is step 7. For instance, it may be that molehills are spotted in the neighborhood. Given this information, it may be decided to implement the ultrasonic anti-mole equipment. Finally, the result of this action may be monitored and further action may be required (step 8). It may be, for instance, that the equipment does not work at all and that several molehills appear. In this case new actions are needed (step 7) or may even have to be designed (steps 6). The last step allows for iteration. Going back to previous steps can mean that other operational regulatory actions may be chosen (step 7) or defined (step 6). It can also mean that other disturbances; ways of attenuating them; different norms; or even different systems (sets of variables) may be selected. Of course, going back to earlier steps is also allowed in other steps of the sequence.[6]

The last step suggests that if a problem is solved (the desired values of the relevant variables are within their specified limits) – the sequence stops. This may not be true for all problems. Dealing with some problems may require a more continuous process of "problem-solving" in which (instead of being solved once and for all) periods of "stability" may be discerned. That is, periods of time in which the relevant variables have their desired values. These periods may, however, be disrupted, after which by means of the problem-solving activities a new period of stability is established.

As we said above, the sequence of steps, describing problem-solving incorporates control, design and operational regulation. And, so, given this "translation," we can refer to problem-solving as a (continuous) process involving control, design and operational regulation. In the course of dealing with problems, iterations between these three types of activities are possible. This is indicated in Fig. 2.2, in which a sequence of control, design and operational regulation activities regarding some problem is given.

On the vertical axis, the three types of activities are stated. A horizontal line in the graph at the level of some type of activity indicates that, for some period of time, only activities of that type are carried out. A vertical line indicates a shift to another type of activity. To deal with a problem one has to start (implicitly or explicitly) with a formulation of the problem (i.e., a definition of the problem in terms of essential variables being outside their desired values). This "control period" is indicated in the figure by the first horizontal line on the left-hand side (under 1).

[6] Evaluation can be regarded as a continuous process, monitoring all the steps. It is also needed at as a separate last step.

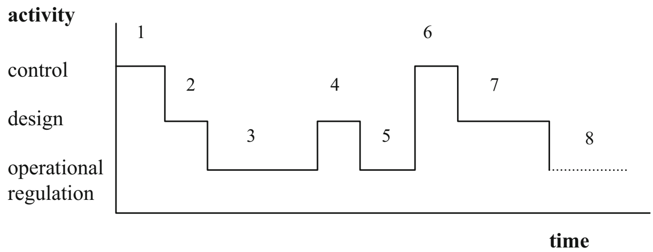


Fig. 2.2 Control, design and regulation activities associated with problem-solving

After control, design takes over and by means of attenuation and/or amplification the regulation table is constructed (the horizontal line under 2). Next, during a "period of operational regulation" (the third horizontal line) one tries to deal with the problem – by means of implementing cause and error-controlled regulatory activities.

In dealing with most problems, these three activities will probably be performed in this sequence (although one may return to control after design to change variables and/or norms). In the rest of the process of dealing with a problem, the three types of activities may follow each other in many ways. One might say that each problem has its own specific history of control, design and operational regulation activities – in terms of how control, design and operational regulation will follow each other, and, of course, in terms of the specific control, design and operational regulation activities needed to deal with the problem.

2.3.2.5 Ashby's Method Underlies Regulation in and of Organizations

To see how Ashby's method underlies "regulation in organizations," we may refer back to Chap. 1, where we introduced organizational regulation in terms of "making sure that organizational transformation processes are run correctly." To do so, goals have to be set for them; infrastructural conditions (from three classes: division of work; human resources and technology) have to be designed for them, so as to ensure that these processes realize their goals, and these processes have to be monitored in order to avoid and repair problems. It is not difficult to see that this interpretation of regulation in organizations is, in fact, a straightforward translation of Ashby's method: it describes dealing with organizational transformation processes in terms of control (setting goals), design (providing infrastructural conditions – from the three different classes) and operational regulation (monitoring of and intervening in organizational transformation processes). Figure 2.3 represents this translation.

In fact, many of the models available in management literature and practice are instantiations of this interpretation of Ashby's method. One well-known model (instantiating these ideas) is the management model of Anthony (1965), who acknowledges three levels of management: strategic; tactical and operational

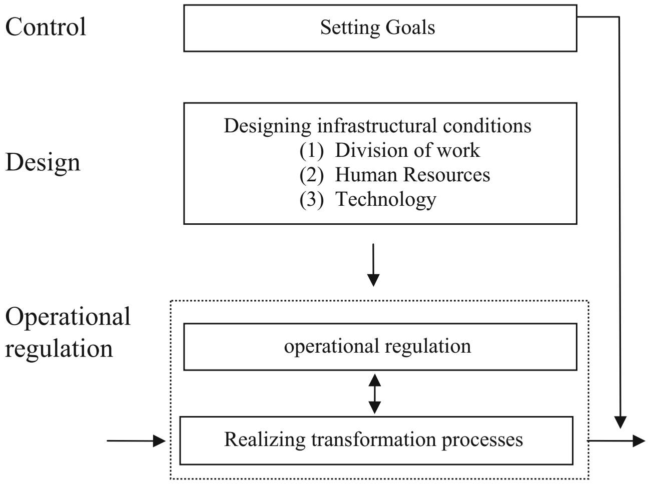


Fig. 2.3 Regulation in organizations based on Ashby's theory

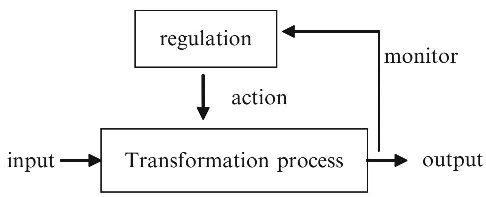


Fig. 2.4 The "control cycle"

management – roughly corresponding with control, design and regulation. Another model, often used in management literature is the "control-cycle" (see Fig. 2.4). In this cycle goals are set for the output of the transformation process. Given these goals, the output is monitored (the current state of the output is compared to the desired state; the goal). If the gap between the two is problematic, actions are taken to intervene in the transformation process (or with regard to the input) in order to make sure that the output will, at some later moment, reach the goals set for it.

Many versions of this model exist in management literature. One might argue that it is not difficult to find examples of Ashby's ideas with regard to what we would call management. It is rather the other way round, it is difficult to imagine something we call management and *not* to find something in it that has been discussed by Ashby in a general and comprehensive way.

Therefore, instead of adding more examples, we would like to stress that underlying all these different models and interpretations of management in organizations is a perfectly general model which can be described in cybernetic terms. This general model describes regulation as the *function* in organizations performing control, design and operational regulation activities with regard to some transformation process.

In mapping Ashby's method onto organizational regulation, the translation of control and operational regulation seems relatively straightforward. However, the translation of design – particularly its result (a "mechanism") - needs some further elaboration: what is it that is designed in the context of organizational regulation?

In Ashby's method, design is said to "construct a mechanism that can be expressed by a regulation table." And, in our explanation of design, we abstracted from the possible physical and social embodiment of this mechanism. However, if we translate Ashby's design to dealing with organizational transformation processes, it becomes relevant to discuss this mechanism in more detail.

As we discussed in Chap. 1, designing "infrastructural conditions" for some transformation process should result in "an organizational structure (i.e., a network of tasks and responsibilities), relating human resources and technological means in such a way that (1) the transformation process can be realized and (2) the transformation process can be regulated operationally – i.e., monitored and intervened in."

To give an example, if a goal of a particular transformation process is to produce a wooden table, design should think of different tasks, relevant for producing a wooden table and relate them into a "network of tasks." Such tasks may be "sawing"; "drilling," "assembling," or "painting." These tasks may be defined and related in sequence and/or in parallel in such a way that, if performed, wooden tables are actually produced. Once tasks are defined, they should be assigned to human resources. For instance, each task may be performed by a different person. Moreover, these human resources should know how to perform their tasks and they should be motivated to do so. At the same time different tools and machines are needed to realize these tasks. In this simple example, a description is given of "a mechanism" needed to *realize* transformation processes: a network of tasks relating knowledgeable and motivated human resources and technological means.

To deal with disturbances influencing the realization of a transformation process, a similar mechanism is needed. That is, to regulate organizational transformation processes, regulatory tasks should be defined and related – e.g., "process-monitoring" or "maintenance and repair." And, similarly, such tasks should be assigned to knowledgeable and motivated personnel and they should be provided with the proper technological means (e.g., tools to maintain or repair the equipment needed for sawing, drilling, etc.).

In short, the "mechanism" that is designed in the context of realizing and (operationally) regulating organizational transformation processes consists of (1) a particular division of work (organizational structure) relating (2) competent and motivated human resources, and (3) technological means (including tools and machinery). This type of mechanism was referred to as an *organizational infrastructure* (with respect to realizing and operationally regulating a transformation process).

Just like Ashby prescribes in his method, designing an organizational infrastructure should start with attenuating disturbances as much as possible, and proceed by amplifying regulatory activities (i.e., installing passive blocks as well as error and cause controlled regulatory actions). Both attenuation and amplification can come about by (changing) a particular division of work; a particular way to manage human resources and a particular configuration of technological means.

2.3.3 *Adaptive Behavior*

In the previous sections, we discussed Ashby's regulatory method and translated it into a description of organizational regulation. In this discussion we introduced control, design, and operational regulation with respect to some operational transformation. Although this treatment already suggested a tight relation between the regulatory activities themselves, and between the regulatory activities and the operational transformation, we have not yet discussed them as functional parts of one concrete system – i.e., one concrete system embodying a transformation process and capable of performing all three regulatory activities *itself*.

However, if one is interested in the question how concrete systems manage of regulating their own behavior in order to survive (like organisms or many organizations), one *should* relate the regulatory and operational functions into a concrete system. Ashby's theory provides several concepts such as self-regulation, adaptation, and ultra-stability to deal with this issue. In fact, one could say that it is by means of these concepts that Ashby's cybernetics delivers its most relevant contribution: it shows how adaptive behavior depends on a particular relation of regulatory and operational functions. Or, phrased in more technical terms: it shows how a concrete ("ultra-stable") system is able to adapt (and survive) due to its "circular organization." In this section, we explain how these cybernetic notions can be used to describe organizations as concrete systems, capable of regulating their own behavior in order to adapt and survive. We treat this topic in two parts: (1) adaptation and self-regulation, and (2) organizations as adaptive self-regulatory systems.

2.3.3.1 Adaptation and Self-Regulation

To explain both adaptation and self-regulation, it is useful to consider a concrete entity consisting of an operational part (producing behavior related to its survival) and a regulatory part aimed at regulating the operational part.

Now, any concrete system, producing operational behavior and capable of regulating its own behavior by means of one or more regulatory activities (control, design, operational regulation) *by itself* is capable of self-regulation. Given this description of self-regulation, it is possible to distinguish several forms of self-regulation based on the kind of regulatory activity the system persforms itself. A concrete system might, for instance, be capable of performing only operational regulatory activities and "regulate" its behavior, given particular goals and some "mechanism" – e.g., a thermostat. At the other end of this self-regulatory dimension are concrete systems performing all three regulatory activities by themselves – such as organizations. One might call these latter concrete systems capable of "complete" self-regulation. Below, we will discuss adaptation based on "complete" self-regulation.

Complete self-regulatory concrete systems can deal with environmental changes that may threaten their essential variables (i.e., with disturbances) in two general ways: they can employ their *existing* regulatory potential or they can *change* their regulatory potential and use it to counter disturbances. Usually, dealing with disturbances by means of changing regulatory potential (and using it) is associated with the concept of adaptation.

In line with the regulatory method discussed in the previous section, the required changes for adaptation can come about in two ways: it is possible to adapt by means of control or by means of design. Adaptation by means of control entails changing goals (i.e., altering the set of essential variables and/or changing their associated norm-values). Adaptation by means of design implies altering the "mechanism" in such a way that it includes:

1. Improved ways to perform the operational transformation. The goal of these changes is to attenuate, i.e., to reduce the probability of errors due to the operational transformation.
2. Improved ways to regulate (either by means of control, design or operational regulation). These may include:
3. Improved ways to control (i.e., change the set of goals)
4. Improved ways to attenuate environmental disturbances. Equipped with these new means to attenuate, the probability of the occurrence of disturbances in the environment of the concrete system may be reduced;
5. Improved ways to amplify operational regulatory potential

Suppose that this concrete system is able to perform by itself control, design and operational regulatory activities *successfully* and *continuously*. Performing these activities *successfully* means that goals are realized (essential variables stay within their limits; or regain acceptable values). Performing them *continuously* implies that it is possible to keep on (1) realizing the goals that are set, and/or (2) adjusting these goals. Successfully and continuously performing all regulatory activities, then, implies the possibility to adapt. It entails being able to set and realize goals effectively, over and over again.

Ashby discusses adaptive behavior and self-regulation, using the concept of "ultra-stability." In essence, an ultra-stable system is a self-regulatory concrete system that, due to its "circular" organization displays adaptive behavior (cf. Ashby 1960). In this text, we will not treat Ashby's ultra-stable system, but, instead, present our own interpretation of it. Our version deviates in that it (1) has a "circular organization" based on the relation between control, design and operational regulation, and (2) is directly tied to (adaptation and self-regulation of) organizations. This is the topic of the next section.

2.3.3.2 Organizations as Adaptive Self-Regulatory Systems

Successful, continuous self-regulation in organizations means that an organization is able to select and reselect relevant goals, "proper" transformation processes and infrastructural conditions to realize these goals. Moreover, it can perform adequate operational regulatory activities vis-à-vis the selected transformation processes. An organization's self-regulatory capacity enables it to adapt and survive.

In Chap. 1, we stated that regulation in organizations is faced with a fundamental uncertainty – it is impossible to be sure about the contribution of the regulatory selections to survival. At the same time, it is impossible to foresee all environmental changes and, hence, regulation in organizations entails, fundamentally, the possibility to *revise* selections. Or, put in the terminology of Chap. 1, it must be possible to perform the (regulatory) experiment continuously. In fact, one may describe regulation in organizations as a particular (ongoing) sequence of control, design and operational regulation activities – just as we described in the section tying Ashby's method to solving problems.

Based on Ashby's notions, experimenting in organizations can now be described as continuous self-regulation (in terms of control, design and operational regulation), aimed at organizational survival. Now, *actual* organizational control, design and operational regulatory activities are realized by a particular organizational infrastructure; i.e., by human resources operating in a network of tasks and responsibilities, using a particular set of technological tools. So, a requirement for continuous self-regulation is a particular set of infrastructural conditions – see also Chap. 1. In Fig. 2.5 the relation between a particular set of infrastructural conditions and self-regulation is given. This particular infrastructure should be part of the concrete organizational system – for without it C, D, and OR-activities are impossible. In fact: adaptive behavior crucially depends on this infrastructure.

On closer inspection, it appears that self-regulation based adaptation in organizations involves five feedback-loops – each tied to a particular type of regulation (see also figs 2.6 and 2.7).

Figure 2.6 depicts the three feedback-loops directly related to realizing operational transformation processes. The first feedback-loop involves operational regulation

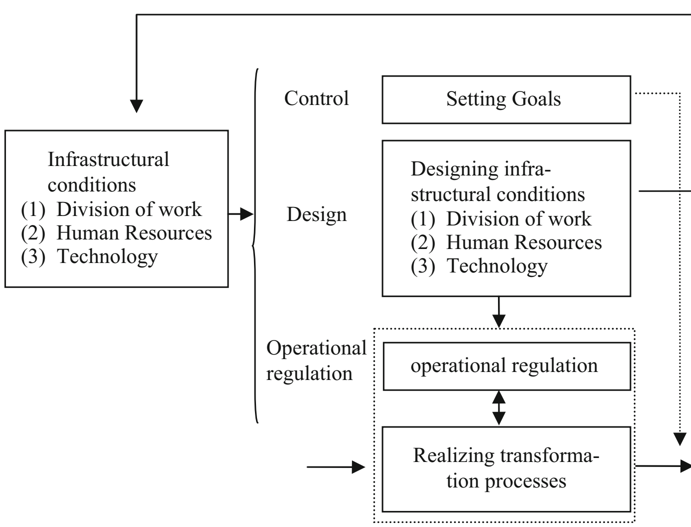


Fig. 2.5 Self-regulation (in terms of a concrete organization performing C, D, and OR-activities) requires a particular set of infrastructural conditions

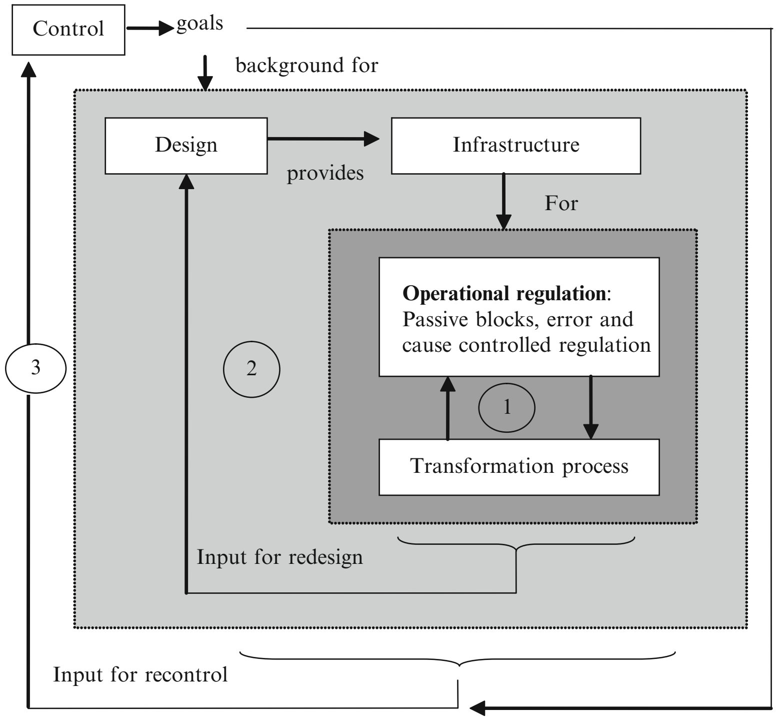


Fig. 2.6 Feedback-loops in adaptive, self-regulatory organizations

vis-à-vis a transformation process. Given goals and infrastructure, this loop involves continuous reacting to disturbances by applying different kinds of operational regulation. In this loop disturbances are blocked by passive blocks; or dealt with in a more active fashion by means of error or cause controlled regulation.

The second feedback-loop may be called the "design-loop." Given the goals, set by control design provides an organizational infrastructure, by means of which disturbances are attenuated, and which is the basis for performing transformation processes and their operational regulation. Monitoring this infrastructure, that is, its potential for attenuation, performing processes and their regulation, may result in redesign. Redesign simply means changing the infrastructure so that its potential for attenuation, performing transformation processes, and their regulation improves. In the figure, an arrow is drawn from operational regulation to design – because some operational regulatory activities may have design-consequences (see earlier).

In the third feedback-loop (the control-loop) goals for the transformation process are set, serving as a background for designing an organizational infrastructure. Monitoring goals (their environmental appropriateness, as well as whether they can be realized by means of a particular infrastructure) can result in resetting goals ("re-" control).

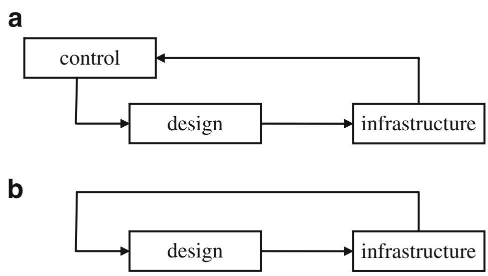


Fig. 2.7 Feedback-loops in adaptive, self-regulatory organizations

With respect to these three loops, two remarks can be made. The first is that all three loops contain a more or less continuous 'monitoring aspect' (implicit in the explanation thus far). This means for the control and design loop that (regardless of actual disturbances) goals may be reset or the infrastructure may be redesigned. For the operational feedback loop this means that one is continuously monitoring events and evaluating their disturbing character, so that cause controlled regulatory actions may be taken.

A second remark concerns the 'nested' relation between the loops, representing a kind of default (but not fixed) loop-sequence with respect to dealing with actual disturbances. That is, if problems occur – they are usually first dealt with within the operational feedback loop. Only if these disturbances persist (and one lacks the prospect of successfully dealing with them within the operational loop), redesign may take place; and if this does not resolve the problem, resetting goals may be seen as a last resort.

In Fig. 2.7 two additional feedback-loops are given that are not directly tied to realizing the operational transformation processes, but to the appropriateness of the infrastructure with respect to supporting control and design activities. In the fourth feedback-loop (Fig. 2.7 a), the adequacy of the infrastructure in supporting control-activities is monitored which may result in a redesign of the infrastructure. In the fifth feedback-loop (Fig. 2.7 b) the adequacy of the infrastructure for performing design-activities is at stake: it monitors whether the division of work, human resources (and HR systems) or technology to design an infrastructure may be improved and redesigns it, if necessary.

2.4 Organizations as Systems Conducting Experiments

After an introduction to Ashby's ideas about effective methods for studying and controlling complex systems, we apply these notions to the main theme of this book: organizations as systems conducting experiments. In this section, we explain that, viewed from Ashby's cybernetic perspective, organizations necessarily conduct experiments. In this way, we use all that was said in the previous sections to describe organizations and explain how cybernetics can be used to study them.

To explain what this experiment entails, it is helpful to repeat the nature of the cybernetic view of regulating organizations we discussed in the previous sections. In short, it means that, in order to deal with problems in/of organizations, one has to:

1. Control: define the relevant variables and their desired values (set the goals)
2. Design: provide a mechanism, which can be expressed by a "regulation table." By means of this mechanism disturbances are attenuated and operational regulatory potential is amplified. In organizations, this mechanism harbors an infrastructure aimed at
   - Realizing transformation processes and
   - Regulating these transformation processes operationally
   - Design
   - Control
3. Regulate operationally: use the designed mechanism / regulation table, to "block the variety of disturbances to the relevant variables" – i.e., to select error or cause-controlled regulatory actions and implement them in order to deal with specific disturbances in specific circumstances.

Moreover, dealing with problems in organizations is a continuous process in which (1) goals are set and reset; (2) disturbances can be attenuated and regulatory activities can be added more than once; and (3) more than one regulatory activity can be selected and implemented. The goal of this continuous process is meaningful survival.

For organizations this process is fundamental. In organizations one continuously adopts goals, creates conditions and actions to realize them, and performs these actions so that the goals are actually met. Moreover, in the course of setting goals and realizing them it may turn out that (due to changing circumstances and/or insights) goals, conditions and actions may have to be adjusted. And, because the ultimate goal of this process is survival, it may be said that organizations continuously translate and realize survival by means of control, design and operational regulation activities – see Fig. 2.8.

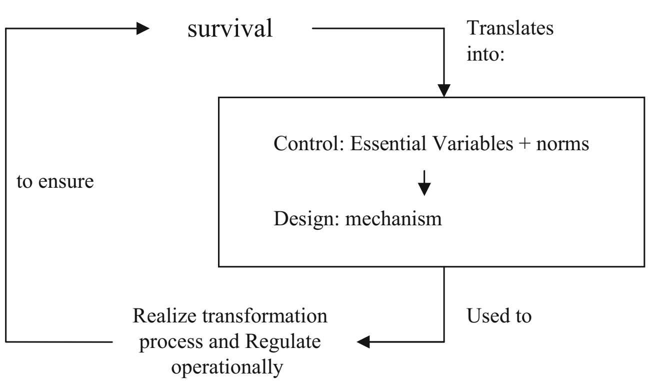


Fig. 2.8 The continuous process of translating and ensuring survival by control, design and operational regulation

By referring to the control, design and regulation activities, organizations can be described as adaptive self-regulatory systems – systems that, by themselves, by means of these activities continuously aim at meaningful survival.

In the course of surviving, organizations face particular difficulties. First, the translation of the "overall idea of meaningful survival" into (a constellation of) specific variables and norms for a particular organization is not straightforward. And, second, the same holds for the design of an infrastructure and the success of its implementation. We will briefly discuss both related problems.

For living organisms many of the essential variables are fixed. A human being cannot "ignore," for instance, its body temperature or blood pressure as essential variables for survival. In a similar vein, the range of desired values is also more or less fixed for many of these variables. These variables and their desired values have stabilized in the process of the organism's evolution and can not be altered freely. For organizations, something similar can be said. For them it is essential that they can adapt and realize their goals. However, what these goals are, and how they are to be realized is not fixed and, therefore, subject to choice. For instance, a company producing information and communication technology may choose its innovative capacity as a variable essential for its survival. It does so, relative to its understanding of the specific current and projected environmental circumstances and its own competencies. For other organizations, this capacity may not be relevant at all – again based on an understanding of the specific circumstances. Given a selection of some essential variable(s), the desired values also are subject to choice. In the example of the capacity for innovation the variable first needs to be "specified" – e.g., in terms of number of relevant patents or the time-to-market of an innovation. Given the specification of variables at such a level that values can be attached to them, one has to choose desired values. This, again, depends on an understanding of the specific current and projected environmental circumstances of an organization and its own competencies. What is the desired "number of patents" and "what are relevant patents"? To determine the relevance of patents, one needs to develop an understanding of the effect of such patents for the meaningful survival of the organization (e.g., in terms of its future market position or profitability, or societal contribution).

A similar argument can be given for design and operational regulation, with respect to the contingency of the elements appearing in it. In fact, the main argument is that in the course of adapting to its environment (by means of control, design and operational regulation) organizations continuously face uncertainty. The relevant variables, parameters and their values entering this process can neither be known *a priori* nor *a posteriori* with certainty. They are contingent and there is no procedure for determining the "most" successful ones. To be more precise: this fundamental uncertainty manifests itself at three levels:

1. At the level of control, the essential variables and their desired values are not a priori known. A selection depends on an understanding of the current and projected environmental circumstances and the organizational competencies.

2. At the level of design, the parameters and their values are contingent
   a. That is, the current set of disturbances can not be known with certainty, for it depends on our knowledge of the environment – which can never be complete.
   b. The projected set of disturbances is contingent – many different futures can be projected, giving rise to many different possible disturbances.
   c. The set of attenuating and amplifying measures and their effect is also not a priori known.
3. At the level of operational regulation, the selection and implementation of a specific regulatory action is also not a priori fixed. It depends, among other things, on the regulator's understanding of the specific disturbance and the projected effect of a regulatory measure in the specific circumstances for a specific disturbance.

To sum up: if we follow Ashby's cybernetic perspective, we see that in their effort to survive, organizations face an inescapable contingency. They are forced to select from the sets of possible variables, desired values, parameters, their values, etc., while the outcomes of these selections (their contribution to the meaningful survival of the organization) is not a priori known. In this sense, one could say that organizations are continuously (forced to) conduct experiments. In this experiment, essential variables, their desired values, parameters etc., can be viewed as "hypotheses." That is, in the experiment organizations hypothesize that if the organization selects *these* variables with *these* desired values it can survive and if the organization selects *these* disturbances, *these* regulatory actions, *these* conditions for selecting and implementing them, and if regulators in the organization are able to select the "right" regulatory actions, given their understanding of disturbances, then, the organization will actually survive (as specified by the essential variables and their desired values).

Moreover, experimenting with the organization's meaningful survival is a continuous process. It is a process in which the organization can select new variables, desired values, other disturbances, other ways to attenuate and/or amplify them, select other conditions, etc. It is, in fact, a continuous process of "cybernetic problem-solving" as we described earlier in this chapter: a continuous process in which control, design and operational regulation activities are performed iteratively.

As we stated, this experiment is central in our cybernetic perspective on organizing in the rest of this book. In this chapter we treated the basic cybernetic elements for understanding the experiment. In the coming chapters we elaborate on the nature of this experiment (the rest of Part I of this book) and discuss the necessary conditions for conducting it (in Part II). The nature of the experiment is elaborated in the next chapter (in which the central issue is the process of *selecting* variables, desired values, parameters, etc. and the role of knowledge in this process). In Chap. 4 we relate the "experimental arche" of organizations to their "social arche." In Part II we discuss the conditions, necessary for conducting organizational experiments in terms of required *functions* (Chap. 6) and *structures* (Chap. 7).

References

Anthony, R. N. (1965). *Planning and control systems: A framework for analysis*. Harvard, MA: Harvard University Press.

Ashby, W. R. (1958). *An introduction to cybernetics* (3rd ed.). London: Chapman & Hall.

Ashby, W. R. (1960). *Design for a brain* (2nd ed.). London: Chapman & Hall.

Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science, 1*(2), 89–97.

Chapter 3
The Experimental Arche Continued: Von Foerster on Observing Systems

3.1 Introduction

In the previous chapter we used Ashby's cybernetic theory to discuss the "experimental arche" of organizations. This arche referred to a continuous and risky process of control, design and operational regulation with respect to organizational transformation processes. At the heart of our discussion of the experimental arche was Ashby's regulatory logic, stating that, in order to regulate a particular concrete system, one has to:

1. Select essential variables and desired values
2. Identify parameters, disturbing the essential variables
3. Design an infrastructure (a "mechanism") by means of which:
   – Disturbances are attenuated
   – The system's transformation processes can be realized
   – Regulatory potential (regulatory parameters) becomes available
4. And, given 1, 2, and 3: select values of regulatory parameters (= select regulatory actions) in the face of actual disturbances.

Moreover, in this Ashby-based notion of regulation, one needs a model of the behavior of the concrete system: a transformation. According to Ashby (1958), a good (conditional, single-valued) transformation relates the selected variables and parameters in such a way that predictions can be made about the behavior of the concrete system. To arrive at such a transformation, the black-box method was introduced – a method enabling a regulator to derive a transformation based only on the values of the variables and parameters that are chosen to describe the concrete system that should be regulated. Ashby's black box method seems to suggest that we can "objectively" select variables and parameters, and derive a transformation connecting them based on trial and error, without, as Ashby puts it, "reference to prior knowledge". If this is what regulating systems is about, one might say that it does not contain much risk. It is "just" a matter of selecting variables/parameters; observation and deduction. The risk attached to it may have

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_3, © Springer-Verlag Berlin Heidelberg 2010

to do with the mistakes we make in selecting variables/parameters or in deducing a conditional transformation from empirical observations; or it may have to do with time-constraints we face while regulating; or with the probabilities governing the behavior of the system.

However, in the introduction of this book, we argued that the organizational experiment was "fundamentally risky." The experiment was said to be risky because one cannot be sure about the effect of the selected variables, parameters and their values. This can now be specified with respect to Ashby's regulatory logic. It means that one can neither *a priori* nor *a posteriori* select with certainty, the "right" variables and their desired values (in terms of their contribution to meaningful survival), the right parameters and their desired values (in terms of their contribution to the realization and regulation of transformation processes). And, at the same time, it is impossible to be certain about the relation between variables and parameters – i.e., about the transformation describing the system's behavior. The fundamentally risky character of the experiment, then, has to do with the fundamental uncertainty of (1) the *selection* of variables and parameters (and their desired values), and (2) the *model c.q. transformation* relating these variables and parameters.

Our argumentation about the fundamentally risky character of the experiment is based on the theory of Heinz von Foerster (1911–2002). In contrast to Ashby, who seems to hold that regulators can "objectively select" variables/parameters and derive transformations based on trial and error, without reference to "prior knowledge," von Foerster claims that for most of the systems we have to deal with in our daily lives, both selecting variables (and parameters) as well as determining a (conditional single-valued) transformation is *impossible* without reference to prior knowledge. As he sees it, any concrete system can be described by means of an indefinitely large number of variables and parameters. There are so many possible variables and parameters to select from, that it would be a downright miracle if we could pick out the relevant ones by trial and error. Likewise, the number of possible conditional single-valued transformations connecting variables and parameters is so large that it is impossible (or sheer luck) to derive one from empirical observations alone. But, if von Foerster is right, and both tasks (selecting variables/parameters and constructing a transformation) are really impossible most of the time, how, then, do we deal with all the systems we are continuously confronted with? How do we regulate them if the necessary prerequisites for regulation (variables, parameters and transformations) cannot be met?

To deal with the problem of selection (of variables, parameters and their values) and the problem of determination (of the conditional single-valued transformation), von Foerster stresses that they are essentially related to how we, as observers, define systems and their behavior. A paradoxical situation arises: on the one hand, selecting variables, parameters and transformations out of an indefinitely large number by means of trial and error is impossible, while on the other hand we do regulate systems, and therefore, we must have made a particular selection of variables/parameters and transformations. The way out of this paradox is to drop the assumption that we select (variables and parameters) and determine (transformations) based on trial and error alone. Instead, we rely on our prior knowledge for this selection and determination. Von Foerster argues that it is a fundamental mistake of so-called "objective approaches" to maintain that "no observer characteristics" should enter the description of the phenomena they observe (von Foerster 1993). This attitude seems to exclude observer-dependent prior knowledge in defining systems and their behavior – an attitude which, von Foerster might say, *causes* the two problems in the first place. Instead, von Foerster sets out to deal with the two problems by examining the role of observers (*including* their prior knowledge) when they define systems and their behavior.

Von Foerster (along with other cyberneticians) proposes to use cybernetics to examine the role of observers – i.e., to research the relation between observer and observed system. To be more precise, concepts from cybernetics (like stability and feedback) are used to describe the regularities in the relation between observer and observed system. If cybernetics is used in this way, it is often referred to as "second order cybernetics."

Ashby's cybernetic theory (which may be called "first order cybernetics") contains a particular set of concepts to deal with observed systems (i.e., concepts to describe the regularities in their behavior and to propose regulatory actions on the basis of these regularities). Second order cybernetics uses this same set of concepts to deal with the *observer* of systems. The title of one of von Foerster's books – "Observing systems" (von Foerster 1981) – harbors an ambiguity that nicely refers to what second order cybernetics entails. "Observing systems" refers to the fact that in (both first and second order) cybernetics one observes systems – i.e., the object of study of cybernetics is "a system." It also refers to the fact that in second order cybernetics the object of study is a *specific kind* of system: the system that observes (the observing system). In one word, second order cybernetics is about observing the systems that observe (with the instruments of first order cybernetics). As von Foerster puts it: instead of "looking at things out there" – which is what first order cybernetics does, second order cybernetics turns to "looking at looking itself" (von Foerster 2002).

Although several authors can be accredited for discussing the role of the observer in dealing with systems (for instance, Maturana and Varela 1980, 1984; Pask 1992; Varela 1988; Varela et al. 1993 or Luhmann 1984), we choose von Foerster as the leading author for this chapter. It was von Foerster who explored and elaborated the role of the observer in cybernetics and its consequences for cognition, regulation, and action. His work can be used to provide a precise delineation of the two problems appearing in first order cybernetics (i.e., the problem of selection and the problem of determining a transformation), and to provide an answer to them. Von Foerster succeeded in making second-order cybernetics and its consequences accessible by popularizing them in a large number of papers, books, interviews and lectures (see, for instance, Segal 1986; Cybernetics and Human Knowing 2003; or von Foerster and Poerksen 2002; for overviews). He is regarded by many as an inspiring and cohesive force in the development of second-order cybernetics. For instance, as founder and head of the Biological Computer Laboratory (1957–1976), he initiated and facilitated leading-edge research on (second-order) cybernetics.

The second-order cybernetic perspective has consequences for studying and regulating organizations. If the choice of variables and parameters, and if the description of the transformation both depend on the observer, it follows that regulation and control of organizations are observer-dependent. In fact, as we will discuss, this perspective highlights the hypothetical nature of the conjectures that are part of "the organizational experiment" (see chap. 1). In particular, the perspective makes apparent that these experiments are fundamentally *risky* experiments.

To describe the contribution of von Foerster's work to understanding the risky nature of the organizational experiment we organize this chapter as follows. In Sect. 3.2 we discuss the two problems of first order cybernetics. In this elaboration we introduce von Foerster's famous distinction between trivial and non-trivial machines. Next, in Sect. 3.3, we unfold his solution to these problems which is based on the cybernetic study of observers. Finally, in Sect. 3.4, we focus on the consequences of von Foerster's theory for the risky character of the organizational experiment.

3.2 Two problems in first-order cybernetics

As we discussed in the previous chapter, Ashby's theory on regulating complex systems relies on the selection of variables and parameters and on the (conditional single-valued) transformation (describing the behavior of the system) relating the selected variables and parameters. However, selecting variables and parameters as well as determining the transformation are problematic. In this section we treat these two problems and introduce von Foerster's way of dealing with them.

3.2.1 The problem of selection

To regulate a concrete system an observer needs to select a (list of) variable(s) used to describe its behavior. In the previous chapter, we said that an "organization" can be described by many variables. For instance, a description containing the variables "net profit" and "number of staff," or one with the variables "number of products" and "net-profit per share," or... In fact, a concrete system "organization" seems to allow for an indefinitely large number of possible variables and variable combinations. Which one should the observer choose?

Ashby acknowledges this problem. He states: "every material object contains no less than infinity of variables and therefore of possible systems" (1958, p. 35). To solve this problem, Ashby proposes that observers "should pick out and study the facts that are relevant to some main interest that is already given". But how does an observer decide what facts are relevant to such a "main interest"?

The observer's choice of variables is, of course, not a random one. The aim is to regulate the behavior of a concrete system. In the case of organizations, it has to do with regulation with respect to "meaningful survival." However, if observers have to select variables relevant for the regulation of a concrete system they need knowledge about regulating concrete systems (both with respect to regulation in general and to regulating the specific system under consideration). Without such knowledge, successfully selecting the right variables from an almost infinite number would be sheer luck. It may thus be conjectured that observers have some kind of "pre-orientation" regarding the relevance of variables, used for describing the behavior of the system the observer wants to regulate (cf. Winograd and Flores, 1986).

This puts the problem of the selection of variables in a new problematic perspective. On the one hand, if we *do not* have such a pre-orientation, we – as observers – are completely in the dark when picking out the variables to describe the behavior of the concrete system. On the other hand, if we *do* have such a pre-orientation, the question arises where this pre-orientation comes from and what its content is. The problem of selection becomes the problem of pre-selection!

The same holds for the selection of parameters. To arrive at a relevant description of the behavior of the system (one that can be used to understand or influence it) one also needs to select parameters. And, again, without prior knowledge, it seems to be a hopeless enterprise to pick those parameters that are actually relevant for the behavior of the system. This task cannot be accomplished without a pre-orientation on what counts as relevant – e.g., a model about how the environment can influence the variables of the system.

To formulate an answer to the problem of selection we can use von Foerster's theory. He proposes that theories about defining and regulating systems should include properties of their observers. In particular, he proposes that a theory explaining how observers select the relevant variables/parameters to describe a particular system should include the observer's pre-orientation (s)he uses to select the variables/parameters to describe the behavior of a concrete system. The important question then becomes how an observer develops such a pre-orientation. To deal with this question, von Foerster uses insights from (first order) cybernetics. With these insights he explains how we build up, during our lives, preferences for defining systems in a particular way. How? Technically phrased: these preferences are the "eigenvalues emerging due to the closure of our cognitive system." An explanation of these concepts and how they help to explain the emergence of the observer's pre-orientation is the topic of Sect. 3.3.

3.2.2 The problem of defining the transformation

To regulate complex systems, Ashby proposes to select variables and parameters and to use them to arrive at a description of their behavior. In the previous section the problematic nature of selecting variables and parameters was introduced. In this section we deal with the problem of deriving a (conditional single-valued) transformation, a description of the behavior of the system.

To introduce this problem, suppose, that we already have at our disposal adequate variables and parameters to describe the behavior of the concrete system. In other words, we ignore the selection problem. For example, suppose that we have a box with a switch. The switch has two positions (on and off). If the observer of the box puts the switch in position "on," the box produces a soft humming. If the switch is in position "off," the humming stops. The variable to describe the behavior of the box may be labeled "Humming" and has two values: "Humming on" and "Humming off." The parameter (switch) also has two values (on and off). The relation between the value of the parameter (the input) and the behavior of the box (output) is given in Table 3.1.

Table 3.1 Transformation describing a humming box

T_{Switch}	Humming on	Humming off
SWITCH = ON	Humming on	Humming on
SWITCH = OFF	Humming off	Humming off

This table describes the behavior of the box in terms of two transformations, dependent on the parameter-value (see also chap. 2). The table describes the behavior in a conditional form, which rests on the assumption that the relation between switch position and behavior (or output) is fixed. Von Foerster calls systems that have a fixed input-output relation "trivial machines" (see, for instance, von Foerster 1970, 1984, 1992).

A trivial machine realizes a function that maps the input-states onto the output-states. In the example: if X is the input (the parameter; the position of the switch – with two possible states) and Y is the output (humming on or humming off), the table realizes a function F(X) that maps X onto Y: F(X) = Y. This function F determines the output (Y), based on a specific value of the input (X) – according to the above table. The possible input-output combinations are listed in Table 3.2.

Table 3.2 Possible input-output combinations of the humming box

Value of X:	Value of Y:
ON	ON
OFF	OFF

In this case, the relation between the input and output states is fixed and an observer can easily derive the machine's function. In general, by experimenting with a trivial machine, observers may –in principle– derive this function.[1]

Von Foerster states that trivial machines have four important features (e.g., in von Foerster 1984, 1992). They are:

[1]This is easy and straightforward with a limited number of input and output states. It quickly becomes tiresome, however, if the number of states increases. If the number of input-states is denoted by #X and the number of output-states by #Y, the total number of possible trivial machines is $\#Y^{\#X}$. In the example, there are 4 possible trivial machines and the observer needs at most 3 trials to determine the right one. If the number of states only moderately increases, the number of possible machines increases rapidly (cf. von Foerster 1970).

1. Synthetically determined

This means that the behavior of the machine is completely determined by the way it is put together ('synthesized') – the behavior is hard-wired, so to speak. In the case of the humming box, one could say that it is possible to make (or model) one that embodies the behavior of the box. And, by means of embodying the function driving its behavior, its behavior is determined.

2. Analytically determinable

This means that it is possible to determine (analyze) the relation between the input and the output of the machine – based only on observing the input-output combinations. As with the humming box, it is possible to draw up a table (embodying a conditional single valued transformation), describing its behavior.

3. History-independent

A trivial machine is said to be history-independent because the relation between input and output is always the same, regardless of its specific history of input-output combinations. In the example, this independence is made apparent by the two columns of table showing its behavior given a certain input value. Both columns are identical, meaning that it does not matter for the determination of the behavior of the box what the previous value of the humming-variable was.

4. Predictable

Once the relation between input and output is known (features 2 and 3), the input determines the output with certainty. So, given a certain input, the behavior of the trivial machine can be predicted with certainty.

A trivial machine shows a fixed relation between input-states and output-states, and only one function needs to be used for their mapping. To determine the transformation describing the relation between input and output for a trivial machine is relatively simple. In fact, the problem of determining transformations would not really be problematic if we only had to deal with trivial machines. However, as von Foerster, states, the vast majority of the systems we encounter in our daily lives are "non-trivial" machines. Non-trivial machines are fundamentally more complex than trivial ones. As he describes: the important difference between a trivial and a non-trivial machine is that the behavior of a non-trivial machine is not only dependent on the input-state, but also on an internal state – that is itself subject to change (von Foerster 1992, p. 62).

Now, what, exactly are non-trivial machines? To explain them and to set them apart from trivial machines, we use Fig. 3.1 as a base.

Just like trivial machines, a non-trivial machine shows behavior that can be described by a sequence of states (the output-states). In the figure, the output states are values of the variable indicated by Y. A non-trivial machine also has one or more parameters influencing its behavior: the input state (X in the figure). In this respect they resemble trivial machines. However, what sets them apart from trivial

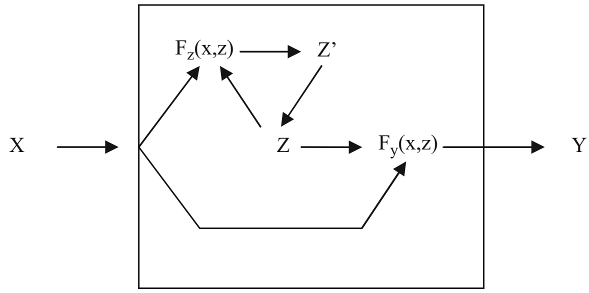


Fig. 3.1 A model of a non trivial machine (cf. von Foerster, 1984, 1992)

machines is that they have an additional parameter (or set of parameters), representing an internal state that also determines the value of the output state. In the example, this internal parameter is denoted by Z.

At any moment in time the internal state Z has some specific value. Now, given a new value of X – the machine calculates (see also Fig. 3.1):

- The output Y, by means of $F_y(x,z)$ – using the current value of X and Z.
- The new value for the internal state: Z', by means of $F_z(x,z)$, also using the current value of X and Z.

The new value of the internal state is used for subsequent calculations of the output and internal state.

To illustrate how a non-trivial machine works, suppose that we have Table 3.3 describing the behavior of a non-trivial machine NTM.[2]

Table 3.3 Example of a non-trivial machine

NTM		X=A	X=B	X=C	X=D
Z = α	Y	1	2	3	4
	Z'	α	α	β	β
Z = β	Y	4	3	2	1
	Z'	α	β	α	β

The functions calculating the machine's output have two types of input: an "external" input X (which can take on the values A, B, C or D – see shaded top row), and an "internal" input (the internal state Z, either α or β – see shaded left column). The machine has one output-variable: Y. A non-trivial machine calculates its output on the basis of the value of the external input X *and* the internal state Z.

[2]Beware: this table is different from the Ashby-based regulation-tables, for both the values of output states and the internal parameter co-determining the output are given in the table.

Given the value of the internal state Z (left column) and a value of the external input (top row) the output Y is given in the table (row 2 and 4). The next value of the internal state (Z') is also given in the table (row 3 and 5).

Suppose, for example, that the machine starts with α as the value of the internal state (i.e., Z = α). Now, if an observer feeds the machine with an A (i.e., the input X = A), then the output (Y) is 1 and the next value of the internal state (Z') is α. This is given in rows 2 and 3 (under X = A). If the next input is C, the machine uses this input and the new value of Z (α, which is Z' as calculated in the previous step) then the output becomes 3 and the next value of the internal state becomes β. The reader may verify that, if the machine starts with Z = α, the input-sequence A, A, B and A corresponds with the output-sequence 1, 1, 2, 1, and the input-sequence A, B, C and A corresponds with the output-sequence 1, 2, 3, 4.

For an observer, trying to match the behavior of the machine with his/her manipulations of the input-variable, these two sequences are highly disturbing. The same input-value (A) may result in different output values: it may result in 1 or 4. In Ashby's terms: the transformation is multi-valued. The reason for this is that the behavior of the machine also depends on the value of its internal state – a state that cannot be observed. And that value, in turn, depends on the previous values of X and of the internal state.

The question arises whether there is any hope of deducing a single-valued transformation (the description of the behavior) from the manipulation of the machine without any knowledge of its internal state. Von Foerster's answer is: no. What the observer should do in such a case is conjecture the existence of one or more internal states and try to deduce the function(s) determining its or their behavior. This is, however, impossible. The number of possible machines is simply too large. Von Foerster (1992) calculates, for instance, that the number of possible machines with four input states and four output states (like the machine we described above), is 2^{8192} (which is, approximately, 10^{2466}). Von Foerster compares (1992, p. 65) this number with the number of elementary particles in the universe (which is estimated to be a mere 10^{72}) and with the number of microseconds the earth exists (about 3.10^{23}) and concludes that finding the right machine is impossible. As he puts it: the number of possible machines is "transcomputational" (von Foerster, 1984). This makes non-trivial machines harder to deal with than with their trivial counterparts. To underline this, and to show the difference between trivial and non-trivial machines, von Foerster (1984, 1992) gives the following four characteristics of non-trivial machines. They are:

1. Synthetically determined

This means that the behavior of the non-trivial machine is completely determined by the way it is put together ('synthesized').

2. Analytically indeterminable

This means that it is impossible (except for very few non-trivial machines with very few input and output states) to determine the relation between the input and the output of the machine – based only on observing the input-output combinations.

3. History-dependent

A non-trivial machine is said to be history-dependent because the relation between input and output depends on the value(s) of (an) internal state(s) (which, in turn, are calculated based on the input). This is the main difference with trivial machines and the reason for its analytical indeterminability and its unpredictability.

4. Unpredictable

Because a non-trivial machine is history dependent and hence analytically indeterminable its behavior cannot be predicted (with certainty).

To summarize: the output of non-trivial machines depends on an external input *and* on an internal state. This makes them transcomputational and therefore indeterminable and unpredictable.

Non-trivial machines have a large impact on the applicability of Ashby's black-box method. Von Foerster would argue that it may be applied to find the transformations of a trivial machine, but that it is impossible to use it to determine the transformations describing a non-trivial machine. This would not be problematic if we did not have to deal with non-trivial machines. However, some reflection on the "machines" we encounter in our daily lives shows that only very, very few are trivial! Most of them and the most interesting ones have the characteristics of non-trivial machines (cf. von Foerster 1984, 1992).[3] The rule for calling a machine a non-trivial one is that it has internal states representing its history (or memory) and co-determining its behavior. Given this rule, it is easy to see that almost all machines we have to deal with are non-trivial: e.g., animals, humans, departments, or organizations. The behavior of all these "machines" depends on some recollection of what happened in the past (represented by changes in internal states). For most living systems this is straightforward: they have a memory. For other systems, however, history is also often represented, but we normally do not call the means by which they do that "a memory." For instance, the time needed by a sawing machine to produce planks, not only depends on the input's seize and hardness, but also on an internal state, representing, for instance, the degree of wear of the machine.

If most of the machines we have to deal with have the characteristics of non-trivial machines, it seems that observers can do nothing but abandoning the hope of analytically determining their transformations. However, if the transformations are undeterminable, how can we deal with the systems we encounter? How can we understand them? How can we regulate them?

[3] Although most of the non-trivial machines we have to deal with have a large number of input and output states, Von Foerster's argument even holds for non-trivial machines with few input-states and output-states (as the one from our example, which only has four input and four output states).

3.2.3 Dealing with Both Problems: The Cybernetics of Observing

If we want to use first-order cybernetics to deal with the systems around us, we face two problems: the problem of selecting the proper variables (and parameters) to describe their behavior and the problem of determining a (conditional single-valued) transformation, describing their behavior. If these prerequisites cannot be met, regulation does not seem to be possible according to Ashby's theory. However, we seem to be quite capable of dealing with all kinds of systems in our daily lives – our friends, children, families, the machines we operate, the organizations we manage. . ..

The question may arise whether first order cybernetics is a good descriptive theory of what happens when we regulate systems. We think it is, but, along with von Foerster, we think that an important assumption regarding its application is missing and needs amending. That assumption has to do with *the role of the observer*.

The role of observers is no issue in first order cybernetics. As von Foerster puts it, in first order cybernetics one seems to believe in "objectivity" (cf von Foerster 1988, 1992, 1993 2002). That is, given some relevant purpose, the variables and parameters present themselves somehow, "objectively," and by using the black-box method the transformation can be deduced. Ashby asserts that observers should not refer to prior knowledge when defining systems. But, as we have already stated above, this attitude towards observers, essentially, causes the two problems! Von Foerster would argue that this prior knowledge is what *enables* observers to deal with systems. Every time observers encounter some system, their prior knowledge of dealing with them co-determines which variables and parameters they (implicitly or explicitly) use to describe it and which conjectures they make about its behavior. Put differently: observers always have a particular pre-orientation enabling them to select relevant variables and parameters and to construct a model describing the system's behavior.

The way to deal with the problems of first order cybernetics, then, is to make explicit what this pre-orientation is, where it comes from, and what its role is in regulating systems. If this is explained, both problems disappear. Von Foerster treats the emergence of this pre-orientation in a specific way. He explains how observers, during their phylogenetic and ontogenetic history build up a pre-orientation consisting of preferences regarding the description of concrete systems and their behavior in a particular way. He states that these preferences can be seen as more or less "stabilized hypotheses" about the concrete system and its behavior enabling the observer to deal with it and he argues that these hypotheses are to be regarded as "*eigenvalues produced by the closed cognitive system of the observer*." In the next section we will explain what this "production of eigenvalues by the closed cognitive system or the observer" means and how it contributes to dealing with the two problems of first-order cybernetics.

3.3 Observers as Closed Systems Producing Eigenvalues

To deal with the problems of first order cybernetics, von Foerster proposes to highlight the role of the observers of systems and use first order cybernetics to explain how "observing systems" come to observe the things they observe. To do so, he treats observers as "closed systems in which eigenvalues emerge". This idea of "eigenvalues produced by the closed cognitive system of the observer" is the key to understanding von Foerster's solution to the two problems of first-order cybernetics. However, at this point it is rather difficult to understand what is meant by it, because almost none of the key's constituents have been treated. The reason for already introducing this key is that it helps to structure the presentation of von Foerster's reasoning and as such it may serve the reader as a marker, pointing at relevant issues that need to be dealt with. This is especially helpful in this section, because the individual ingredients of the key are in need of a rather intricate explanation, and one might easily lose track.

So, in this section, we work towards an understanding of von Foerster's notion of "eigenvalues produced in the closed cognitive system of the observer" and how this idea helps in solving the problems of first order cybernetics. In order to do so, we organize this section as follows. We first start with dealing with "eigenvalues produced in closed systems." In particular, in Sect. 3.3.1. we treat three related questions: (1) what are closed systems? (2) what are eigenvalues?, and (3) how do closed systems show eigenvalues? In Sect. 3.3.2 we move on to discuss the observer as a closed system in which eigenvalues emerge. Relevant questions are: how is the (cognitive system of the) observer regarded as a closed system? And, what does it mean that eigenvalues emerge in the observer's closed cognitive system? These two sections uncover von Foerster's key to the solution of the two problems of first order cybernetics – but not *how* it leads to the solution itself. This is the topic of Sect. 3.3.3.

3.3.1 Closed Systems and Their Eigenvalues

The closure of a system is a central concept in von Foerster's theory. In essence, in a closed system its output also serves as its input. If systems are closed in this sense, cyberneticians argue that they may develop recurrent patterns of behavior. Cyberneticians use the term "eigenvalues" or "eigenbehaviors" to refer to these patterns. To explain these phenomena more thoroughly, von Foerster introduces a simple formalism (e.g., von Foerster 1984, 1992). In this section we introduce this formalism and use it to explain what the emergence of eigenvalues in closed systems amounts to. Moreover, we discuss several aspects of the formalism in more detail in order to explain "observing systems" in Sect. 3.3.2.

3.3.1.1 A Formalism to Explain the Emergence of Eigenvalues

Von Foerster's formalism to explain the behavior of closed systems contains:

1. A set of elements (A)
2. An operation OP(. . .), acting upon the set of elements A (notation: OP(a) : A → A). This means that the input for the operation (a) comes from A, and it produces as output elements from A.

If the set is the set of all natural numbers, then the operation may be given by, for instance, OP(n) = n+1. In this example the operation takes a natural number as its input and produces a natural number as its output. Another example is the transformation in Table 3.4

Table 3.4 Example of an operation

OP↓	0	1
	1	0

In this example, the set of elements is {0,1} and OP(0) = 1 and OP(1) = 0. Here, too, input and output of the operation belong to the same set. This example also shows that von Foerster's formalism can be cast in Ashby's transformation-notation from the previous chapter.[4]

Because the operation produces elements belonging to the same set as the elements that it can take as its input – it is said to be a *closed* operation. In Ashby's terminology, these operations are called "closed transformations."

Von Foerster is particularly interested in recursively applying operations. That is, the output of the operation is used as input for the same operation. Suppose, we have the operation OP(x) = x+1, acting on the set of natural numbers. We may take as the first argument the natural number 1. The output will be 2 (notation: OP(1) = 2). Now, if we use this output as input for the next application of the operation, we "recursively apply" the operation. This results in the output 3. This output may again serve as the input of the next application of the operation, etc. So, we obtain the following results:

$$\begin{aligned} &\mathrm{OP}(1) = 2 \\ &\mathrm{OP}(2) = \mathrm{OP}(\mathrm{OP}(1)) = 3 \\ &\mathrm{OP}(3) = \mathrm{OP}(\mathrm{OP}(\mathrm{OP}(1))) = 4 \\ &\text{And so on.} \end{aligned}$$

[4] To see the resemblance, it is important to note that what von Foerster treats as input for the operation is called "operand" by Ashby, and von Foerster's output is Ashby's transform.

If we denote the n-th application of the operation on an element x by $OP^n(x)$, we see that

$$\begin{aligned}&OP^1(1) = 2\\ &OP^1(2) = OP^2(1) = 3\\ &OP^1(3) = OP^2(2) = OP^3(1) = 4\\ &\text{Et cetera.}\end{aligned}$$

If we also denote the first element we use as input by x_0 (also called the "initial value") and further agree that x_i is the result of the i-th application of OP on the initial value, that is: $x_i = OP^i(x_0)$, we get:

$$\begin{aligned}x_{i+1} &= OP^1(x_i) \text{ and}\\ x_{i+1} &= OP^{i+1}(x_0).\end{aligned}$$

So, regarding the example, the following holds:

$$\begin{aligned}&x_0 = 1;\\ &x_1 = OP^1(x_0) = OP^1(1) = 2;\\ &x_2 = OP^2(x_0) = OP^1(x_1) = OP^1(2) = 3;\\ &x_3 = OP^3(x_0) = OP^2(x_1) = OP^1(x_2) = OP^1(3) = 4;\\ &\text{And so on.}\end{aligned}$$

Now, with respect to this "recursive formalism," von Foerster considers the following limit:

$$\lim_{n\to\infty} OP^n(x_0) \tag{3.1}$$

He states –along with many other cyberneticians – that, for some operations, under some conditions, this limit does indeed exist. If it does, von Foerster proposes to call this limit the *eigenvalue*[5] of the operation (von Foerster 1981, p. 275 ff).

Von Foerster uses several examples to illustrate "eigenvalues." An example he often uses (cf. von Foerster 1984, 1992) is the operation "$\sqrt{\ldots}$". He claims: whatever its input (greater than 0), the eigenvalue is always: 1. He demonstrates his audience the emergence of this eigenvalue by repeatedly "pressing" the square-root button of a calculator for some initial value. After several recursions of the

[5] From standard algebra, one can learn that, for any function f, if a and λ exist such that $a = \lambda f(a)$, then a is called the eigenvector and λ the eigenvalue. In von Foerster's formalism, λ seems to be set to 1, and a is called the eigenvalue (e.g., Lipschitz 1987).

operation (by pressing the button) the display shows "1." With this operation and this set of elements, von Foerster claims:

$$1 = \sqrt{\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{\ldots}}}}}}}} \tag{3.2}$$

In other words, the eigenvalue of the operation "$\sqrt{\ldots}$" acting upon the set of positive real numbers is 1. Or, in the limit-notation given above:

$$\lim_{n \to \infty} \sqrt{(x_0)}^{n} = 1 \qquad 3.3$$

3.3.1.2 Three necessary elaborations of the basic formalism

The above formalism introduces basic elements for understanding the behavior of closed, complex dynamic systems – i.e., how eigenvalues may emerge in closed systems. However, to understand *observers* as complex systems with the aid of this formalism, some elaboration is needed – after all the operation of "observing" seems to be a bit more complex than the square-root operation. In particular, before we apply the formalism to observers, it is necessary to discuss (1) the diversity of eigenvalues emerging in closed systems (2) the role of input and (3) the complexity of the systems involved in producing eigenvalues.

The diversity of eigenvalues

Above we only talked about one specific numeric value as eigenvalue. If we want to describe the products of observing ("experiences of some kind") as eigenvalues we have to allow for different types of invariances (other than numbers) that may emerge. Moreover, since observers can experience many things we should allow for the fact that a closed system may produce more than one eigenvalue.

To discuss these issues, suppose that we have a machine realizing some operation OP acting upon some set of elements. The machine has three buttons, labeled "a," "b" and "c." An observer of the machine can press one of them, after which the operation of the machine is presented with some (random) initial value from one of three value-regions (region A, B or C – see Fig. 3.2). The machine then starts to apply its operation recursively to this initial value. Suppose further, that this operation produces one of two eigenvalues, dependent on the initial value, and that for some other initial values no eigenvalue appears. The machine's behavior is depicted in Fig. 3.2. If button a is pressed, the operation is presented with an initial value from region A and it computes (i.e., the recursive application of the operation

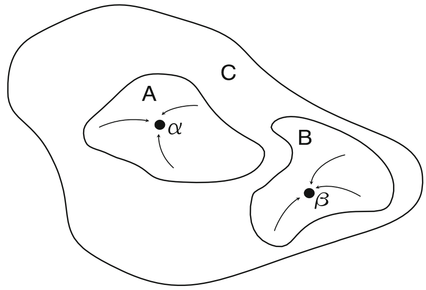


Fig. 3.2 Field of behavior of a particular machine showing two eigenvalues

on its initial value results in) eigenvalue α. If button b is pressed, the initial value comes from region B and eigenvalue β is computed. Initial values from region C (pressing button c) do not lead the computation of an eigenvalue.

The behavior of this machine shows that some initial values do, while others do not lead to an eigenvalue. Moreover, the machine has two *different* eigenvalues. One could say: the behavior of the machine is "α", "β" or something unpredictable, dependent on the initial value. Put in the terminology of von Foerster: dependent on the initial value the eigenvalue may differ or even be non-existent. In this example, an operation may lead to more than one eigenvalue. Moreover, different sets of initial values leading to different eigenvalues can be acknowledged.

A variation on the "multiple eigenvalue" theme is that an operation may lead to a *sequence* of reappearing values. Suppose, for example, that the above operation computes OP(α) = β **and** OP(β) = α. In this case, the resulting behavior of the machine is, eventually, α, β, α, β, α, β, α, β,. . . von Foerster (1981, p. 279 ff.) would say: both α and β are eigenvalues implying each other.

Until now the formalism is illustrated by using numbers (or letters) to indicate the eigenvalues. However, von Foerster explains that the formalism is quite general and can be applied to many closed domains. Dependent on this domain, different *types* of eigenvalues may emerge: "eigenfunctions", "eigenoperations", "eigenalgorithms", "eigenbehavior", etc. (1981, p. 279).

In all, if we want to study observers with the given formalism, we need to allow for different eigenvalues and for different types of eigenvalues. We need to allow different types of eigenvalues, because if von Foerster states that observations can be treated as eigenvalues, we need to go beyond the domain of numbers. And, because observers may have many different experiences (which are to be seen as the eigenvalues of observing systems), we need to allow for *many* eigenvalues to settle down.

The role of input

The behavior of the machine in the example shows that its operation can "absorb" certain displacements from the eigenvalue. Such eigenvalues are said to be "stable."

To see this, suppose button a is pressed, and the operation is given an initial value from region A. It then calculates eigenvalue α. If button a is pressed again, a new initial value from region A is randomly selected (which probably differs from α). By pressing button a again, we displace the value of operation from the calculated eigenvalue, and it will recalculate the previous eigenvalue. If this displacement is small, few computations may be necessary to regain this value. If it is large more may be needed. If we press other buttons, the displacement from eigenvalue α becomes so large that either another eigenvalue emerges (β) or non can be calculated. In such a case, it is said that the new initial value lies outside the stable region of the original eigenvalue. Eigenvalues can thus be used to explain the occurrence of specific behavior of a machine (calculating an eigenvalue) and to explain its "absorptive capacity" – the re-occurrence of the same specific behavior in spite of "disturbances" (displacements). Moreover, they can also be used to explain a shift in behavior, given a larger displacement (outside the "region of convergence").

Now, one may wonder where these displacements come from. In cybernetics, displacements are modeled by changes in "parameter-value." A parameter may cause the system to change its mode of operating. For this machine, this means that another initial value is selected. Based on this new value, however, the (or a new) eigenvalue may be (re) computed. When the parameter remains constant for some time – the machine is allowed to recursively compute some stable value. When the parameter changes its value, and again remains constant for some time, a stable eigenvalue (or eigenbehavior) may again be computed (or in some special cases the system would go on computing indefinitely).

How different parameter-values can trigger different eigenvalues is illustrated by the effect of pressing the buttons in the example. The observer may press several buttons and thus present the machine with different initial values. This can be modeled by using "button" as a parameter, whith possible values: "a," "b" or "c." Given a specific value of the parameter, the machine shows particular behavior – see Table 3.5.

Table 3.5 Relation between parameter-value and behavior (of the machine from the example)

parameter value:	Effect: an initial value is selected from region:	The resulting eigenvalue is:
a	A	α
b	B	β
c	C	Non

In this example, the machine has one parameter with three different values. And, if the parameter-value is held constant at a (or b) for some time, the operation computes eigenvalue α (or β). A very important observation is now that the behavior of this machine depends on parameter-values. Parameters and their role in determining the behavior of a system have already been discussed in Chap. 2. It was stressed that a change in parameter-value meant a change from one transformation to another. The parameter-value selected another transformation. This is also what happens in this example – see Table 3.6, where the above reasoning is cast in the Ashby-notation.

Table 3.6 Relation between parameter-values and behavior (continued)

T_p	α	β	Some value from region C
P = a	α	α	α
P = b	β	β	β
P = c	Some value from region C	Some value from region C	Some value from region C

In the table, only the end result of the application of the operation is given (an eigenvalue, or some unknown value from region C). Such tables do not make apparent how the change in behavior comes about. What is specific about the machine from the example is that the change in behavior is brought about by selecting a new initial value. This is an important specification for observing systems, for which each sensory impression may serve as a (a constellation of) parameter-value(s) changing the computation of eigenvalues in the nervous system – see below.

From the example, it can also be learned that there are two types of "input": (1) input for the operation, and (2) input for the machine as a whole.

Input for the operation – i.e., the value used by the operation to compute its output – comes from the three different regions. The first input is the initial value (x_0), while all the subsequent inputs for the operation are the consecutive outputs of the operation. That is:

$$\text{The first input} = x_0$$
$$\text{The second input} = OP(x_0) = x_1$$
$$\text{The third input} = OP(x_1) = x_2$$
$$\ldots$$
$$\text{The nth input} = OP(x_{n-2}) = x_{n-1.}$$

"Input for the machine as a whole" is the value of a parameter. In the example, a, b or c are values of the input for the machine (because they are values of the parameter). So, a change in "input for the machine" (a change in parameter-value) means that "a new input for the operation" is selected. And, a change in input for the machine may lead to the computation of a new eigenvalue.

In cybernetics, the influence of the environment of a machine on its behavior is modeled by the effect of parameters (see also Chap. 2). One could say that the last type of input (the parameter-value) *triggers* the machine to "behave in a particular way" – i.e., to compute a specific (or no) eigenvalue.

The difference between the two types of input is important for several cyberneticians. They emphasize that parameter-input (from the "environment") is no input for the operation itself. Changes in parameter values ("changes in the environment") are "perturbations" in the sense that they trigger the (re) computation of (different) eigenvalues (cf. Varela 1984; Maturana and Varela 1980; Segal 1986).

That is, they displace the input-value of the operation to some other value, which, in turn, is used as an initial value for its recursive computations.

In all, parameter-values (representing the effect of the "environment") serve as input for the machine as a whole. They cause the machine's operation to use particular input (initial values) which may lead to particular eigenvalues. In this way, it can be said that a specific set of parameter-values may give rise to specific behavior of the machine.

Eigenvalues and the complexity of systems producing them

Until now, the formalism introduced for describing the behavior of closed systems was illustrated by referring to some very simple systems. However, to use the formalism to describe how we observe, it should be generalized to more *complex systems* (with a more complex operation, computing a large number of eigenvalues) in more *complex environments* (consisting of a large number of parameters).

What can be said about such complex systems is that different sets of parameter-values may cause them to compute (different) eigenvalues. An early example of the emergence of eigenvalues (or eigenbehaviors) in large complex systems is a series of computer-simulations performed by Ashby in the 1950s. In these experiments "...Ashby connected [...] 1000 non-trivial machines" (von Foerster, 1984, p. 19). All non-trivial machines performed logical functions on their input and used as their inputs the outputs of (other) non-trivial machines. Von Foerster continues: "...after setting them at an initial value [Ashby] let them loose." And, after a while, "the systems settled into various eigenbehaviors". These considerations cover, in fact, a central thesis of first-order cybernetics: stability arises through closure – the recursive application of operations within a system –whatever its complexity. Varela (1984, pp. 25–26) calls this his "principle 1: Every operationally closed system has eigenbehaviors". According to Varela, this principle is "derived from a combination of both empirical observations and cybernetic considerations" (ibid. p. 25). Elsewhere, Varela (1992, p. 21) is even more explicit about this: "Provided that there is any form of structural coupling with any form of contingency, a rich [self-organizing] network will establish regularities. It cannot but bring about regularities". So, the cybernetic claim is that regularities will emerge, even though the system in question is very complex, and even though it is confronted with a very complex environment.

The aim of Sect. 3.3 is to understand how von Foerster's treatment of observers ("observing systems") can help in appreciating the nature and origin of our pre-orientation with respect to selecting variables and parameters and to determining the transformation of the systems we regulate. Since von Foerster claims that the key to this understanding has to do with "eigenvalues, emerging in the closed cognitive system of the observer," we set out, in subsection 3.3.1, to explain what closed systems and eigenvalues are and how eigenvalues may emerge in closed systems, using von Foerster's formalism. In particular, we discussed how systems in which operations recursively act on their output (can) show eigenvalues and how the emergence of these eigenvalues depends on specific sets of parameter-values.

In the next section, we apply these insights to understanding cognition. In particular, we deal with von Foerster's idea that the cognitive system of any observer is a closed (nervous) system producing eigenvalues that are linked to cognitive content.

3.3.2 Closure, cognition and observing

In von Foerster's treatment of observing systems two steps can be discerned. The first is to explain that observation is based on our nervous system and how we should treat our nervous system as a closed system and the second is to argue how cognition can be treated as the result of "recursive computations in the closed nervous system." Below, we discuss both steps.

3.3.2.1 Closure of the nervous-system

To understand the closure of our cognitive system, von Foerster proposes to treat our nervous system (producing our cognitions) as a closed system. Von Foerster identifies the closure of the nervous system by referring to this system as a large whole of interconnected nerve-cells (e.g., von Foerster 1981).

Before it can be appreciated how this "large whole of interconnected nerve-cells" can be regarded as a closed system, it is helpful to have some basic knowledge about nerve-cells and their connections. A nerve cell (or neuron) can be said to consist of a cell-body, and a nerve-fiber (or axon) – see Fig. 3.3 (cf. Thompson (1976) or Lindsay and Norman (1977; p. 195 ff)).

A cell-body generates an electrical impulse that travels along the axon. An axon has one or more endings that connects it to the cell-bodies (or their extensions, called dendrites – not shown in the figure) of other neurons. If an electrical impulse reaches the axon-ending, a chemical neuro-transmitter substance is released which moves across the small gap between the axon-ending and the cell-body or dendrites of another neuron (cf. Lindsay and Norman 1977, p. 197). This transmitter-

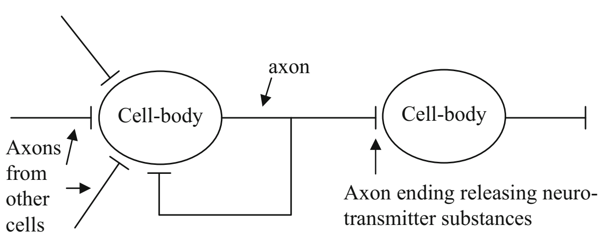


Fig. 3.3 Schematic representation of a nerve-cell

substance serves as input for the other neuron. Typically, neurons may connect to many other neurons and they may also connect to themselves – in this case the axon-ending of a neuron connects to its own cell-body (or its extensions) – see also Fig. 3.3. Based on the transmitter input of a set of neurons, the electrical potential of a cell-body may reach a threshold causing the cell-body to generate an electrical impulse to move along its axon. In such a case a cell-body is said to "fire." Since an impulse traveling along an axon means that transmitter substances are passed on to another neuron, we can also take the cell "fires" or the cell "does not fire" as shorthand for the cell's output.

So, basically, each cell can be regarded as an operation that computes, on the basis of its inputs (transmitter substances, caused by the electrical impulses of (other) nerve-cells), an output (which is again transmitter-substance caused by electrical impulse traveling across its own axon, that serves as an input to (other) nerve-cells). So, a nerve-cell may be modeled by:

$$\text{Output } A_{t+1} = \text{OP}(\text{input } X_t,\ \text{input } Y_t,\ \text{input } Z_t,\ \ldots)$$

Here Output A_{t+1} means: the output of nerve-cell A produced at moment t+1 is calculated by OP on the basis of several inputs: input X_t (from nerve-cell X at moment t), input Y_t, etc. (one of these inputs may also stem from nerve cell A itself). Because nerve-cells also have internal states, relevant for calculating their output, they can be viewed as non-trivial machines.[6] Now, if all these machines are coupled (in the sense of the Ashby-experiment – see earlier) closure emerges because the output of the nerve-cells at one moment is the input for nerve-cells at another. The nervous-system can thus be viewed as a huge whole of interconnected non-trivial machines. Or, alternatively, as a large non-trivial machine, embodying a closed operation, operating on (large) patterns of electro-chemical impulses. This is, of course, a rather simplified representation of what is going on in the central nervous-system. However, the crux of the matter is that it models the nervous system as a closed system in which eigenvalues may emerge – which is von Foerster's main thesis. To see this, consider Table 3.7

Table 3.7 Representation of impulse-patterns at different moments in time

		Moment:						
		T	T+1	T+2	T+3	T+4	T+5	...
Nerve – cell :	1	0	0	0	1	1	1	
	2	1	1	1	1	1	1	
	3	1	1	0	1	0	0	
	...	...	...	...	...	...	...	...
	N	0	1	0	1	0	1	

[6] In many formal representations of nerve-cells, such internal states are represented by threshold-functions.

At each moment in time a nerve-cell produces output – it either fires (and releases transmitter-substances) (1) or not (0). In addition, one may also consider the output of all nerve-cells together. In the table, a column represents the whole pattern of nerve-cell output at a specific moment in time. A rather complex operation can be defined that transforms an impulse-pattern at some moment in time into another pattern at the next moment in time (resulting in the change of one column into the next). This is a closed operation, because its output (a particular impulse pattern, resulting from the joint operation of all its nerve-cells at one moment) is also its input (at a "next" moment). This is captured by Fig. 3.4.

$$\mathrm{OP}(\text{impulse pattern}_t) = \text{impulse pattern}_{t+1}$$

Fig. 3.4 Recursive determination of impulse-patterns (cf. von Foerster, 1981, p. 273 ff.)

According to the central thesis of first-order cybernetics, in such a system eigenvalues may emerge... This means that some impulse-patterns (or parts of it) may stabilize in time – i.e., invariant configurations of impulse-patterns may occur. The nature of these invariances is not elaborated in this text. It is clear that it does not amount to a simple repetition of entire impulse patterns. Instead, it may have something to do with invariances (eigenvalues) emerging in parts of these patterns.

To understand how these invariances come about and to appreciate their role in understanding cognition, von Foerster draws our attention to three additional systems to which the nervous system is connected: the sensory system, the motor-system and the endocrine system. All these systems participate in the generation of eigenvalues, and, hence, according to von Foerster, to the emergence of cognitive content.

To start with, the sensory system transforms sensory impressions from the environment into input for nerve-cells and may thus be regarded as a huge complex of parameters to the nervous system. For instance, every time we "see" an object the sensory system transforms electro-magnetic waves from the environment into a pattern of input-values for the optic nerves. In this way, a part of the environment (certain electro-magnetic waves) serves as a set of parameters for the sensory system. And, the sensory system, in turn, translates these environmental parameters into a pattern of input-values – it translates the environmental parameter-values into a pattern of parameter-values that can affect the nervous system. Based on this pattern of input values for the optic nerves, the nervous-system may compute (an) eigenvalue(s), which is/are associated with "seeing the object (in its environment)." In general, every constellation of sensory parameter-values (which, in turn, corresponds to certain environmental stimuli) may give rise to the calculation of eigenvalues in the nervous system. Moreover, different constellations of parameter-values may lead to the computation of different eigenvalues in the nervous system. This relation is depicted in Fig. 3.5.

The nervous system is also coupled to a motor system. That is, nerve-cell output, or constellations of nerve cell output, connects to motor-nerves, which may cause different kinds of motor activity. Put differently, the nervous system acts as a set of

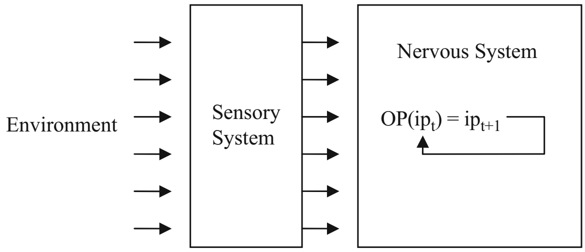


Fig. 3.5 Environment as a set of parameters for the sensory system; the sensory system triggers recursive computation of impulse-patterns in nervous system (see text)

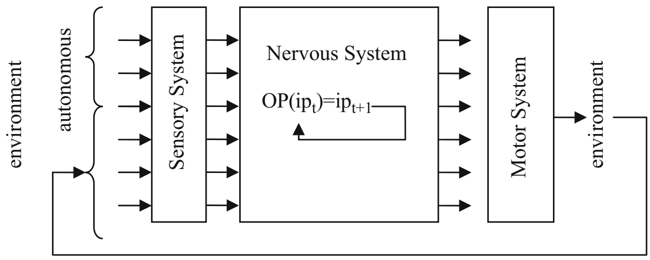


Fig. 3.6 Interplay between environment, sensory, nervous and motor-system

parameters to the motor system. More specifically, each eigenvalue, calculated by the nervous system, may result in specific behavior (-al patterns). This behavior, in turn, has an effect on the environment (see Fig. 3.6). So, if an eigenvalue is reached under specific sensory conditions, and this eigenvalue gives rise to a specific behavioral pattern, it may be said that the connection between sensory and motor system is mediated by the nervous-system.

The behavior produced by the motor system has an effect on the environment and thus alters the constellation of parameter-values affecting the sensory system. Moreover, the environmental parameter values may also be changed by "autonomous" factors – i.e., factors that do not depend on the system's motor activities. These factors represent autonomous environmental changes. In Fig. 3.6 these two different sets of environmental parameters are indicated by the distinction between autonomous environmental parameters and the environmental changes due to motor output (arrow below).

The changed parameter-value constellation may trigger the calculation of yet another eigenvalue or it may induce the nervous system to recalculate the same eigenvalue. The newly calculated eigenvalue may, in turn, connect to specific behavior, that again, may lead to a change in sensory parameter-values, etc. According to von Foerster, this continuous cycle of perception (setting the sensory parameter-values), computation in the nervous system, behavior, perception, computation constitutes another form of closure of the nervous system (von Foerster, 1981). This system, von Foerster states, "recursively processes what it perceives" (cf. von Foerster 1981, p. 294 ff.). This is a very important point in von Foerster's theory because in this cycle the "robustness" of eigenvalues vis-à-vis different constellations of parameter-values is tested, so to speak. In this process, the stable eigenvalues ultimately remain.

To illustrate this process, suppose that the environmental input for the sensory system (a constellation of parameter-values) at a specific moment in time (t) is abbreviated with: P_t. This constellation of parameter-values may trigger the computation of an eigenvalue – say EV_t. And this eigenvalue may connect to motor-output M_t.[7] With these abbreviations we can interpret the sequence as presented in Fig. 3.7.

This sequence indicates that the constellation of sensory parameter-values at

Fig. 3.7 Representation of parameters relating to eigenvalue and to motor-output

$$P_1 \longrightarrow EV_1 \longrightarrow M_1$$

moment "1" causes the nervous system to calculate eigenvalue EV_1, which causes the motor system to produce behavior M_1 (Fig. 3.7). In addition, behavior M_1 may change the parameter-value constellation P. As we have said earlier, the set of parameter-values P can also be changed by an environmental sub-system generating "autonomous" changes to the parameter-values (which will be indicated with A; their influence on the parameter-values at moment t with A_t.) The sequence then becomes as in Fig. 3.8.

This indicates that both A_2 and M_1 may cause the parameter-value constellation

$$\begin{tikzcd}
A_1 \arrow[r] & P_1 \arrow[r] & EV_1 \arrow[r] & M_1 \arrow[dll] \\
A_2 \arrow[r] & P_2 & &
\end{tikzcd}$$

Fig. 3.8 Representation of parameters relating to eigenvalue and to motor-output (continued)

to change. The changed parameter values, in turn, may lead to the calculation of eigenvalue EV_2. In Fig. 3.9 a chain of these events is given.

In this chain several consecutive eigenvalues may be computed that are identical/comparable (i.e., $EV_i = EV_{i+1} = EV_{i+2} = \ldots$). In such a case the same eigenvalue is recomputed. This means that this eigenvalue is (re) computed in the face of (at least) two different sets of parameter-values. In other words: the eigenvalue endures the changes in sets of parameter-values. An interpretation of this may be that the system as a whole "learns to treat different environmental circumstances as the same". Or, alternatively, the system "learns" about the robustness of an eigenvalue. This mechanism enables "observing systems" to

[7] Note that only the index in P_t refers to a moment in time. The other indices indicate that both the eigenvalue and the motor output are based on P_t.

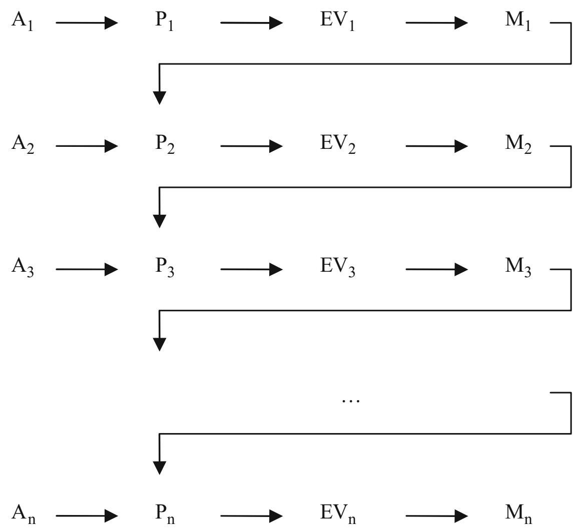


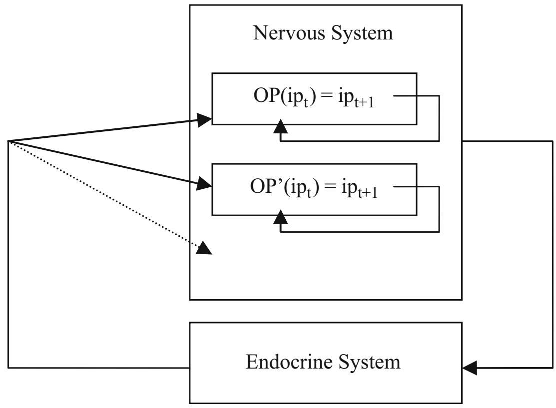
Fig. 3.9 Representation of parameters relating to eigenvalue and to motor-output (continued)

associate certain eigenvalues with many different parameter-value constellations, something that is crucial for the emergence of cognition – as we will discuss below.

The last system, contributing to the emergence of eigenvalues in the nervous system is, as von Foerster explains, the endocrine system. The nervous system is connected to the endocrine system that also produces neurotransmitter substances that travel (via the bloodstream) into the nervous system and influence the way nerve cells operate. For instance, if we experience a threat, our nervous system causes the endocrine system to produce adrenaline (a hormone acting as neuro-transmitter), which, in turn, reaches nerve-cells via the bloodstream and causes all kinds of behavioral and emotional responses. Von Foerster calls this connection the "synaptic-endocrine" closure of the nervous system. (cf. Segal, p. 133; von Foerster 1981, p. 304 ff). By virtue of this connection, the nervous system recursively changes (via the endocrine system) its own modus operandi. In this way, it participates in changing (selecting) its own operations – see Fig. 3.10. In terms of the above formalism and its extensions, this amounts to selecting another (set of) operation(s) –which may compute different eigenvalues. Put somewhat more picturesquely: due to this connection the nervous system is capable of reprogramming itself.

In this section we discussed how the nervous system, with its closed organization computes eigenvalues, and how several other systems are related to this process. In Fig. 3.11 an overview of all the participating systems in the process of the emergence of eigenvalues is given. Although this argument shows how eigenvalues may emerge in the nervous system, it only leaves us with *neuronal* eigenvalues – i.e., invariants at the level of impulse patterns. What we need to describe observing systems, however, is the emergence of cognitions (in terms of observations, experiences, thoughts, etc.). We treat this issue in the next section.

Fig. 3.10 Relation between nervous system and endocrine system



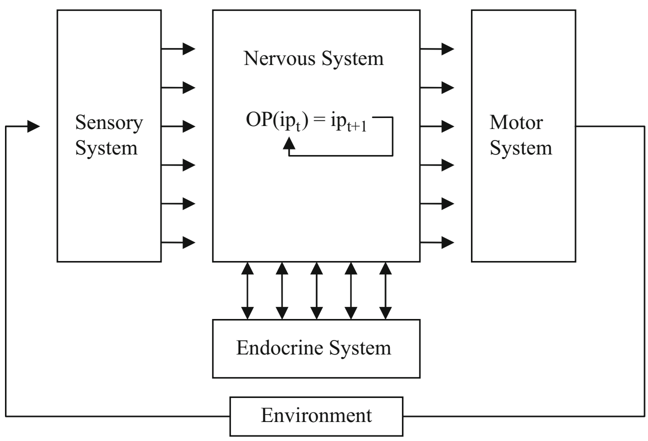


Fig. 3.11 Overview of systems participating in the emergence of eigenvalues

3.3.2.2 The emergence of cognitive content

In our daily lives we have all kinds of sensory experiences – we see all kinds of colorful objects, hear all kinds of sounds, experience different tastes, etc. Von Foerster explains, however, that our sensory receptors do not encode the "quality" of these experiences. That is, all sensory receptors do is translating electromagnetic waves, pressure waves, etc. into impulse patterns (that only differ in intensity and periodicity). The "quality" we experience (e.g., a color or a sound) is not encoded by our sensory receptors – but is something that emerges in our experiential world (von Foerster 1981). In this context, von Foerster often refers to the "principle of undifferentiated coding" (e.g., von Foerster and Poerksen 2002; von Foerster 1992), i.e., the idea that our senses do not encode the nature of that which stimulates them.

A particular taste, for instance, can be triggered by eating something (stimulus a) or by using "an electrode stimulating a taste bud with a view volts" (cf. von Foerster and Poerksen, p. 18). Since both ways of stimulating nerve cells may produce the same taste experience von Foerster explains that the "nature" of the stimulus is not encoded – it remains unspecified what caused the particular taste: "The only thing we know is that there is a stimulus" (ibid p. 18).

From this, von Foerster derives a fundamental question regarding cognition: how does our brain enable us to experience (a) reality? This is what he calls the "problem of cognition." And he immediately proposes an answer: since the qualities of our sensory experiences are not encoded by the sensory receptors, it is clear that the nervous system is organized in such a way that it *computes* these qualities (von Foerster 1993, p. 71). In several of his papers he introduces this "problem of cognition" and presents his way of dealing with it in terms of the emergence of eigenvalues in the closed nervous system. However, as we stated above, in this explanation these eigenvalues emerge at a neuronal level – not at a cognitive level. There still seems to be a gap.

To deal with this gap between the two levels, it is possible to postulate a connection between states of the nervous-system and "cognitive contents." This connection is widely accepted. Many psychologists, philosophers, neuroscientists etc. strongly believe that there is some form of mapping of patterns of impulses in the brain and our experiences, thoughts, etc. Supposing this association between neuronal states and cognitive contents, it is a *cybernetic* insight that these cognitive contents emerge because of the recursive computation of eigenvalues in the nervous-system – triggered by parameter-values. Now, if the eigenvalues emerging in the nervous-system are seen to be associated to cognitive processes (and to our experiences and other cognitive contents), one may state that – as von Foerster does (1981, p. 306): "The nervous-system is organized in such a way that it computes a stable reality."

In a modest attempt to tie nervous-system eigenbehaviours to cognitive processes he states: "I have presented my ideas on molucular calculatory processes only to indicate that there are perspectives that hint at the contribution of molecules to conscious thought" (von Foerster 1991, p. 93, our translation).[8]

Von Foerster relies on the association between neuronal states and cognitive states when he presents his "(gr)aphorism," (see Fig. 3.12) capturing the essence of his theory (cf. von Foerster 1981, p. 296):

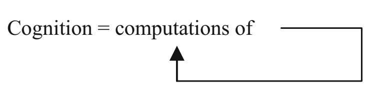


Fig. 3.12 Von Foerster's view on cognition

[8]"Ich habe meine Vermutungen über molekulare Rechenprozesse nur vorgelegt, um anzudeuten, dass es Perspektive gibt, die auf eine Mitwirkung der Moleküle an dem großen Drama des bewussten Denkens hindeuten [. . .]" (von Foerster 1991, p. 93).

Given the above reasoning about the closure of the nervous-system and given the association between eigenvalues in the nervous-system and our cognitive contents – this may now be understood.

Throughout his work, von Foerster elaborates the statement that cognition is a process of recursive computations. In one particular example, he shows how this is related to the cognitive development of children. He draws from the work of Piaget when he states that the experience of the invariance of objects (object-constancy as Piaget calls it) only slowly emerges in a continuous process in which sensory "activities" like seeing, feeling and touching an object are correlated to motor-activities with the object (chewing on it, throwing it away, etc.), that, in turn, change the sensation of the object, etc (e.g., von Foerster 1992, p. 141). Von Foerster sees this process as a recursive computational process, in which object-constancy emerges as an eigenvalue.

To explain this, he describes (p. 149 ff.) an infant playing with a ball. The child interacts with the ball in several ways – it sees it, reaches towards it, touches it, moves it, etc. The results of the actions of the infant are perceived by it, and based on this perception new actions are performed, again changing its perception of the ball. Von Foerster proposes to treat the child as operator (1981, pp. 273 ff.), acting on what we, as observers, see as a ball. Through the recursive application of the operation, von Foerster explains, an eigenvalue emerges that enables the child "to experience the ball as an invariant" (cf. Segal 1986, p. 142). In this case, the infant reaches, what von Foerster calls "sensori-motor competence" regarding the object (1981, p. 281 – that is: "the motor behavior that must arise in order for the child to develop a stability of behavior with the object. Then it makes full use of the object, controls the object, knows what to do with the object" (Segal 1986, p. 138). In fact, this argument can be cast in the notation we used to describe the connection between sensory, nervous and motor system – see Fig. 3.13. Sensori-motor competence arises when successive eigenvalues in the figure "equal" each other – i.e., when the child learns to "calculate" the "same" eigenvalue in many different circumstances. It then learns, in this example, that something (we as observers call a ball) remains constant in different circumstances.

Von Foerster proposes to take the name of the object (as it appears to us, as observers) as the name for the eigenbehaviour, emerging when a child reaches "sensori-motor competence." (von Foerster 1992, p. 68). This may seem strange – but he argues: the toy (ball, rattle, etc.) is something that we as observers can identify when an infant plays with it. For the child, however, there is no such "toy" – it only "sees," "understands" it as the invariant emerging through its sensory-motor interactions with it. He explains: "Ontologically, Eigenvalues and objects, and likewise, ontogenetically, stable behavior and the manifestation of a subject's "grasp" of an object cannot be distinguished. In both cases "objects" appear to reside exclusively in the subject's own experience of his sensori-motor coordinations" (1981, p. 280).

Glanville (2003, pp. 97–98) summarizes the importance of these insights for understanding cognition: "von Foerster showed us a model of a mechanism that would generate [. . .] stable-under-change objects". This mechanism of recursively

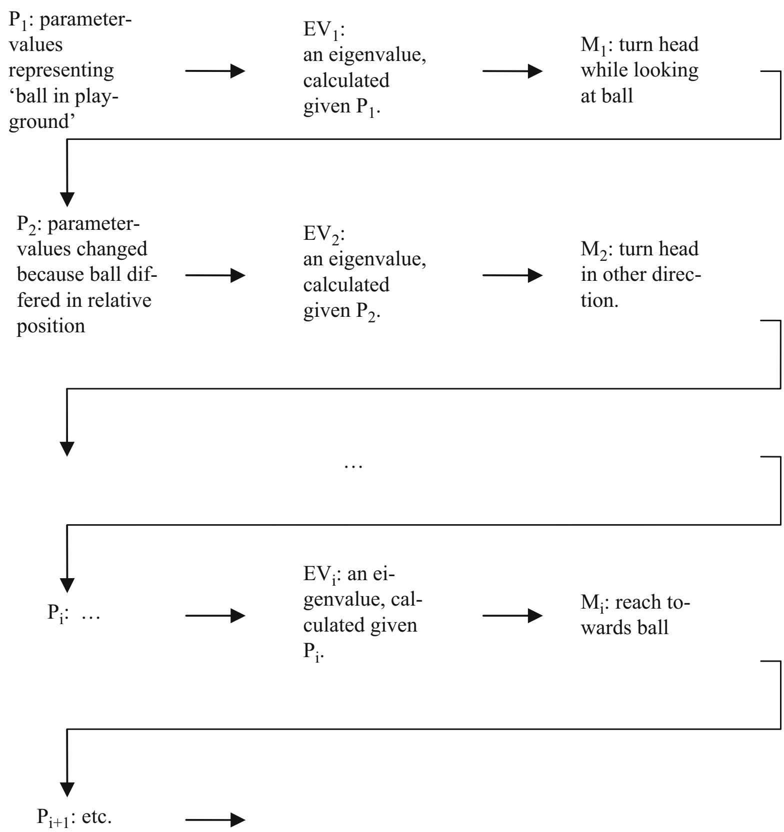


Fig. 3.13 The emergence of "sensori-motor competence" (example)

computing our percepts "shows us a way in which we can understand how [. . .] we may develop constancy in our experience [. . .]." In sum, von Foerster asserts that through processes of recursive computations our experiential world emerges. Moreover, as von Foerster argues, once experiential invariances appear (i.e., once we reach "sensori-motor competence" regarding an object) they can be referred to in language and thought (they can, for instance, be given a name; inferences can be made regarding the objects, etc.). In this way, higher order cognitive processes also depend on the experiential invariances we produce.

Based on these ideas one could claim that all knowledge is brought about by recursive computations in the nervous system. This holds for experiences of objects or events, but also for "higher order cognitions" such as remembrances of objects or events, or associations regarding them, inferences, conjectures, models, theories, etc. All cognitive content is the result of recursive computations. This insight can be taken to be the central thesis of second-order cybernetics and is the main contribution of von Foerster's work.

3.3.3 Problems of First-order Cybernetics Revisited

In the previous section we presented von Foerster's central thesis that cognition is a process of recursive computation. This means that cognitive content emerges due to recursive computations in the nervous system. Moreover, these computations are triggered by a specific set of parameter-values, which changes because of our actions, which, in turn, are based on our perceptions, which are triggered by a specific set of parameter-values, etc. In this subsection we will use this thesis to discuss the two problems of first-order cybernetics concerning the definition of a system and its behavior, i.e., how do we select the variables and parameters for defining a system and how do we determine its behavior?

As may be derived from the previous sections, a specific focus on the *observer* of systems, i.e., as a closed cognitive system, is central to von Foerster's treatment of the two problems. As discussed, the observations or experiences, arising within this system are the result of recursive computations; they are "emerging invariances" in our cognitive system. In particular: an object emerges in our experiential world when we reach "sensori-motor competence" – i.e., when an eigenvalue settles regarding an object. This is von Foerster's way of describing how an observer comes to "see" an object – in fact how any system is "observed." Defining a system (and its behavior) in its environment is thus tied to emerging eigenvalues in the nervous system. All observed systems, as they appear in our experiential world, are tokens for our "sensori-motor competence." They represent eigenvalues regarding the system, which come about in a process of recursively dealing (observing, acting) with it, as described earlier. Given a particular (phylogenetic and ontogenetic) history of interactions of an observer with his/her environment a specific "observed" system as invariant emerges. In this sense, observers build up a pre-orientation towards the systems we "encounter" consisting of preferences with respect to describing them and their behavior. And, observers cannot but "use" this emerging description per hypothesis to deal with the system.

The problems of first order cybernetics now disappear. These problems arose, essentially, because observers were not to refer to prior knowledge when they defined systems and described their behavior. In this view, selecting variables, parameters and transformations describing a system's behavior is an "unguided" process of trial and error. Von Foerster would argue that such prior knowledge cannot be excluded – but, instead, that it is an essential aspect in defining (and dealing with) the non-trivial machines we encounter in our daily lives. Due to the eigenvalues emerging in the process of interacting with non-trivial machines variables and parameters (associated with the emerging eigenvalue) present themselves immediately and implicitly due to our interaction with the system. In the same fashion, the emerging eigenvalues point at how to see the behavior of the system and how we can deal with it.

The process in which eigenvalues emerge is not a conscious process – only the results of the process can be experienced. So, when we start to deliberate about the definition and behavior of systems, we already have at our disposal a particular implicit definition. However, this does not mean that we cannot change what presents itself to us. What we can do – in a deliberative or reflexive mode – is to reconstruct variables of the system that presents itself to us, make inferences about them and change them. And if we deliberatively change a description of a system, it may change our interaction with it, which, in turn, may change our pre-orientation with respect to this kind of system.

In all, defining systems is not a matter of only consciously searching for variables and parameters and using them to describe their inner workings. Dealing with the non-trivial machines we encounter in our daily lives, is based on the hypothetical invariances that emerge in the process of observation and action. We should acknowledge that we have and cannot escape a pre-orientation towards all the systems we observe. In contrast with Ashby, then, who argued that observers should not refer to prior knowledge when defining systems (which, essentially, causes the two problems) von Foerster emphasizes that observers and their prior knowledge cannot be excluded when they define systems.

3.3.3.1 The Hypothetical Nature of Knowledge

As discussed, von Foerster holds that our experiences are produced by our nervous system itself – although triggered by constellations of parameter-values. Put differently: our experiential world is a recursively computed construction, triggered by specific sets of parameter-values. In this line of reasoning a difference is made between the realm of parameters and the realm of experiences/cognitions. Parameter-values trigger the computation of cognitions. So far, nothing special has been said – we just rephrased an aspect of his theory. Somewhat more striking is that von Foerster claims that it seems that *nothing else* can be said about the realm of parameters except that they trigger the computation of cognition (cf. von Foerster and Poerksen, 2002) We can only experience the results of the recursive computations triggered by the parameter-values – not the parameters that triggered them. In this sense, von Foerster's theory may be seen as a contribution to epistemology. His theory is one in which, as Luhmann puts it: "the theory of self-referential enclosed knowledge, for the first time, gets a form in which the inaccessibility of the outer world "in itself" can be expressed" (Luhmann 1990, p. 33, our translation).[9]

If it is accepted that a specific constellation of parameter-values triggers the computation of some cognitive content, one could say, as we did above, that a change in this constellation tests its robustness. The originally triggered eigenvalue is (implicitly) used "per hypothesis": if it holds, the system "learns" about its stability (i.e., given the new set of parameters the same eigenvalue is calculated) – if it does not hold the system learns about the borders of the stable region (the recalculation of the same eigenvalue is not allowed by the new parameter-value constellation).

[9] "die Theorie des selbstreferentiellen, in sich geschlossen Erkennens erst jetzt die Form erwinnt in der [. . .] die Unzugänglichkeit der Außenwelt "an sich"[. . .] zum Ausdruck gebracht [werden kann]".

Put differently: the implicit hypothesis regarding the applicability of a specific eigenvalue, given a change in parameter-values, is rejected. Each time an eigenvalue is calculated, it is used (implicitly) per hypothesis "to endure change" – Schütz (cf. Schütz and Luckmann, 1994) might say that they are constructs, used (calculated) "bis-auf-weiteres". This leaves us with the conclusion that knowledge is *essentially* hypothetical – allowed and tested (during our phylogenetic and ontogenetic development) by large sets of parameters-values (different environments). Moreover, different histories of interaction of the closed "observing system" with its environment (sets of parameter-values) may lead to different eigenvalues, and hence, to different experiences. So, not only is knowledge essentially hypothetical, but also many different hypotheses may exist regarding the same set of parameter-values – i.e., all those eigenvalues that are *allowed* by it. In this sense these eigenvalues are contingent – they could have been different.

"Essentially hypothetical and contingent" knowledge does not mean we can always "experience" things "differently." In fact, one may identify degrees of "stability" of eigenvalues. If an experience is "firmly" stabilized, it means that is has endured many different sets of parameter and that it is (virtually) impossible to "experience" it otherwise – or even downright nonsensical to talk about experiencing it otherwise. For instance, most sensory experiences cannot be experienced differently (although, of course, there are such things as illusions). This also holds for many of the assumptions we have about the world – premises like "all men are mortal," or "one needs food to survive" cannot be negated seriously. However, many other concepts may appear to be less stable – and can be changed. For instance, "problem-definitions" or "solutions to problems" are rarely firmly stabilized and need to be open for change.

In organizations, many of the "models-of-the-world" we use to select goals, infrastructures and regulatory schemes fall, in our view, in the category of such unstable constructs. Models of the organization, its strategy, its competitors, customers, personnel, of the efficiency of processes, of the effect of ICT on the efficiency of a process, and so on, are volatile constructs, and, hence, the selections we base on them are *risky* (see below).

Given the presented outline of von Foerster's theory, we are now in a position to discuss the main contribution of von Foerster's theory to understanding "organizations as social systems conducting experiments", i.e., a description of the risky nature of these experiments. An elaboration of this understanding will be presented in the next section.

3.4 Von Foerster and the Risky Nature of the Organizational Experiment

An important implication of von Foerster's theory is that all knowledge is essentially hypothetical and contingent. This is a relevant implication for the main thesis of this book, because it says something about the nature of our perceptions and knowledge of and in organizations – knowledge we use to set organizational goals, design infrastructures, transformation processes, and the schemes we use to regulate these processes operationally. These are all hypothetical constructs, "stabilized within" organizational observers awaiting refutation. And, although they have been stabilized, their effect cannot be determined with certainty, neither a priori nor a posteriori. This definitely gives a specific meaning to describing organizations as *risky* experiments!

In this section, we set out to explain what this "von Foerster related risk" means. To do so, it is helpful to distinguish two types of understanding risk that is associated to organizational experimentation: a "common sense understanding" and a "fundamental (von Foerster type of) understanding" of risk. After discussing these types of understanding risk (and explaining the kind of risk attached to organizational experiments according to von Foerster), we turn our attention to some consequences for dealing with this risk in organizational experimentation.

3.4.1 A Common-sense Understanding of Risk

In a sense, every manager already knows that organizational decisions are risky. If a new product is launched, there is always the risk – despite all intelligence, marketing research or campaigns – that sales disappoint. If a new method for monitoring and rewarding human resources is implemented there is always the risk that it does not work as well as it was intended to. Every manager knows that one cannot know everything with respect to organizational decisions, and, so, one cannot foresee all possible circumstances and disturbances. Therefore, one always runs the risk of failure. In one word, as common sense already has it: organizational decision making is risky. In this sense, von Foerster's reasoning can be said to be rather superfluous, for in day-to-day practices the risk attached to decisions seems to be accepted as quite natural.

The common sense view of risk can be taken one step further to include the risk attached to the models we use in organizational decision making. In the organizational experiment, decisions concern the selection of goals, infrastructural elements and regulatory schemes with respect to transformation processes in organizations. Cast in Ashby's terminology our organizational decisions concern control, design and operational regulatory activities with respect to organizational transformation processes. Selections with respect to all three organizational experimental activities (setting goals, designing an infrastructure and regulate operationally) all require *models* relating behavior (in terms of selected variables) to specific parameter-values. Our "experimental decisions" are based upon such models containing, in essence, the hypotheses tying parameters to behavior.

For instance, in order to set goals for an organization or an organizational process, one needs to establish an idea about the "effectiveness" of these goals, which, in essence, is based on a model tying these goals to their intended effect (i.e., to the meaningful survival of the organization) given certain environmental circumstances (which may facilitate or hinder the realization of the goals). The same holds for designing an infrastructure and for operational regulation; both activities require a model relating relevant variables and parameters. To design an infrastructure, a model is needed to determine (build hypotheses about) how each infrastructural element can contribute to attenuate disturbances or to amplify regulatory potential. Without, for instance, a model describing the effect of a particular division of work (a parameter) on the efficiency of a transformation process (a variable to describe the behavior with respect to some process), it is impossible to make a decision about how to divide work. Or, without a model describing the effect of a particular HR-instrument on the motivation of human resources (parameter), it is impossible to select it as an infrastructural instrument attenuating disturbingly low levels of motivation. Similarly, to regulate processes operationally one needs models describing the effect of particular regulatory activities on actual disturbances.

In short: as part of conducting organizational experiments (select goals, design an infrastructure and regulate operationally) one needs models relating the behavior of variables to (constellations of) parameter-values. These models enable us to make predictions about the contribution of a set of goals, infrastructural elements and operational regulatory activities, and to make decisions, based upon these predictions. With respect to these models, the common-sense view of risk rests upon two notions: (1) we cannot know everything (in advance), and so our models are necessarily incomplete, and (2) we make all kinds of mistakes when we build these models and make predictions based on them. The first notion is straightforward: since we cannot know everything, "unforeseen" events cannot be included into our models and can therefore always affect the outcome. The second notion is backed up by a large amount of psychological research, showing that we are not very good at tasks, needed to make "good" models and predictions. Such tasks include estimating the probabilities of intended and unintended outcomes of alternatives, explaining causes of success and failure, judging whether events or problems are similar, and specific types of logical reasoning. We are prone to make all kinds of mistakes when we try to perform these tasks, e.g., errors in judgment, errors in perception, errors in selection; miscalculations of probabilities; logical errors; and so on; (see, for instance, Kahneman et al. (1982), Nisbett and Ross (1980), or Hogarth (1994) for extensive treatments of such errors). These biases and shortcomings may lead to an *increased* decisional risk; for if we are unaware of them, we may believe that our models were helpful in predicting success or disaster, while in fact, they were not.

So, a common-sense explanation of risk, attached to our organizational models, includes the idea of imperfectness in two ways: (1) unforeseen events can always disturb things, and (2) our models (and the decision we base on them) are problematic due to our biases and shortcomings.

3.4.2 A More Fundamental Understanding of Risk

The common-sense idea of (the causes of) risk seems to be commonly accepted – in literature as well as in practice. Less obvious, however, is that our organizational "constructs" (the variables, parameters we select, the cognitive "models" we employ to select them and understand their connection, and our pre-orientation towards variables and parameters) are *fundamentally* uncertain and contingent, and that this contingency is something that belongs to the nature of "decision making in organizations." This puts risk at the heart of understanding the organizational experiment. Let us explain.

Based on von Foerster's ideas, it can be said that (1) the selected variables, parameters, and models relating them, as well as (2) our implicit pre-orientation, are hypothetical and contingent, and therefore risky. They are all more or less stabilized eigenvalues, recursively calculated in a process of environmental coupling. To elucidate this, Fig. 3.14 may be helpful.

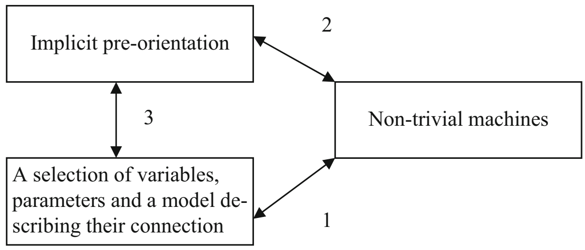


Fig. 3.14 Dealing with non-trivial machines

The figure depicts the relation between the variables, parameters and model we need to deal with non-trivial machines (most of the "systems" we encounter in our daily lives) and our implicit cognitive pre-orientation towards these machines. To deal with any non-trivial machine, we have to select variables and parameters and make a (more of less explicit) model to understand their behavior in terms of these variables and parameters – arrow 1. This is just common Ashbian regulatory logic. However, as von Foerster explained, because these machines are non-trivial, we cannot be sure about the completeness and correctness of such models. We can only employ them by hypothesis. Based on von Foerster's theory it can also be derived that selecting "good" or "relevant" variables and parameters is based on our pre-understanding of the non-trivial machine (arrow 2). These are more or less stabilized "cognitive eigenvalues," we tacitly and implicitly use to understand the machine and which have been developed based on our previous interactions with it (or with allegedly similar "machines"). This pre-understanding itself is also necessarily contingent – because of the machine's non-triviality (and because of the necessarily hypothetical nature of these eigenvalues). Moreover, based on this implicit pre-orientation, there may (given some purpose) still be different possible ways to describe a non-trivial machine explicitly – from which a relevant one should be chosen (arrow 3).

The above reasoning introduces a kind of risk that is fundamentally attached to regulating non-trivial machines – risk that comes in two ways (1) risk due to the contingency of the implicit pre-orientation (in the course of interacting with some non-trivial machine only a *particular* implicit understanding of it has stabilized), and (2) the risk due to the contingency of the particular explicit selection of variables, parameters and the model describing their relation (given some purpose).

At first sight, this may all seem a rather academic exercise, but it has important consequences for organizational experiments and for the practice of managing organizations. For example, if we go back to the example of the model, required for setting goals, one might stress (referring to von Foerster's ideas) that this model does *itself* contain hypothetical and contingent models of all kinds of systems – all kinds of non-trivial machines like competitors, customers, financial institutions, governments, etc. In our interaction with these non-trivial machines (competitors, customers, etc.) a model of them (relating particular variables to describe their behavior and parameters affecting them) has emerged; but just because of their non-triviality we cannot be certain about these descriptions; and so we have to content ourselves with contingent models of them. This introduces a kind of risk that cannot be taken away by including more "knowledge" into our models, by training or by becoming aware of it – for each inclusion of a description of yet another non-trivial machine also introduces the description's contingency and hypothetical nature.

It is important to see that the risk related to non-trivial machines is something different than the one implied by the common-sense notion of "the impossibility to know everything." That is, the latter "just" seems to suggest that, given our models-of-the-world, there is always a possibility that some unforeseen event will occur. Von Foerster would probably not disagree, but his theory points at a more fundamental kind of risk: the model-of-the-world *itself* necessarily includes contingent, hypothetical models of non-trivial machines. And, although some way of understanding them may have stabilized, our understanding of them is fundamentally uncertain: it is impossible to gain full knowledge of the inner workings of non-trivial machines. Decisions, then, become at least "twice" as risky as compared to the common sense notion: unforeseen events may always happen *and* the models on which we base our decisions are themselves hypothetical contingent models.

In all, von Foerster might say that models may stabilize in the course of time, in the sense that they are used successfully in different circumstances (as stable eigenvalues they "endure different sets of parameters" and enable us to structure our experiences repeatedly). But they are and remain contingent. And, because of their contingency, our models of the world contain risk, for they could have been different. They are always *hypotheses* of what the world is like and because they are hypotheses, they might be rejected. Every (organizational) decision, then, always has the form of a hypothesis because it contains a contingent selection based on a necessarily imperfect, contingent "model of the world."

3.4.3 Two Consequences of von Foerster's View on Risk

Given the treatment of the risky nature of conducting organizational experiments, two important consequences for the practice of "experimenting" in organizations (i.e., for the practice of setting goals, designing infrastructures and regulating operational processes) can be discerned: one concerning the danger concerning "reification and reduction" and one concerning "responsibility."

3.4.3.1 Reification and Reduction

A first consequence is that one should be aware of and accept the contingency and hypothetical nature of our experimental "selections." This entails that one should try to uncover this contingency and avoid trivializing "the world." Revealing contingency could be accomplished, quite simply, by, for instance, actively trying to construct alternative models; or by inviting others to comment on your models and decisions. This requires a particular infrastructure – for instance, an infrastructure (1) enabling experimenting with alternative approaches to problems, (2) securing an atmosphere in which making mistakes is tolerated – for they are necessarily part of conducting organizational experiments, or (3) supporting active participation in discussions about models and decisions. Conversely, one may also list infrastructural features that may frustrate revealing the contingency of our models and decisions. For instance, in an infrastructure in which work is broken up into small short-cycled tasks there is often not much left to experiment with (because the object of the task only contains a very small fragment of the whole transformation process). More (and less obvious) examples of infrastructural configurations enabling or frustrating the awareness of the contingency of our models and decisions can be given, but we will postpone this discussion to Chap. 7, in which we treat the division of work in detail.

Even though we will get back to the role of the infrastructure with respect to conducting the experiment; a specific aspect of the technological part of the infrastructure deserves attention in this section, because it may play a role in "trivializing the world." Many methodologies exist (either ICT-supported or not) to help managers to deal with problems in organizations (cf. Rosenhead 1989, for overviews). Among these are "decision analysis," tools derived from (subjective) expected utility theory, or from game-theory, "multiple criteria analysis," "system-dynamics," "soft-system methodology," or "scenario-analysis."

Although most of these methodologies have been used successfully, some of them – especially the ones that are mathematically oriented – harbor the danger of trivialization and may install a false sense of "exactness" with respect to the models we rely on when we make decisions. Many of these decision-aids require that we express our ideas of value, of what we think that counts as problematic, and of uncertainty (with respect to the occurrence of events) into variables, constraints, probabilities, rules, or equations. This may trivialize our way of dealing with problems because it may lead to (1) reification and (2) unjustifiable reduction.

Reification is problematic if a decision-situation is "frozen into" a model that – once established – remains unchanged. This can happen, for instance, if one builds a model for deriving "an optimal solution." To arrive at this solution, the assumptions on which it is based have to be fixated. Of course, modelers may come to realize that some variable should be added to the model, and once it is included, the calculation may restart. But after a while one may feel that one has built a model that "really captures the essence of the problem," a model that no longer needs to be changed. Reification can also come about in non-mathematical modeling – if one fixates, for example, a specific formulation of a problem, e.g., in order to start working on solutions.

Of course, at some point in time it is very helpful to fixate the formulation of a problem, but, as many authors point out (Ackoff 1978; Rittel and Webber 1973), if one does not allow for reformulating the problem, this can easily lead to solving the wrong one. Reification, then, trivializes the world in the sense that it fixates certain assumptions about the world into a model, which may turn out to be problematic if changes do occur, or if other problem formulations could also have been used. This sort of trivialization conceals the basic contingency of the models we actually use to make decisions.

The problem of reduction may occur when we are forced to express our experiences with respect to values, uncertainty, or to what counts as problematic or as an alternative into variables, probabilities, rules or equations. The essence of a model is that it only contains a limited number of variables and relations (so reduction always occurs), but one may ask (along with many philosophers and authors in the field of decision making) whether our experiences can really be captured by them (cf. March 1978; Winograd and Flores 1986; Dreyfus and Dreyfus 1986). This doubt gives rise to acknowledging the *problem* of reduction, which occurs if our cognitions, feelings, intuitions, etc. with respect to problematic situations (and their solutions) are reduced to only a limited number of model-elements that may not capture the essence of them (e.g., when "what we consider to be of value" is translated into quantifiable variables; or that our values are treated as compensatory; or when our intuitions with respect to future events are translated into a single probability-measure; etc.). Based on their idea of expertise, Dreyfus and Dreyfus summarize the problem of reduction as follows: "conventional mathematical modeling [. . .] requires that the expert informants, whose expertise is supposed to be captured in the model, provide the sort of decomposed and decontextualized information that concerns beginners, but not experts" (1986, p. 177). So, the main issue with reduction is that the model-representation falls short of representing the way we experience aspects of a problem.

This paints a rather gloomy picture of the many decision-aids used in management, and we should perhaps say that it is a bit overstated. Reduction and reification do not have to be problematic if one is aware of them. However, installing this awareness is seldom a high priority of proponents of the model or consultants using it. Most decision aids are used to deal with risk in the common-sense way: it is stressed that they are helpful in constructing and evaluating alternatives and that, by doing so, they reduce uncertainty. However, they always do so while fixating and reducing "reality as we experience it," and they rarely focus on questioning the contingency of our assumptions/models.[10] If we are unaware of these problems, these decision aids may install a false sense of certainty, suggesting an unjustifiable

[10]Of course, a large number of methods for divergence of ideas (e.g. tools for fostering creativity) exist (and are) used for supporting problem-solving. The main focus here is on more traditional decision aids e.g. system dynamics or multi-criteria analysis, that have a long tradition as management decision-aids.

belief in the model's predictions while ignoring the fundamental uncertainty attached to our experiences.

In all, some parts of the technological infrastructure (to which these decision-aids belongs) may hinder uncovering the contingency of our models-of-the-world – and may even play a role in concealing it.

3.4.3.2 Responsibility

Above, we dealt with the contingency of our models of the world as a consequence of von Foerster's model for the practice of experimenting in organizations. We will now turn to a second consequence, related to the responsibility for our decisions. To discuss responsibility for our decisions, it is necessary to see that making a decision involves *choosing from a set of options*. Besides this rather obvious aspect, it also involves *constructing* the set of options to choose from. Now, in a nutshell, responsibility for a decision can mean two different things. It can be (and normally is) described as "being responsible for choosing a particular alternative from a set of options." This view of being responsible for a decision entails that one willingly and more or less deliberately (i.e., based on a (moral) assessment of the situation) chooses a particular course of action (from a set of possibilities). If one, for instance, chooses a particular alternative that turns out to be harmful, someone can be held responsible for the decision and its consequences, if this alternative is chosen willingly and if these harmful consequences could have been foreseen (or were already known). In this view of responsibility, selecting willingly *from a particular set of alternatives* is at stake.

However, responsibility can also refer to the distinctions we make that bring about the set of options in the first place. In that case it refers to using particular preferences, cognitions, feelings, intuitions, etc. to interpret some situation as a situation in need of a decision and to its associated "solution space." This extended type of responsibility is related to the problems of reification and reduction we have mentioned above. In order to make a decision, one *must* reduce and reify – one simply cannot go on deliberating for ever. However, this may become problematic if one does not investigate the consequences of a particular reification and reduction and if one does not try to deal with negative ones (which may occur for different reasons – such as ignorance or opportunism).

So, even though one cannot escape reduction and reification, one can at least try to come up with and compare different interpretations of a problem and different "sets of options" associated with them. This is reflected by the idea of "extended responsibility" of our decisions: the responsibility for interpreting a particular situation as problematic situation and formulating its associated solution space. If we reexamine the above notion of being responsible for a decision as choosing willingly and more or less deliberately (i.e., based on a (moral) assessment of the situation) from a set of possibilities, it can be said that such responsibility should also apply to the models-of-the-world/distinctions we use to define and assess the situation.

In the case of experimenting in organizations, this means that responsibility for experimental selections refers to two things. It refers to a particular selection (choosing a particular goal; a particular infrastructural aspect), and it refers to the choice for a particular interpretation "of the world" based on which a particular set of options appear. If one knows that such "ways-of-seeing" are contingent, one also knows that one has to *select* one, a choice for which one can be held responsible, and which entails the obligation of assessment.

The main theme in this chapter is to discuss the risky nature of our selections in the course of the organizational experiment. As we have argued, every (organizational) decision always has the form of a hypothesis because it is a contingent selection based on a necessarily imperfect, contingent "model of the world." Moreover, in this last section we pointed at two consequences for organizational experimenting.

But, even though one agrees with von Foerster's reasoning and accepts its consequences for conducting the organizational experiment – his ideas are not yet sufficient to describe organizations. They help in uncovering the risky nature of the experiment, but they do not treat their social dimension. In fact, his theory may be regarded as a specific theory of knowledge creation, and describes how cognitive content may emerge at the level of the individual. And in this sense it can be used to describe how individuals in organizations select, apply, and reselect goals, infrastructural aspects and regulatory schemes. But, as we have argued in the introduction of this book, such selections in organizations are not a purely individual matter: they come about in organizational communication. Or, put it in von Foerster's terminology: the emergence of individual eigenvalues is co-determined by processes of social interaction (communication). One might say that the opinions, reflections, interpretations, intuitions, emotions, etc. of *others* should be taken as input (parameters) to the individual process of knowledge stabilization; and at the same time, our own reflections, emotions, opinions, interpretations, etc. are input to the process of knowledge stabilization of others. Thus viewed, a *network* of individuals is defined, in which the emergence of cognitive content of each individual is co-dependent on the output of the members of the network. And, in the spirit of von Foerster, one might introduce the stabilization of cognitive content with respect to the network as a whole – i.e., a process in which *individual* eigenvalues stabilize despite and/or because of the relations they have with the other members in the network. In such networks, then, the individual stabilization of hypotheses is co-dependent on processes of communication. A relevant question for understanding organizations, then, is how "models" and "selections" relevant for the organizational experiment come about (stabilize) "intersubjectively" in processes of social interaction.

In some texts, von Foerster explicitly deals with the issue of stabilization in social networks and the role of communication (e.g., von Foerster, 1984) in which he discusses different network-configurations with respect to their suitability to support the process of stabilization. However, to understand organizations as social systems in which this kind of stabilization takes place one needs to understand the concept of social interaction (communication) and what processes of stabilization at the level of the network entail. And, although von Foerster's theory provides some basic concepts for going into these issues, we will use Luhmann's theory to discuss them – for, as will become apparent in the next chapter, his ideas explicitly treat the *social* "arche" of organizations.

References

Ackoff, R. L. (1978). *The art of problem solving*. New York: Wiley.

Ashby, W. R. (1958). *An Introduction to cybernetics*. London: Chapman & Hall.

Cybernetics & Human Knowing (2003) 10(3–4). Special issue on von Foerster

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over Machine*. Oxford: Basil Blackwell.

Glanville, R. (2003). Machines of wonder and elephants that float through air. *Cybernetics and Human Knowing, 10*(3,4), 91–106.

Hogarth, R. M. (1994). *Judgment and choice*. Chichester: Wiley.

Hughes, G. J. (2001). *Aristotle on ethics*. London: Routledge.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Lindsay, P. H., & Norman, D. A. (1977). *Human information processing*. New York: Academic Press.

Lipschitz, S. (1987). *Linear Algebra*. New York: McGraw-Hill.

Luhmann, N. (1984). *Soziale systeme*. Frankfurt am Main: Suhrkamp.

Luhmann, N. (1990). Das erkenntnisprogramm des Konstruktivismus und der unbekannt bleibende Realität. In N. Luhmann (Ed.), *Soziologische Aufklärung (5)* (pp. 31–58). Opladen: Westdeutscher Verlag.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. Reprinted In D. E. Bell, H. Raiffa & A. Tversky (Eds.). (1988) *Decision Making* (pp. 33–57). Cambridge: Cambridge University Press.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht: Reidel.

Maturana, H. R., & Varela, F. J. (1984). *The Tree of knowledge*. Boston: Shambhala.

Nisbett, R. E., & Ross, L. (1980). *Human inferences: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Pask, G. (1992). Different kinds of cybernetics. In G. van de Vijver (Ed.), *New perspectives on cybernetics* (pp. 11–31). Deventer: Kluwer.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169.

Rosenhead, J. (ed). (1989). *Rational analysis for a problematic world*. Chichester: Wiley.

Segal, L. (1986). *The dream of reality*. New York: Norton.

Schütz, A., & Luckmann, T. (1994). *Strukturen der Lebenswelt (I)*. Franktfurt am Main: Suhrkamp.

Thompson, R. F. (1976). *Introduction to physiological psychology*. New York: Harper & Row.

Varela, F. J. (1984). Two principles of self-organization. In H. Ulrich & G. J. B. Probst (Eds.), *Self-organization and management of social systems* (pp. 2–24). Berlin: Springer.

Varela, F. J. (1988). *Cognitive science: A cartography of current ideas*. Cambridge, MA: MIT Press.

Varela, F. J., Thompson, E., & Rosch, E. (1993). *The embodied mind*. Cambridge, MA: MIT Press.

von Foerster, H. (1970). Molecular ethology: An immodest proposal for semantic clarification. In G. Ungar (Ed.), *Molecular mechanisms in memory and learning* (pp. 213–248). New York: Plenum.

von Foerster, H. (1981). *Observing systems*. Seaside, CA: Intersystems Publications.

von Foerster, H. (1984). Principles of self-organization – in a socio-managerial context. In H. Ulrich & G. J. B. Probst (Eds.), *Self-organization and management of social systems* (pp. 2–24). Berlin: Springer.

von Foerster, H. (1988). *KybernEthik*. Berlin: Merve Verlag.

von Foerster, H. (1991). Was ist Gedächtnis, dass es Rückschau und Vorschau ermöglicht? In S. J. Schmidt (Ed.), *Gedächtnis* (pp. 56–95). Frankfurt am Main: Suhrkamp.

von Foerster, H. (1992). Entdecken oder Erfinden: Wie lässt sich verstehen? In H. Guman & H. Maier (Eds.), *Einführung in den Konstruktivismus* (pp. 41–88). München: Piper.

von Foerster, H. (2002). *Understanding understanding*. Heidelberg: Springer.

von Foerster, H., & Poerksen, B. (2002). *Understanding systems*. New York: Kluwer.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Norwoord, NJ: Ablex.

Chapter 4
The Social "arche," Organizations as Social Systems: Luhmann

4.1 Introduction

In the previous chapters, we explored the first organizational "arche," i.e., we discussed organizations as conducting risky experiments with meaningful survival. We based this discussion on insights taken from first- and second-order cybernetics.

In this chapter, we shift our attention to the second "arche," to organizations as *social* systems. As argued in Chap. 1, experiments in organizations are characteristically social. Selecting goals, designing infrastructures, and performing operational regulation are all communicative events belonging to the "system of connected communications" we call the organization.

This "addition" of the social character to the experiment changes nothing of what we said about control, design, and operational regulation in earlier chapters. At the same time, it also changes everything. Because organizations are a *particular* type of social systems, we need to *specify* the social character of the experiment.

For instance, in Chap. 2 we argued that to study and control complex systems, we need to determine goals, attenuate possible disturbances and amplify regulatory potential, and select regulatory actions to deal with actual disturbances. In Chap. 3, we learned that machines that are dependent on their own history are synthetically determined and analytically indeterminable or that defining variable systems for the purpose of regulation is a risky process involving "observers." However, we do not yet know what all of this means for organizations as social systems. In the present chapter, we explore this question by explaining the social systemic character of organizations.

To this purpose, we discuss the work of the German sociologist Niklas Luhmann (1927–1998). Luhmann is regarded by many as one of the most important and prolific writers on sociology of the twentieth century. However, the importance of Luhmann's work for the development of sociology is not the decisive reason for selecting his work for discussion in this chapter (we could have selected authors such as Bourdieu or Giddens as well). What is decisive is that Luhmann's work is firmly rooted in the tradition of system theory and (second order) cybernetics.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_4, © Springer-Verlag Berlin Heidelberg 2010

More in particular, Luhmann takes concepts developed in cybernetics and modifies them to conceptualize organizations as social systems. By selecting Luhmann's work, we are in a position to combine a deep and up-to-date conceptual understanding of the social character of organizations with an application of insights from cybernetics.

To formulate the social "arche" of organizations, the rest of this chapter is structured into three sections. In Sect. 4.2 we discuss Luhmann's general theory of social systems by focusing on his major work "Soziale Systeme" (Luhmann 1984; in the rest of the text we refer to the English translation from 1995). Section 4.3 deals with the social "arche" of organizations in terms of Luhmann's theory of organizations. In this section, we argue that organizations are a particular type of social systems consisting of *decisions*. This argument is based on Luhmann's papers on organizations (Luhmann, 1988b, 1993; Luhmann 1998) as well as on his book "Organisation und Entscheidung" (Organization and Decision), which appeared after his death (Luhmann 2000). In Sect. 4.4, we prepare our discussion in Chap. 5 of the link between Luhmann's ideas about the social "arche" of organizations and the experimental "arche" of organizations.

4.2 Luhmann's General Theory of Social Systems

This section is devoted to a discussion of relevant concepts from Luhmann's "general theory of social systems" ("Soziale Systeme," 1984; "Social Systems," 1995). The reason for paying attention to this "general theory" is that, according to Luhmann, organizations are a *particular type of social systems*. Because of the complexity of Luhmann's work, it is perhaps a good idea to charter the theoretical landscape we intend to navigate. Figure 4.1 can function as our "roadmap." The topics we discuss in this and the next section are printed in bold.

In the figure, we can see that the starting point is a "system." Note that the systems at the center of Luhmann's theoretical endeavor are quite different from the systems we discussed in Chaps. 2 and 3.

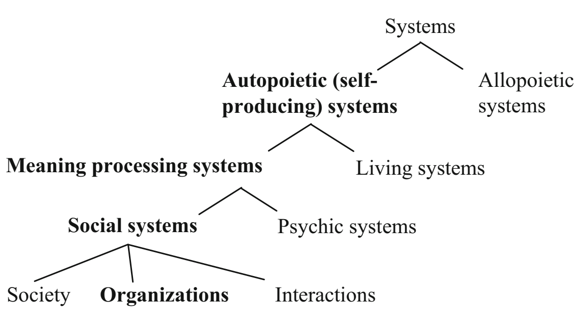


Fig. 4.1 Luhmann's classification of systems (adapted from Greshoff, 1999, p. 70)

In these chapters, we dealt with *variable* systems defined by an observer designed to study or control the behavior of "things" in the world.

Luhmann's theory is not about variable systems. It is about *real* systems that exist in their environment. Examples of such systems are "chairs," "tables," "organisms," "social systems" or "organizations." Unlike variable-systems, these systems do not consist of "variables," i.e., features of real systems or "things" selected by an observer as relevant for some purpose. The systems discussed by Luhmann consist of "*related elements"*; they are *element-systems*. Of course, element-systems have features that can be selected as variables or parameters by observers to study and control their behavior. Within the class of systems Luhmann distinguishes between "autopoietic" and "allopoietic" systems.

According to Luhmann, organizations belong to the class of autopoietic (or self-producing) systems. To explain how organizations as social systems produce themselves, we take the class of "autopoietic" or "self-producing" systems as our point of departure. In Sect. 4.2.1., we describe the general process of self-production. Next, we follow the tree down to the level of organizations, discussing "meaning processing systems" (Sect. 4.2.2), "social systems" (Sect. 4.2.3) and "organizations" in Sect. 4.3.

4.2.1 Autopoietic or Self-Producing Systems

Autopoietic systems are "self-producing" systems; this is what the term "autopoiesis" literally means: "auto" = "self," "poiesis" = "production." The term "autopoietic system" was coined by Chilean biologists *cum* cyberneticians: Humberto Maturana and Francisco Varela. In their book "Autopoiesis and Cognition" (1980) and in other publications, Maturana and Varela (1980) describe "living systems" as autopoietic, i.e., self-producing, systems. Luhmann "borrowed" this concept and applied it to the domain of social systems.

To make this application possible, Luhmann first "generalized" Maturana and Varela's concept of "autopoiesis." This means that he "stripped" it from its biological roots, formulated it in general terms so that it could be applied to other types of autopoietic systems. In terms of Fig. 4.1: Luhmann reformulated the concept of autopoiesis that was originally developed for the level of "living systems" in *general* terms so that it could be used for all types of "autopoietic systems." After generalizing it, he *specified* the general concept of autopoiesis for the class of so-called "meaning processing systems," "social systems," and the different types of social systems, such as "society," "organizations," and "interaction systems."

As all real systems, autopoietic systems have their own existence in their environment. Moreover, as all real systems they consist of related elements. However, it is characteristic of them that they exist as *self-producing systems*. This means that the related elements out of which they consist continuously produce new related elements. To quote Luhmann (1988a, p. 283), an autopoietic system is a system that "reproduces the elements out which it consists by means of the elements out of which it consists” (Luhmann 1988a, b, p. 283; our translation).[1] For example the cells constituting a multi-cellular organism are related in a way that allows for the production of new related cells. By actually producing new cells, the organism reproduces itself, staying alive in its environment. Luhmann argues that the same holds for social systems. They maintain themselves in their environment because their elements, “communications,” produce new elements, “communications.”

To explain autopoietic systems we discuss (1) their elements and how they are produced, (2) their regulatory structures, and (3) their emergent characteristics.

4.2.1.1 Elements and the Production of Autopoietic Systems

The elements of an autopoietic system, to use a simple metaphor, can be regarded as its “building blocks.” This notion of building block has to be taken in two senses. Elements are building *blocks*: they are what the system consists (is built) of. At the same time, elements are *building* blocks: elements actively producing (building) new elements.

As building *blocks,* elements are what the system consists of. Given some autopoietic system of a particular type, its elements constitute the system. For instance, an organism consists of cells as its elements, or a social system of communications. Because, as we will explain below, these elements are the smallest *systemic* units that in relation with each other produce new units of the same kind. If one would “divide” an element of an autopoietic system into its ingredients, destroying its internal structure, the element would stop to be a *systemic* element, and the system would perish. By themselves, the ingredients without being arranged into elements cannot be the system’s building blocks.

The “indivisibility” of elements does not mean that elements of self-producing systems do not have an inner structure or that their constitution does not depend on a complex of conditions. They do. Elements of self-producing systems presuppose all kinds of “ingredients” and complex causal processes between them and these “ingredients” and causal processes do play a role in the production of elements by elements. However, the elements of an autopoietic system cannot be reduced to these “ingredients.” Although elements of an autopoietic system require a *particular* arrangement of “ingredients” and causal processes between them, they also have emergent features, i.e., features that only occur at the level of elements and cannot be reduced to their ingredients in isolation. For instance, “cells” of a multi-cellular organism consist of complex interacting intra-cellular components that play a crucial role in the reproduction of cells. However, taken by themselves, i.e., without their particular interrelations, these components are not yet a cell, and do not function as building blocks of the organism. In general, relative to an autopoietic system, its elements are the ultimate units the system consists of.

[1] “das die Elemente, aus denen es besteht, mit Hilfe der Elemente, aus denen es besteht, reproduziert” (Luhmann 1988, p. 283).

Elements are not only the system's building *blocks* in the "passive" sense that, in the end, real systems consist of related elements. They are also "*building* blocks" in an "active," productive sense, for they play a crucial role in the production of new related elements, and thereby, in the reproduction of the system as a whole.

To understand what this means, it is important not to misunderstand the term "building block." It must not be taken as if the elements of self-producing systems are "there," "given" as elements of the system, before or independent of their relation to other elements (e.g., like a stone that exists independent of its being built into a wall). Rather, their existence depends on their being part of the system they are building blocks of. To explain, the following (considerably simplified) example may help.

Suppose that alien bacteria enter a multi-cellular organism. The cells of this system "recognize" the bacteria as "alien." The bacteria do not "fit" into the reproductive relations between the cells that constitute the organism. As a result the bacteria are rejected as elements of the system. They are "complexes of things and processes" (consisting of membranes and intra-cellular components) that are not accepted *as* elements contributing to the reproduction of both new elements and thereby of the organism as a whole. The system's own cells, on the contrary, are "complexes of things and processes" that *are* generated, recognized, and connected *as* elements of the network of related elements.

This example points at two things. First, elements are not just "there" independent of their relation to other elements. They are "complexes of things and processes," or, in short: "complexes," that somehow have to be generated, recognized, and connected *as* elements by other elements, to actually become an element of the self-producing system. They are not "there" as elements, before or independent of their relations to other elements. The reverse is true; it is by being involved in the productive relations between elements that "complexes" become elements of self-producing systems.

In the second place, if we say that self-producing systems are systems consisting of "related elements producing related elements," it is exactly this process of being generated, recognized and connected as an element by related elements constituting the system that is intended.

"Elements producing elements" can be understood as follows.

A. In the world self-producing systems exist, i.e., real systems consisting of related elements producing related elements. These systems can maintain their existence, given all kinds of material conditions and processes – in short material conditions (Fig. 4.2a).

For instance, there are organisms consisting of cells producing cells or social systems consisting of communications producing communications. These systems maintain themselves in their environment by means of elements producing elements. They can do this given all kinds material conditions, allowing for the production of elements by elements.

B. From the material conditions, elements of the system "*generate*" "complexes" that have an inner structure and emergent features allowing them to be recognized *as* elements and connected to the system. These "complexes" with the right inner structure are "potential systemic elements" (see Fig. 4.2b).

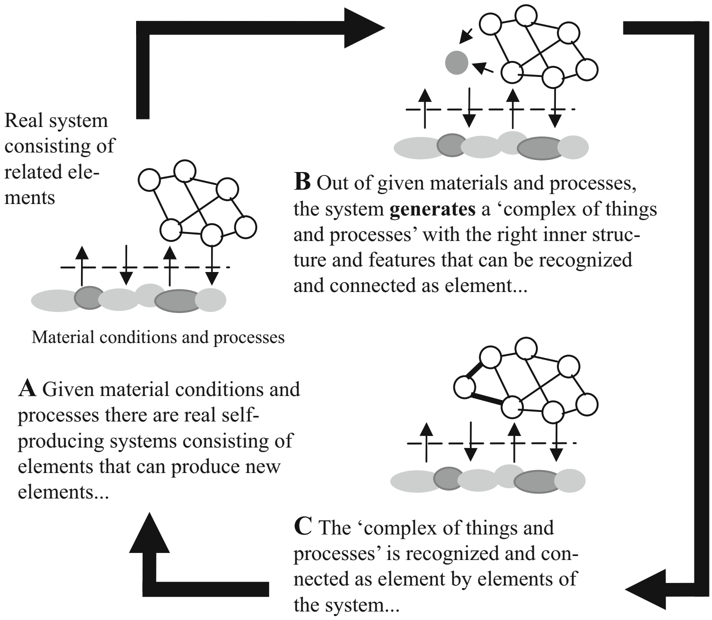


Fig. 4.2 Elements of a self-producing system, producing a new element

For instance, given all kinds of conditions, cells of an organism contribute to the "generation" of "complexes" with a particular inner structure and particular features. Given their inner structure and emergent features these "complexes" can be recognized *as* cells, and connected to the network of cells constituting the system. Other "complexes" (e.g., "alien" bacteria) do not have the required inner structure and features. They cannot be recognized *as elements* and connected to the network of related cells.

C. The "complex of things and processes" is actually recognized as element and connected to the set of related elements. If this happens, the *potential* systemic element becomes an *actual* systemic element. As such, it can contribute to the "generation," "recognition" and "connection" of new elements (see Fig. 4.2c).

For instance, if a "component" has the "right" inner structure and is recognized *as* a cell and connected to the network of cells constituting the organism, it can take part in the "generation," "recognition" and "connection" of new cells.

The process of "elements producing elements" is a cyclic process (as indicated by the bold arrows in Fig. 4.2). Once a new element is produced, it becomes a factor in the production of new elements. Moreover, from what has been said until now, it follows that elements cannot exist in isolation. They are elements by virtue of their relation to other elements. For instance, once a cell is taken out of an organism, it is no longer an element of the organism. For a short while it may remain a "complex of things and processes" with the "right" inner structure and features (that under strict conditions may perhaps be re-inserted as an element). However, as a loose "complex of things and processes" taken out of its autopoietic context, the cell tends to disintegrate, loosing its inner structure, so that it even stops being a "potential element."

To summarize, "elements producing elements" means that given all kinds of pre-existing materials and processes, elements "generate" "complexes of things and processes" – with the "right" inner structure and emergent features – that are recognized and connected as actual elements to elements constituting the system. From now on we use the term "production" to refer to this longer explanation.

The above explanation of autopoietic systems provides an answer to three basic questions:

> Question 1: "What is an autopoietic system?"
> Answer: "An autopoietic system is a system that maintains itself in its environment by means of the production of the elements it consists of *by* the elements it consists of"
>
> Question 2: "What are elements of autopoietic systems?"
> Answer: "Elements of autopoietic systems are the smallest units the system consist of, producing new elements"
>
> Question 3: "What does it mean that elements produce elements?"
> Answer: "Related elements 'generate' 'complexes of things and processes' that, given their inner structure and emergent features, can be and actually are recognized as elements and connected to the system, contributing to the production of new elements"

4.2.1.2 Structures Regulating Autopoietic Production

We explained that an autopoietic system consists of elements producing elements, but we have not yet discussed how this production is *regulated*. However, it is easy to imagine that regulation is needed. For instance, imagine that the production of cells by cells in a mouse is not somehow regulated. This may lead to at least two problems.

First, suppose that there is nothing to regulate the generation, recognition or connection of "complexes" as cells. In this case, there would be nothing to prevent the generation of "alien things" lodging themselves in the mouse's organism, possibly, doing their destructive work from within. Somehow, the difference must be made between "complexes" that *can* contribute as cells to the mouse's self-production and "complexes" that *cannot*.

Second, suppose that there is also nothing regulating *which* cells can link up with *which* cells at what moment in time. Nothing would prevent that the mouse grows, for instance, ears in its entrails. Or nothing would stop the unbounded growth of, for instance, the mouse's nose. Apparently, if the self-production of the mouse is to continue, not all possible cells of the mouse should be generated, recognized and connected to all possible cells all the time. If the mouse's autopoiesis is to continue, there must be a "mechanism" selecting which particular cells should be generated, recognized, and connected to which particular cells at what moment in time.

In general, we can say that if the autopoiesis of a system is to continue, "something" must regulate that:

1. out of all the "complexes of things and processes" that can be produced those should be selected that can contribute as possible elements to the autopoiesis of the system;
2. out of all possible elements that can be produced those particular elements are selected that should be generated, recognized, and connected to particular elements at a particular moment in time.

Luhmann calls this "something" the system's *structure*. According to Luhmann it is the function of the structure of an autopoietic system to constrain the set of all possible "complexes of things and processes" that can be produced by the system to a subset of *particular* elements that are *likely* to keep the autopoietic process going, and, therefore, should be generated and recognized by and connected as elements to *particular* elements of that system at a particular moment in time.

Luhmann formulates the regulatory function of the system's structure in terms of "constraining complexity." Without structure all "complexes" that can be produced can count as elements. We call this the set of all possible elements. The variety of this set is called V[Possible]. However, if the system's autopoiesis is to continue, only particular "complexes" should link up as particular elements with particular elements (at a particular moment in time).

The structure of a system *constrains* the variety of "complexes" that can be produced by the system to a subset of elements that are likely to contribute to the autopoiesis of the system, and, therefore, should be produced by particular existing elements. We call this the set of "likely" elements (its variety is V[Likely]). If the system's autopoiesis is to continue V[Likely] should be smaller than V[Possible]. In this sense, a system's structure regulates the autopoietic production of elements by elements by functioning as an enabling *constraint* on the set of all "complexes of things and processes" that can be produced by the system (V[Possible]), limiting it to a set that should be produced at a given moment in time by given elements of the system (V[likely]).

Taking into account the constraining role of the system's structure, the production of an element by the system's related elements can be described as follows (see also Fig. 4.3).

1. At a particular moment in time, the system consists of a set of related elements. At this moment, the system has a particular structure; let us call it S_t (= the system's structure at time t).
2. This structure constrains the set of all possible "complexes of things and processes" that can be produced by the system (set A in the figure) to a subset of elements that (1) are likely to contribute to the system's autopoiesis, and, therefore, (2) should be generated and recognized by and connected as new elements to particular existing elements of the system (set B in the figure).
3. As a rule, a potential new element out of this subset of elements is generated.
4. This element is recognized and connected as a new element of the system.

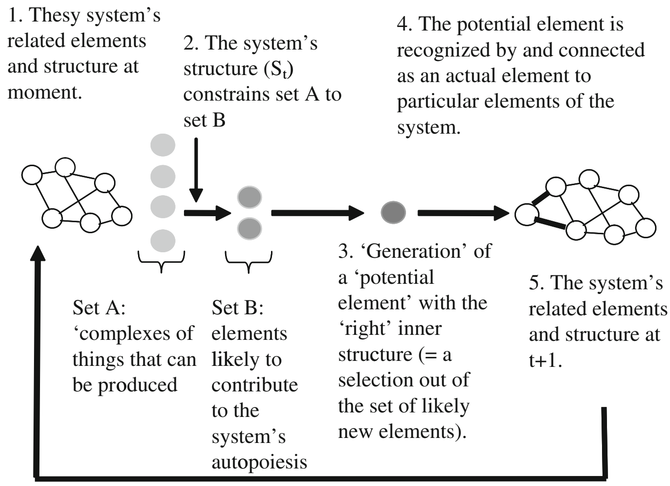


Fig. 4.3 The system's structure and the production of elements by elements

5. These related elements embody the system's structure at the next moment in time (S_{t+1}) and the process loops back to 1.

In our explanation, we emphasized the word "likely" to underline that the reproduction of an autopoietic system is a *risky* business. A system's structure does not *guarantee* its continued autopoiesis. Something may be "wrong" with its structure, allowing for the production of elements that are harmful instead of beneficial to it. In this sense, autopoietic reproduction can be described as a risky experiment with the system's structure as a "hypothesis" about what elements should be produced to maintain the process of self-production.

The system's structure regulates the production of elements by elements. But where is it located and how does it develop?

According to Luhmann, a system's structure is not a self-existent "thing." It cannot exist apart from the related elements constituting the system. It is rather an aspect or a dimension of these related elements. Therefore, it is impossible to destroy an autopoietic system's elements (and their relations) and keep its structure in tact.

Moreover, a system's structure is produced by and co-evolves with the related elements of the system. *New* elements may introduce changes to the system's structure: due to the production of new elements, the structure of a system at time = t+1 may differ from that at time = t. This new structure implies a change in the regulation of the production of elements by elements. Of course, it is also possible that new elements do not introduce changes to the system's structure. In this case, elements come and go while the system's structure, and thereby the regulation of its self-production, remains the same.

In sum, we can say that the structure of an autopoietic system takes care of the regulation of the production of its elements by its elements. It is produced by and co-evolves with the system's elements. It takes care of its regulatory function by constraining the set of all "complexes of things and processes" that can be produced by the system to a subset of elements that are likely to contribute to the system's autopoiesis, and, therefore, should be produced.

4.2.1.3 Emergent Characteristics of Autopoietic Systems

We know now what *elements* of autopoietic systems are, how elements of such systems *produce* new elements, and that systemic *structures* regulate the production of elements by elements. However, elements, structures, and processes are only "aspects" of autopoietic systems, but what can be said about these systems as such?

1. Autopoietic systems are real systems existing in a world of material conditions and processes. From this world, these systems "take" what they need for their processes of self-production. In this sense, autopoietic systems both need and belong to the "furniture of the world." However, within the constraints of the causal processes within this world, autopoietic systems, by means of their own structure, regulate their own self-productive processes.
2. By means of these self-productive processes, autopoietic systems produce and maintain their *boundary* with their environment. This boundary is both a result of and a condition for the autopoietic reproduction of the system. In a way, autopoietic systems can be described as systems that actively produce and reproduce their boundary with their environment.
3. *Within* (and *including)* this boundary we find the related elements, structures, and production processes constituting and reproducing the autopoietic system. These elements, structures, and production processes determine what the (relevant) environment of the system is; i.e., which "material conditions and processes" are either needed for the autopoietic production of the system or can influence the system's development.
4. *Within* its boundary, an autopoietic system can be decomposed in two ways. The first is into elements and their relations. This kind of decomposition focuses on *system complexity:* on the number of elements and of their possible relations. The second is into subsystems. This kind of decomposition focuses on *system differentiation*, i.e., on parts of the system devoted to dealing with particular problems. For instance, in a mammal, subsystems may be devoted to dealing with the "outside" temperature.
5. On the *outside* of the autopoietic system's boundary, we find its environment. As indicated, the system's elements, structures, and production processes determine the (relevant) environment of the system. Everything outside the system that is needed for the autopoietic reproduction of a system or that can influence its development belongs to its "environment." In this sense, each autopoietic system has its "own" environment (although environments of autopoietic systems can overlap). So, according to Luhmann, a system's environment is not "something" existing "before" the system as that "in which" the system exists.

It has no "unity" or "being" independent of the system. It is not a system, but rather a system's correlate.

6. An autopoietic system (say system "A") may find other (autopoietic or allopoietic) systems or things (say, "B," "C," and "D") in its environment. These other systems or things may either be needed for system A's autopoietic reproduction or they may influence the system's development. As such, they are a part of the "material conditions and processes" related to the autopoiesis of system "A." Drawing from Maturana and Varela, Luhmann calls this relation between system "A" and these systems and things, a relation of "*structural coupling*." So, if an autopoietic system A needs things/systems B, C, or D for its autopoiesis or if these things/systems influence A's autopoiesis, system A is said to be structurally coupled to systems/things "B," "C," and "D." It is also possible that autopoietic systems exist in *each other's* environment, conditioning each other's processes of self-production (e.g., symbiosis). Luhmann calls this relation of *mutual* conditioning between autopoietic systems "*interpenetration*." Systems that interpenetrate co-evolve in their environment. The development of the one is bound up with that of the other. As will become clear below, according to Luhmann, entire classes of systems may depend on interpenetration.
7. We explained that to become an element of an autopoietic system, a "complex" must be "produced" that has an inner structure and features that can be recognized and connected as a systemic element by the system's elements. Now, suppose that some "complex" is wholly *unlike* the elements of a given system. In this case, Luhmann argues, it could not be recognized as a potential element and, as a result, it could not be connected as an actual element of the system. In general, Luhmann argues that autopoietic systems of a different type have elements that are so dissimilar that they cannot become an element of another type of system. In short: autopoietic systems of a particular type are "closed" to elements of other types of autopoietic systems. Each type of system has its "own" type of elements connecting to its own type of elements.

Based on this last feature, there are three ways to characterize different types of autopoietic systems. A first way is to describe key properties of their elements. Because structures of autopoietic systems are bound up with their elements, different types of autopoietic systems also have different "mechanisms" functioning as structure. Thus, a second way of characterizing autopoietic systems is to describe the characteristics of the *structure* regulating the production of elements by elements. Finally, because the autopoietic production of elements by elements is governed by the elements and structures of the system, processes of autopoietic production will also differ among system types. So, a third way of describing the specificity of autopoietic systems is to describe characteristics of their *processes of self-production*.

In the subsections below, we characterize the three system-types we need to discuss to explain the social "arche" of organizations by describing key properties of their elements, structure, and processes of self-production.

4.2.2 Meaning Processing Systems

In the previous section we discussed Luhmann's general concept of autopoiesis. However, according to Luhmann, organizations are a *particular* type of autopoietic systems. So, we need to specify the general concept in such a way that it fits the particular elements, structures, and processes of self-production of organizations. A glance at Fig. 4.1 shows that we need to take three steps for this. Starting with meaning processing systems, we need to discuss social systems, and finally organizations. In this section, we take the first step and explain what "meaning processing systems" are. To this purpose, we address two key properties of the elements of meaning processing systems and we characterize their structure and the process of their self-production.

4.2.2.1 Two Key Properties of the Elements of Meaning Processing Systems

In Fig. 4.1 we can see that both social and psychic systems belong to the class of meaning processing systems. The main reason for classifying these two system types into this overarching class is that their elements share two characteristic properties:

1. The elements of all meaning processing systems, i.e., the elements of all social and psychic systems, have a particular "inner structure." Luhmann calls this inner structure "meaning."
2. The elements of all meaning processing systems are "event-like."

Below, we discuss these two properties.

The Inner Structure of the Elements of Meaning Processing Systems: Meaning

To explain what Luhmann understands by "meaning," we can inspect the inner structure of a communication (as indicated, communications are the elements of social systems, and social systems belong to the class of meaning processing systems). For instance, suppose that Martha says to Graham "Let's go on holiday to Greece this summer." This communication is about something (going to Greece this summer holiday): it has a particular focal content. In general, according to Luhmann, it is a property of the inner structure of all elements of meaning processing systems that they are *about something;* that they have a particular *focal content*.

At the same time Martha's communication "marginally" refers to many other possible contents of follow-up communications: e.g., "other possible destinations," "other possible countries or regions that may be visited," "other possible times of the year to go on holiday," and implied in these references, it refers to the "rest of the world"; its countries and regions, its seasons, the relation between work and holiday, etc. In general, the elements of meaning processing systems not only have a focal content; they also marginally refer to all other possible focal contents of follow-up elements, to "the rest of the world."

So, at a particular moment in time, an element of a meaning processing system has a particular inner structure: (1) it has a particular focal content, and (2) it marginally refers to all the possible contents of possible follow-up elements. According to Luhmann, this marginal reference is highly important for the autopoiesis of meaning processing systems. It "opens up" the space of all other possible contents of communication. In this way, it allows for the production of new elements with the same inner structure – focal content / marginal reference to all other possible focal contents of follow-up elements –, and so on.

This inner structure sets the elements of meaning processing systems apart from the elements of other autopoietic systems. It sets the elements of social and psychic systems apart from, for instance, the elements of multi-cellular organisms. We do not say that a cell has a focal content and marginally refers to all possible focal contents of follow-up cells.[2]

This "property," that, given a focal content, each element of a meaning processing system marginally refers to all possible contents of further elements, is called "meaning" by Luhmann. "The phenomenon of meaning appears as a surplus of references to other possibilities of experience and action. Something stands in the focal point, at the center of attention, and all else is indicated marginally as the horizon of an "and so forth' of experience and action" (Luhmann 1995, p. 60).

In Luhmann's technical vocabulary "meaning" should therefore not be confused with its everyday connotation as the meaning of a word or a gesture. It rather refers to the characteristic "inner structure" of the elements of meaning processing systems. This "inner structure" is illustrated in Fig. 4.4.

Figure 4.4 depicts one element of a meaning processing system at one moment in time (time = t). In the figure, we can see this element at the left hand side. The small black inner circle stands for the "focal content" (e.g., for what the communication is about). The dark grey area around it stands for the marginal references to all other possible contents. This margin "opens up" the space containing all possible follow-up contents (and their margin); it opens up the space containing all *possible* follow-up elements. In the figure these possible elements are represented by the two circles on the right hand side connected by a dotted line. Out of this set of all possible follow-up elements, an actual new element will eventually be produced (at time = t+1). In this way, each element of a meaning processing system carries within itself the seeds of the next element, thus enabling the autopoiesis of the system.

[2] By introducing the class of "meaning processing systems" (comprising the classes of social and psychic systems) Luhmann links up with the phenomenological tradition that considers conscious states as both "about something" (the intentional object, as it is called) and marginally referring to both the rest of the world and the past and the future. Luhmann seems to 'borrow' these ideas and "applies" them to characterize "conscious states" and "communications," the elements of respectively psychic and social systems.

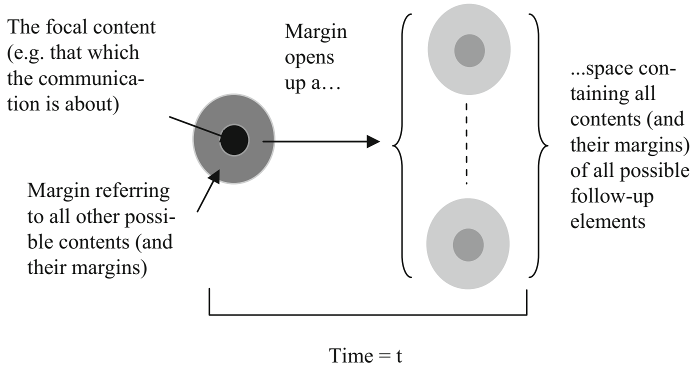


Fig. 4.4 "Meaning": the form of the elements of meaning processing systems

The "event-like" Character of the Elements of Meaning Processing Systems

A second property of meaning processing systems is that their elements are "event-like." The moment they appear, they (start to) disappear. For instance, the moment Martha says, "Let's go on holiday to Greece this summer", this communication appears ... and disappears. So, unlike organisms, consisting of related cells that have a longer time span, meaning processing systems consist of elements that are "event-like."

An important consequence is that in meaning processing systems, new elements do not connect to elements existing *at the same time*, but rather to *prior* elements. For instance, Martha's suggestion "Let's go on holiday to Greece this summer" may connect to Graham's prior question, "Where shall we go on holiday this year?" However, at the moment that Martha's connecting suggestion is produced, Graham's question (to which Martha's suggestion connects) has already disappeared. This "event-like" character implies that "connections between elements" are not connections between old and new elements existing at the *same* moment in time (as for instance is the case with the cells of a multi-cellular organism), but between elements existing at *different* moments in time (see Fig. 4.5).

In the figure, a multi-cellular organism is depicted at the left hand side. At a particular moment in time, it consists of old and new cells. The right hand side depicts the meaning processing system. In case A, the actual element at time = t relates to a prior element that existed at time = t−1. Some meaning processing systems can, at a particular moment in time, consist of multiple elements. This is case B.

4.2.2.2 The Structure and Self-Production of Meaning Processing Systems

In the example provided above, Martha suggested to go to Greece this summer. As argued, this communication has this particular suggestion as its focal content and marginally refers to possible contents of follow-up communications. Now, suppose

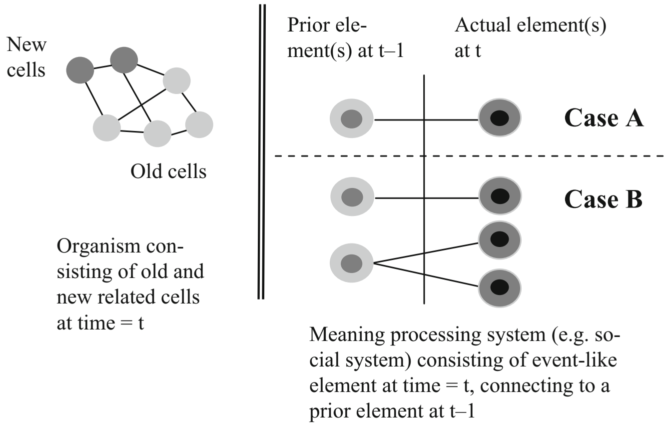


Fig. 4.5 Connection between elements in meaning processing systems

there is nothing to constrain the set of all possible follow-up communications. In this case, "everything" is possible as follow-up. Graham might, for instance, "react" by producing a low rumbling sound or by saying "Yes, systems theory is a great subject", to which Martha might "reply," "The pope is elected by the assembly of cardinals", etc. Soon, the "conversation" would end because it would have become fully incomprehensible and pointless.

If the production of elements by elements is to continue, the set of possible operations that can be produced by the meaning processing system somehow must be constrained to a subset of elements that are likely to contribute to the autopoiesis of the system and therefore "should" be generated, recognized, and connected to prior elements.

In the previous section, we explained that this is the role of the system's *structure*. It is the task of this structure to confine the set of all possible contents of follow-up elements to contents that are likely to contribute to the autopoiesis of the meaning processing system.

How this works for social systems will be described in the next section. Here, we confine ourselves to arguing that because the elements of meaning processing systems marginally refer to focal contents of all possible follow-up elements, there must be something that confines this set to a subset that is likely to contribute to the autopoiesis of the system. This is illustrated by Fig. 4.6.

In the figure, we follow the production of a new element of a meaning processing system in time. As before, the inner structure of elements is represented by a dark inner circle and a lighter grey outer circle. A black center and a dark grey surrounding area stand for an "actual" element. A dark grey center and a light grey surrounding area stand for "prior," "possible," or "potential" elements.

At time = t we have an actual element connecting to a prior element (that existed as an actual element at time = t−1). The actual element has a focal content and marginal references "opening up" at time = t the indefinitely large set of all

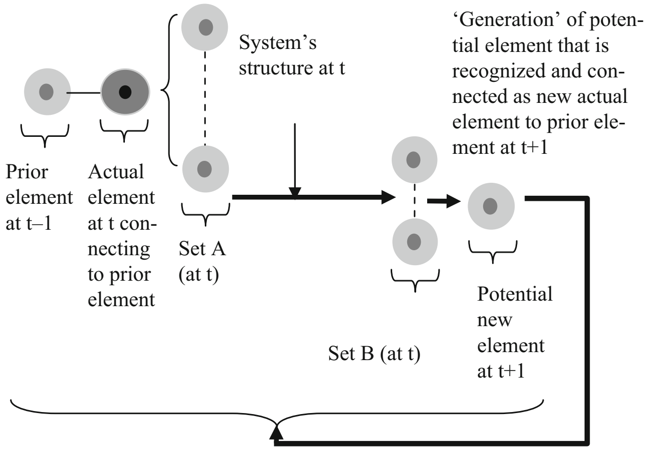


Fig. 4.6 The autopoiesis of elements of meaning processing systems

possible follow-up operations (Set A in the figure). The system’s structure at time = t constrains this set to a subset of elements that is likely to contribute to the autopoiesis of the system and therefore should be produced (Set B in the figure). As a rule a potential element is generated out of set “B” at time = t+1. At the moment of its generation this potential element is recognized and connected as the actual new element at time = t+1. It “connects” to the now gone “prior” element that existed at time = t. In the figure, this whole process of self-production is described as a loop. Each next state of the system can be determined by replacing t by t+1. Note that the “organization” of this process of self-production is quite similar to the one described in the previous subsection (see Fig. 4.3). The “only” difference is the “event-like” character of the elements and the nature of the connection between “prior” and “actual” elements.

As we said in the beginning of this section, Luhmann classifies psychic and social systems as meaning processing systems. Their elements – respectively, conscious states and communications – have the same inner structure: a focal content and a marginal surplus of references to the rest of the world, allowing for new elements with the same form. In spite of this formal similarity, the elements of psychic and social systems still differ, “conscious states” are not “communications” and vice versa, neither can the one be reduced to the other. Although their elements have “meaning” and their event-like character in common, psychic and social systems are autopoietic systems in their own right, with their own elements, structure, and processes of self-production. In the next section, we take the next step down the “tree” presented in Fig. 4.1 to discuss the elements, structure, and self-productive processes of social systems.

4.2.3 Social Systems

In this section, we have - at last – reached the class of systems organizations are a subclass of: the class of *social systems*. To explain what Luhmann understands by a social system, we follow the same "procedure" as in the two previous subsections. First we deal with their *elements*: *communications*. Second, we explain what *structures* the production of communications by communications in social systems: expectations about expectations. Third, we model the *self-productive process*, the process by which communications actually may be said to *produce* new communications.

4.2.3.1 Elements of Social Systems: Communications, their Inner Structure

In the section on Luhmann's generalized concept of autopoiesis (section 4.2.1), we explained that given particular "material conditions," elements of an autopoietic system generate "complexes of things and processes" that have the "right" inner structure to be recognized and connected as elements to elements of the system. If we apply this to communications as elements of social systems, the question arises what the inner structure of potential communications is, so that they can be generated, recognized and connected as actual communications to prior communications.

According to Luhmann, communications are a synthesis of three "components" or "selections" as he calls them: "information," "utterance," and "understanding." Together, these three selections form the inner structure of all communications.

The first selection is "information." By information Luhmann refers to what the communication is about, to its focal content. This focal content results from a selection out of the set of possible contents. Someone (Ego) selects a focal content out of possible contents. This focal content is the communication's information. For instance, from possible contents Martha selects the suggestion to go on holiday to Greece this year as the focal content. This is the "information" of the communication.

The second selection is called "utterance" (German: Mitteilung). According to Luhmann, communication involves not only what is communicated (the information), but also the selection how this "what" is communicated. This selection he calls "utterance." So, besides selecting a particular focal content – the information – Ego also has to select how to utter it. For instance, Martha selects to speak to Graham about their holiday destination. She selects to utter the information by means of spoken face-to-face communication. Most theories of communication stop after mentioning information and utterance as components of communication (Seidl 2005, p. 29). However, for Luhmann, one last "selection" is crucial.

"Understanding" is the third selection, completing the inner structure of a communication as element of a social system. Information and utterance are selections made by Ego. Understanding is something someone else (Alter) does. "Understanding," in Luhmann's technical vocabulary means that Alter understands

Ego's operation as "uttering" this particular "information" in this particular way. There is a difference between this technical meaning of the term "understanding" and its common meaning. Normally, we use the term understanding to refer to understanding what is said, to understanding the communicated information. In Luhmann's technical vocabulary "understanding" refers to Alter "understanding" that Ego utters information. To distinguish Luhmann's technical use of the word understanding from the common use (e.g., I didn't understand you correctly), we will refer to Luhmann's technical meaning by writing "understanding" between parentheses.

Why is "understanding" important according to Luhmann? Suppose Ego operates in a particular way, for instance, Ego slams the door. *Alter* may observe this operation and "see" it as:

1. Ego accidentally slamming the door
2. Ego slamming the door because she feels like it
3. Ego slamming the door because she wants to *convey in this way* (utterance) her *anger about Alter's behavior* (information).

In the first case, Alter does not even "see" Ego's operation as a selection by Ego. Alter sees the door slamming as a chance occurrence. Alter does not "understand" Ego's operation as uttering selected information. There is no communication.

In the second case, Ego's operation is "seen" as a selection by Ego. Out of all possible operations, Ego selected "slamming the door." However, the operation is not "seen" as Ego selecting a particular means to utter particular information. Once again, there is no communication.

Only in the third case, Alter "understands" Ego's operation as a particular utterance of a particular information. According to Luhmann, only if this happens, only if Alter "understands" Ego's operation as Ego uttering this particular information in this particular way, the inner structure of a communication is complete. In Luhmann's view, all communications are this unity of information, utterance, and "understanding." This is their inner structure.[3]

As "understanding" is concerned, three things may go wrong. Suppose that Ego operates in a particular way. Alter may "understand" this operation as Ego "uttering" particular information. Now, there is a communication: the three components are in place. However, (1) Alter may misunderstand the information that is uttered. Alter errs regarding what Ego intended to say. Misunderstanding the information does not mean there is no communication. On the contrary, to misunderstand the information uttered by means of a communication, there has to be a communication, i.e., it has to be "understood" that information was uttered. Misunderstanding the information uttered may surface in and be corrected by follow-up communication

[3] Against other theoreticians on communication (e.g. Habermas), Luhmann holds that this "threefold unity" of "information," "utterance," and "understanding" is sufficient as the "inner" structure of communication. Alter, for instance, does not need to "accept" or "reject" the "information" that is "uttered." "Acceptance" or "rejection" can ensue in a follow-up communication (which has the structure "information," "utterance," "understanding").

("Oh that is what you meant to say"). (2) Alter may "misunderstand" a mere operation by Ego as the utterance of information. Once again, there is a communication; the three selections are in place. This "misunderstanding" may again surface in and be corrected by follow-up communication ("Oh, you didn't mean to say something at all"). (3) Alter may fail to observe that Ego's behavior should be "understood" as the utterance of particular information. In this case, there is no communication. Alter fails to "understand" that Ego utters information. Now Ego may try again (Hey! I'm talking to you!").

In these three cases, either the understanding of the information or the "understanding" that information is uttered is checked by means of follow-up communication. Every follow-up communication, whatever else it does, implicitly checks whether either the information is understood or the operation is "understood" as the utterance of information. Often, this "test" will yield a positive result. Sometimes, it fails. For instance, it appears that the information was misunderstood. If this happens, communication about communication may ensue to remove the misunderstanding (e.g., "Huh, I thought you meant to say that...."). So, at the level of communications, both "understanding" (in the technical sense) and understanding (in its common sense use) are checked and corrected by follow-up communication, connecting to a previous communication.

It is easy to see that each single communication always involves Ego and Alter. Information, utterance, and "understanding" result from three selections. Two of them are made by Ego: the information and utterance. Alter makes the last selection: "understanding" Ego's operation as Ego uttering this particular information in this particular way.

This means that communication cannot be reduced to either Ego or Alter. As Seidl (2005) points out, "[. . .] a communication – as this unity of the three selections – cannot be attributed to any one individual (psychic system). Instead, communication constitutes an emergent property of the interaction between many (at least two) psychic systems" (p. 29). So, it is neither Ego (in isolation) nor Alter (in isolation) who "communicates." They are rather both involved in the production of a new type of operation: "communication." According to Luhmann, this new type of operations constitutes a new systemic level - the level of social systems. Communications are the elements of social systems.

4.2.3.2 The Structure of Social System: Expectations about Expectations

In Luhmann's theory, communications are the elements of a "new" class of autopoietic systems: the class of social systems. As elements of (autopoietic) social systems, communications produce communications. This means that they generate operations that are recognized as communications and connected to prior communications.

This, however, presupposes that operations are generated that *can* be recognized and *can* be connected as communications to prior communications. If the autopoiesis of elements by elements is to continue, the set of all possible operations should be constrained to a subset of communications that are likely to contribute to the autopoiesis of the social system and therefore should be generated, recognized and connected as follow-up communications to prior ones. And as we know by now, this is the function of the system's structure. So, before discussing the process of the self-production of social systems, we discuss their structure. To introduce this theme of social structure a simple example may be helpful.

Suppose that you have *never* learned to communicate, let alone what it means to buy something in a shop. You do not know what "clients" or "shop attendants" are. You do not know what behavior "shop attendants" expect from "clients" or what behavior "clients" expect from "shop attendants." Now suppose, you enter a space that actually is a "shop." Because you have not learned what type of behavior is expected by "shop attendants" from "clients," it is impossible to generate an operation that can connect to prior operations of the "shop attendant" or to which the "shop attendant" can connect follow-up operations. Because you did not develop expectations about what is expected of you by "shop attendants," you are unable to select information and utter it in a way that can be "understood" by the shop attendant as the utterance of information. As a result, there is no communication, let alone that communications "connecting" to communications can be generated. Now, suppose you just randomly *do* something in this situation, e.g., you make a low rumbling sound. Probably, the "shop attendant" will not "understand" this operation as the utterance of information, and no communication takes place. Perhaps, the "shop attendant" asks if you are feeling all right, but as you do not "understand" this operation as the utterance of information, you will be unable to "respond" in a comprehensible way. Soon the system of operations connecting to operations will break down. No self-producing social system consisting of communications producing communications will emerge.

Now suppose that you *have* learned to communicate and know what it means to "buy something in a shop." In this case, you expect the "shop attendant" to expect "client behavior" from you. In Luhmann's words, Ego developed expectations about the expectations that Alter has of Ego's behavior. *Guided* by these expectations, you, as a "client," for instance, ask, "Could I have an apple pie please?" Luckily, the "shop attendant" developed expectations about the behavior "clients" expect from "shop attendants." She understands your behavior as that of a client uttering information. She checks her supplies and replies, "Sorry, the apple pies are out, would you perhaps like another pie?" In Luhmann's words, Alter developed expectations about the expectations Ego has of Alter's behavior, and operates accordingly.

From this example we can learn that if operations are to be generated that can be recognized and connected as communications to prior communications:

1. Ego must have expectations about the expectations Alter has concerning Ego's behavior (the "client" must have expectations about the expectations the "shop attendant" has concerning "client behavior")
2. Alter must have expectations about the expectations Ego has concerning Alter's behavior (the "shop attendant" must have expectations about the expectations "clients" have concerning the behavior of "shop attendants").

One could ask, "Why are 'expectations about expectations' required?" "Why are just simple 'expectations' not sufficient?" To answer this question, suppose that Ego enters the shop. Ego expects a particular type of behavior of the shop attendant. However, Ego has no expectations about the "shopping behavior" the shop attendant expects from Ego. Because Ego has no expectation about what behavior is expected, Ego still cannot generate an operation that can be "understood" by the shop attendant as the utterance of information. So, both Ego and Alter can have expectations of each other's behavior. But this is not sufficient, for expectations only indicate what the other might do – they do not regulate *my own* behavior. What is needed is that both Ego and Alter have expectations about what behavior is expected of *them*. In Luhmann's words, "Ego must be able to anticipate what Alter anticipates of him to make his own anticipations and behavior agree with Alter's anticipation" (Luhmann 1995, p. 303).

Guided by expectations about expectations both Ego and Alter can generate operations (selections of information and utterance) that can be recognized ("understood") as communications, that can connect to prior communications, and to which follow-up communications can connect. In fact, these expectations about expectations constrain the set of all possible operations to a subset. In this subset, we find communications that, in a given situation, are likely to contribute to the autopoiesis of communications by communications, and for this reason, should be generated, recognized, and connected to prior communications. Expectations about expectations structure the production of communications by communications.

In Sect. 4.2.1 we mentioned that the structure of autopoietic systems does not determine what elements will be produced. This is also true for social systems. "Social structures" only constrain the set of operations that can be produced to a subset containing communications that are likely to contribute to the social system's autopoiesis, and for this reason should be produced. In spite of these structures, Ego or Alter can still wittingly or unwittingly produce operations deviating from the expected. If this happens, the operation in question may be communicatively "marked" as "deviant." Communication may ensue to set things right. For instance, someone enters a flower shop, asking for a loaf of bread. Probably, this operation will be marked as deviating from "normal" shopping behavior, and communication will ensue directed at "saving" the situation.

In Sect. 4.2.1 we also argued that structures of autopoietic systems are produced by the system they help to produce. Again, this is also true for the structures of social systems. Of course, Ego and Alter, their memories, their imagination, etc., are needed for the development of social structures. However, they are needed as "material conditions and processes" in their environment.

Social structures are reinforced or changed by means of the social system they help to produce. Existing social structures may be *reinforced* by communications successfully "complying" with them. For example, if I succeed in buying flowers, the social structures supporting the communications involved in this transaction are probably reinforced. Reinforcement may also be achieved by communication about communication. For example, deviations from expectations about expectations are socially treated as "deviations" that should not occur. This, then, reinforces the original expectations. Social structures may also be *instituted* or *changed explicitly* by communications about mutual expectations. For example, parents teaching their children how to behave “properly” in shops. Social structures may also change *implicitly* over time. For example, customs “defining” “polite” communication may gradually change. In each of these senses, social structures can indeed be said to be a product of the social systems they help to produce.

4.2.3.3 The Autopoietic Production of Social Systems

Basically, the production of communications by communications is quite similar to the production of the elements of meaning processing systems as discussed above. This is not surprising, for social systems belong to the class of meaning processing systems (the other member is the class of psychic systems). This means that the *elements* of social systems have meaning as their form (focal content and marginal references to all other contents). Their *structure* allows for the processing of meaning by constraining the set of operations that can be produced by the system to a subset of communications likely to contribute to the autopoiesis of the social system. What is *specific* for social systems is that *communications* as the unity of information, utterance, and “understanding” are their elements and *expectations about expectations* their structures. If we take this as our point of departure, the production of communications by communications can be illustrated as in Fig. 4.7.

In the figure, we see an actual communication at time = t, connecting to a previous one. This actual communication has a focal content (a “theme” as Luhmann calls it) and marginally refers to all follow-up operations that can be produced

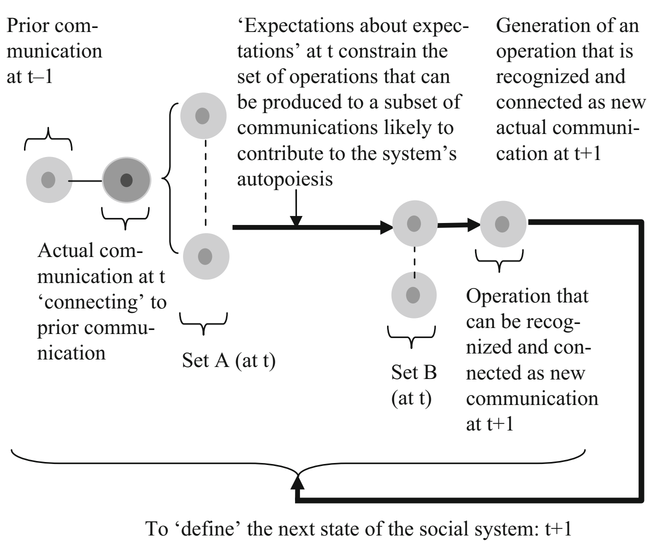


Fig. 4.7 The autopoiesis of communications, constituting social systems

by the system (Set A in the figure). Given the system's structure at time = t, the actual communication is "*connectable*" to a follow-up communication. However, if a new communication is to connect to the given communication, a lot needs to happen.

Given prior communication (e.g., the one at time = t−1), the system's structure at time = t constrains the set of operations that can be produced to a subset containing communications that are likely to contribute to the autopoiesis of the social system (set B in the figure). Set B is the set of communications out of which a communication should be generated in order to be recognizable and connectable as a new communication (at time = t+1) to the actual communication (at time = t).

To understand how such a connectable communication is generated we have to remember that the structure of the social system consists of expectations about expectations. These expectations allow Ego to select information and utterances that Ego expects Alter to expect from Ego. If these expectations about expectations are "correct," Alter is able to "understand" Ego's operation as utterance of particular information. In this way, the system's structure allows for the generation of an operation that is recognizable and connectable as a communication to the prior communication. If this operation is generated, recognized, and connected to the prior communication, it becomes the new communication at time = t+1. Because of its form (content and marginal references) and the system's structure at time = t+1, this new communication in principle is connectable to the next one, and so on.

Based on this description, one could ask how social systems "start" or how they "end." This question is relevant, for the loop presented in Fig. 4.7 always seems to presuppose "prior" communication and always seems to lead to the production of a follow-up communication "connecting" to the prior one. This is counterintuitive. For instance, I may meet someone for the first time "Hello, pleased to meet you". A string of communications may ensue that eventually ends, "Goodbye, nice to have met you". How can this intuitive understanding of communication and the "never ending" process illustrated by Fig. 4.7 be reconciled?

One way to do this is by means of a likeness. We know that we as living organisms were born and are bound to die. In this sense, – relative to us – there is an "absolute" beginning and end to life. However, we are a part of a larger process. In this process the autopoiesis of life continues. Living organisms, consisting of cells producing cells, produce living organisms, consisting of cells producing cells. Although concrete individual living organisms may be born and perish, the production of cells by cells, the autopoiesis of living organisms, continues. In the course of this process, living organisms may co-evolve with each other and their environment, allowing for the huge variety of species we know. So, although there may be concrete living organisms, that are born and die, the larger process of the autopoiesis of living organisms continues. Cells continue to connect to cells. However, what about this larger process? There is evidence that it had a beginning, and perhaps it once ends (for instance, by self-inflicted causes, by a large meteor hitting the earth, or by the sun exploding). According to Luhmann, this larger process "ends" when "the material conditions and processes" disappear that are needed for the production of potential elements (with the right inner structure). It begins when these "material conditions and processes" are in place.

With respect to social systems, something quite similar can be said. Just like an individual living being may be "born" and "die," a chance meeting in the street may "begin" and "end." This chance meeting, however, is a "short lived" social system contributing to and appearing and disappearing in the total process of the autopoiesis of the social. Just as in the case of the autopoiesis of living organisms, this total process "begins" and "ends" with the appearance or disappearance of the right "material conditions and processes" needed for the production of communications. This means that there have to be psychic systems; that these systems come to have dealings with one another; that they develop expectations about each other's behavior, and expectations about expectations. If all these conditions are "in place" social systems, consisting of communications producing communications can emerge. So, indeed, some social systems may be shorter lived. However, they are a part of and contribute to the larger process of the autopoiesis of the social that for its beginning and end depends on the occurrence of the "right material conditions and processes.

This account of "the larger process of the autopoiesis of the social" and shorter lived social systems being a part of and contributing to it implies that Luhmann's social theory should accommodate for both the larger process and the shorter lived social systems. If we take stock of the types of social systems discussed by Luhmann, we can see that it does. In his theory Luhmann distinguishes three types of social systems: "society" (including its differentiation into societal subsystems), "organizations," and "interactions" (see Fig 4.8).[4]

Luhmann defines "society" as the system encompassing all communications. "Society is the all-encompassing social system that includes everything that is

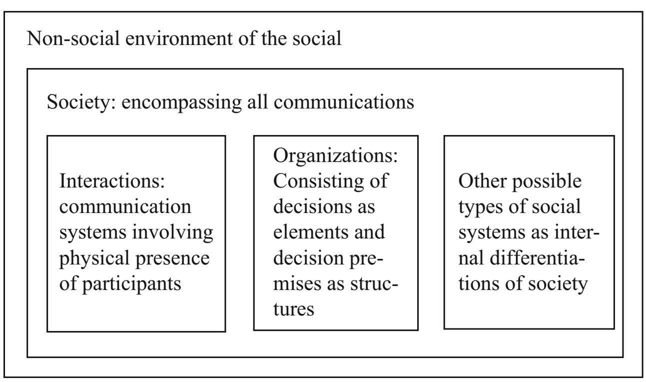


Fig. 4.8 Society, its non-social environment and internal differentiations

[4]This does not necessarily imply that, according to Luhmann, these system-types provide an exhaustive list of all possible social systems.

social and therefore does not admit a social environment. If something social emerges, if new kinds of communicative partners or themes appear, society grows along with them. They enrich society. They cannot be externalized or treated as environment, for everything that is communication is society. Society is the only social system in which this special state of affairs occurs” (Luhmann 1995, p. 408). Apparently, society is the larger social process in which shorter lived social systems can differentiate, emerge, develop, and perish.

Within (modern) society, i.e., as shorter lived internal differentiations of society, Luhmann distinguishes “interactions” and “organizations.” “Interactions” are social systems in which the participants are physically present. For instance, a chance meeting in the street, a meeting in an organization, or a cocktail party may be regarded as “interactions” (for an extensive discussion of this type of social systems see Kieserling (1999). As a rule, “interactions” are short-lived, drawing heavily on already developed expectations about expectations. The second type of social systems that differentiated within (modern) society is the “organization.” What, according to Luhmann, organizations are, is the subject of the next section.

4.3 The Social “arche”: Organizations as a Particular type of Social Systems

Luhmann conceives of organizations as a particular type of social systems within (modern) society. As social systems, they have communications as their elements. However, being a *particular type* of social systems, their elements are a particular type of communications. Luhmann calls these communications: *decisions* (Luhmann 2000, p. 63). Something similar can be said for the structure of organizations. As social systems, organizations have expectations about expectations as their structure. However, as a *particular type of social system*, they have particular expectations about expectations as their structure. Luhmann calls them: *decision premises*. Together, decisions and decision premises contribute to the self-production of organizations as a particular type of social system.

Before we continue to discuss decisions as elements of organizations and decision premises as their structure, we want to underline that Luhmann’s outlook on organizations is quite different from definitions we find in standard textbooks. There, organizations are, for instance, defined as “goal searching systems,” “goal realizing systems,” “goal directed combinations of capital, people, and information,” et cetera.

According to Luhmann, such definitions are only secondary to his own theory of organizations (Luhmann 2000, p. 63). In his view, “goals,” for instance, *result* from decisions and function as decision premises for follow-up decisions. The same holds for the “means” (e.g., capital, people, or infrastructure) to realize these goals. These “means” are a result from decisions, and can function as premises for the production of follow-up decisions. According to Luhmann, searching or realizing goals are social processes that in organizations take the form of either decisions or decision premises. In Luhmann's view, to argue that organizations consist of goals and/or means to realize them is to mistake *results* of organizing for organizing itself. As a sociologist, he is interested in organizing itself, the processes of self-production of organizations in modern society. According to him, a mistaken focus on results of organizing harbors the danger that practical problems and instrumental solutions become predominant in organization theory rather than a systemic and sociological analysis of organizing and organizations.

By treating organizations in terms of the autopoiesis of decisions, Luhmann introduces a "non-standard" view on organizations. He wants to clarify the inner logic of their self production, i.e., he wants to explain how organizations produce themselves by producing their own elements and structures. In his view, such a clarification should be distinguished from (1) theoretical attempts to find the "essence" of organizations, (2) normative theories of organizations, and (3) empirical descriptions of organizations.

Theories aiming to penetrate to the "essence" of organizations, according to Luhmann, have had their day. Instead of actually finding the answer to "what" organizations in the end "are" (what their "essence" is), attempts to formulate this type of theory only resulted in an irreducible variety of supposed "essences." According to Luhmann, this is not helpful. So, instead of asking "*what*" organizations "are," Luhmann proposes to answer the question "*how*" organizations as social systems "produce themselves."

Clarification, as Luhmann conceives it, should also be distinguished from normative theories of organizations. Clarification does *not* provide prescriptions of what organizations *should* do, for instance, to survive successfully. According to Luhmann, normative theories provide a deceptive image of organizations, for instance, depicting them as "rational," aiming at "effectiveness" and "efficiency." In his view, such theories do nothing to clarify the actual processes of their self-production. And clarifying these processes is what Luhmann is after. Therefore, we should not expect recipes for success from Luhmann's theory of organizations. He is happy to leave them to, for instance, business consultants and management gurus (people he never fails to mention with a touch of irony). So, unlike normative theories of organizations, "clarification" means: bringing to light *actual* (not "desired") processes by means of which organizations as social systems produce their own elements and structure.

Finally, clarification should be distinguished from descriptive theories of organizations. This type of theories empirically describes organizations in terms of variables such as "size," "age," "lay-out," "coordination mechanisms," "centralization," etc. Such theories describe *results* of ongoing processes of self-production, rather than the *actual* processes themselves. To this purpose, they make use of contingent sets of variables that need not be related to the inner logic of the self-production of organizations. As a contrast, Luhmann's project of clarification focuses on the processes of self-production of organizations, using a conceptual framework especially designed to this purpose.

To unfold what Luhmann's "non-standard" view entails and how it contributes to a clarification of the inner logic of organizations, we first focus on decisions as elements of organizations. Second, we discuss decision premises as their structure. Third, we unfold the process of self-production of organizations and we deal with some frequently asked questions concerning Luhmann's theory of organizations. Finally, in Sect. 4.4 we connect Luhmann's account of the social "arche" of organizations to the experimental "arche" discussed in Chaps. 2 and 3.

4.3.1 Elements of Organizations: Decisions

According to Luhmann, decisions as elements of organizations are *communications that communicate a selection as a selection*. What does he mean by this?

We know that communications as elements of social systems consist of three selections: the selection of "information," the selection of "utterance," and the selection of "understanding." For this reason, it can be argued that all communications communicate selections. For instance, when Martha suggests going on holiday to Greece this summer, she obviously selected this "information" as a theme for a conversation, and she obviously selected face-to-face communication as the particular way to utter this information. Given their particular inner structure – i.e., their being made up out of three selections – all communications, communicate selections. However, according to Luhmann, not all communications are decisions. Not all communications are communicative events in which a selection is communicated *as* a selection. So, what is the difference between "normal" communications and decisions as elements of organizations? What does it mean to communicate a selection *as* a selection?

We argued that, according to Luhmann, decisions have a particular inner structure: out of a set of selectable options, one is selected. To communicate a selection *as* a selection means exactly this: to communicate that out of a set of selectable options one is selected and the others are not. To illustrate what Luhmann means, consider the following example. Suppose you are a project leader and things are not running smoothly. You are behind schedule, the quality of the finished work is bad, and you are losing control. You reviewed different options with your manager: investing in extra personnel, introducing measures supporting quality control, terminating the project. Now, you get an email in which it is stated that "... we are not going to continue your project. Investing in additional personnel or a quality system is throwing good money after bad". The email is a communication consisting of "information," "utterance," and "understanding." However, it is not just a "normal" communication. It rather communicates a selection as a selection, for it communicates:

1. a selection (we are terminating the project)
2. that this is a selection from a set of selectable options
3. that the other options are not selected

If a communication has this "inner structure," i.e., if it communicates that a selection has been made from a set of selectable options, it *can* be recognized as a decision and connected to prior decisions and thereby contribute to the self-production of the organization (Luhmann 2000, p. 67).

Once the decision is actually recognized as a decision and connected to prior decisions, it may be accepted or not. If it is accepted, it structures follow-up decisions that are in line with it (for instance, a decision to re-allocate project staff, to write down certain investments, to start up a substitute project, or to sack the project leader). If it is not accepted, it may cause all kinds of turmoil in the organization, e.g., it may trigger a debate about continuing or discontinuing the project, about the capacities of "higher management levels" etc. In this sense, non-acceptance may also lead to follow-up decisions, for instance, to deal with non-acceptance.

With this possibility of either "acceptance" or "non-acceptance" of what has been decided by a decision, we touch a tricky point. We explained that, according to Luhmann, decisions as elements of organizations have a particular inner structure: they communicate a selection *as* a selection. This means that they emphasize that one of the options was actually selected and the other options were not. This emphasis on selection may trigger all kinds of questions concerning the "rationality" and "legitimacy" of the decision. "Why is the project discontinued?", "Why don't we change the project leader?" or "Why don't we lower our absurdly high quality standards?"

So, because of their inner structure, because of their emphasis on selection, decisions may trigger questions undermining the "acceptance" of the selection that was communicated. Was it the "right" decision, taken at the "right" moment, and "uttered" in the right way? Every decision, because it communicates a selection *as* a selection, may trigger such questions. Viewed in this way, decisions have an inherent tendency to "undermine" or "deconstruct" themselves. Because decisions communicate selections *as* selections, they "trigger" questions about the selected option, possibly undermining the acceptance of that selection as a premise for further decisions. Each decision communicates the contingency, uncertainty, and risk involved in the decision.

According to Luhmann, contingency, uncertainty, and risk are not "contingent" properties of decisions. They are fundamental. They cannot be "organized away." The decisions needed to do this, because of their inner structure, would immediately "reintroduce" contingency, uncertainty, and risk.

To explain why this is so, Luhmann refers to one of von Foerster's more "paradoxical" statements. Von Foerster argues that "problems that can be decided do not need to be decided and problems that are impossible to decide must be decided" (von Foerster 1993, p. 72 ff.).

To understand von Foerster's statement we need to separate two classes of problems. The first class consists of those problems that can be solved fully by means of calculation. For instance, to establish whether 768931 is a prime number, one only needs to calculate whether 768931 is divisible by another number than one and itself. In the case of this type of problems, we can decide, i.e., select an option (e.g., "768931 is not a prime number") from a range of options ("768931 is a prime number" and "768931 is not a prime number") by means of calculation. The problem can be decided, but a decision is not needed. Mere calculation suffices.

According to von Foerster, there is also a second class of problems. These problems cannot be fully solved by means of calculation. An example of this type of decisions is a decision about going to the Zoo or going to a faculty meeting. I can decide to go to the faculty meeting because of reasons of career opportunities. Yet, I can also decide to visit the Zoo because of reasons of spiritual well being. To decide, I now have to weigh career opportunities and spiritual well being. Before I know it, I am sucked into a quasi-philosophical "debate" with myself about the relative merits of my career and spiritual well being, or about their mutual conditioning or destructive relation. This "debate" may well take the entire afternoon without resulting in a stringently calculated selection. In this particular case, I cannot make a decision by means of calculation, and it is for this reason that I must decide.

Given these two classes of "decision situations," we can understand von Foerster's paradoxical statement: there are problems that can be decided (by means of rational calculation) and for this reason do not need a decision and there are problems that cannot be decided (by means of rational calculation) and therefore must be decided.

According to Luhmann, decisions as elements of organizations belong to the second class. It is impossible to fully ground the selection of an option by means of calculation or deduction leading with necessity to that option. If this were possible, no decision would be required, mere calculation would do the job. In his view, decisions as elements of organizations cannot be grounded in a deductive argument, showing with necessity that the selected option was the "right" one. In Luhmann's view, all decisions involve a "rationality deficit" and for this reason require a risky "jump" from the set of selectable options over an abyss of possible further reasoning to the selected option. So, if one takes "to be decidable" as equivalent to "to be calculable," decisions as elements of organizations are "undecidable." They cannot be decided with necessity by calculation alone. They always require an uncertain and risky "jump." Hence, contingency, uncertainty, and risk are inherent features of decisions as elements of organizations.

Because decisions communicate selections *as* selections, they amplify the uncertainty and risk inherent in them, possibly provoking questions about their "rationality." Because *all* decisions can be contested, the "acceptance" of decisions as a point of departure for follow-up decisions in organizations seems quite problematic. "Why should I accept a selected option and use it as a basis for follow-up decisions, if the selected option can't be fully grounded with necessity by rational argument?"

According to Luhmann, all this may endanger the autopoiesis of organizations. If all decisions can be contested, and if in principle no decision needs to be accepted as a premise for the production of follow-up decisions, then the production of decisions by decisions may become a problem. Because of their being "communications of selections as selections," i.e., because of their inner structure, decisions seem to undermine the self-production of the very system they are the elements of!

If organizations are to be possible, the contingency, uncertainty, and risk involved in decisions somehow need to be "absorbed," making their "acceptance" as a point of departure for follow-up decisions *probable* instead of *improbable*.

According to Luhmann, this is a function of the *structure* of organizations. Their structure functions as a background of (relatively) "undisputed" *decision premises* contributing to the "generation" of operations that can be recognized and connected as decisions to prior decisions. Moreover, they increase the probability of the "acceptance" of decisions as a point of departure for the "generation" of follow-up decisions. For instance, suppose that the "production of luxury cars" has been selected as an organizational goal. In the course of time, this goal has become deeply entrenched in the organization. Although the decision to manufacture luxury cars was and is a contingent selection, with all the uncertainty and risk involved, it has become a relatively "undisputed" background for other decisions, e.g., about "how to approach customers" or "requirements to production control standards." Although the goal remains contestable, the chance that it is actually contested in new decisions (e.g., to hire a new CEO) is quite low.

Therefore, given a "background" of relatively "undisputed" decision premises, new decisions may be generated that have an increased probability of being recognized and connected as decisions and being accepted as a point of departure for follow-up decisions. In this way, the production of decisions by decisions is "safeguarded" by the organization's structure. In the next subsection, we further examine this subject of decision premises as the structure of organizations. Here, we want to conclude by summing up our findings regarding "decisions" as elements of organizations.

According to Luhmann, organizations are a particular type of social systems consisting of decisions as their elements. Decisions are not "mental" but "social events": they are communications. However, they are not just any communication. Decisions communicate a selection *as* a selection. They communicate that out of a set of contingent and selectable options one contingent option has been selected.

If decisions are "real" decisions, the selection of the option cannot be "resolved" with necessity by means of rational calculation alone (for in this case no decision would be required). For this reason, all decisions involve an uncertain and risky "jump" from the set of contingent and selectable options to the contingent selected option. In this sense, all decisions suffer from a "rationality deficit." This deficit makes their acceptance as a basis for follow-up decisions at least problematic and at worst improbable. As such, it endangers the production of decisions by decisions. The organization's structure increases the probability of the acceptance of decisions, thus contributing to the autopoiesis of organizations.

4.3.2 The Organization's Structure: Decision Premises

As all autopoietic systems, organizations need a structure to support the production of elements by elements. According to Luhmann, the structure of an organization is an arrangement of “decision premises.” Decision premises do what all structures of autopoietic systems do: they constrain the set of operations that can be produced to a subset of elements that are likely to contribute to the autopoiesis of the system and therefore should be produced. Below, we begin by examining the main function of decision premises in organizations. Then, we unfold and elaborate the different types of decision premises distinguished by Luhmann.

4.3.2.1 The Main Function of Decision Premises

In his organization theory, Luhmann distinguishes different types of decision premises. To explain their function, we select one of them: organizational goals, and more in particular the goal of producing and selling solar energy.

Based on Ashby, we know what goals do from a cybernetic perspective: they specify what should be realized by transformations and function as a point of departure for the design, operational regulation, and performance of transformations. Here we ask what their function is from the perspective of social system theory: how do organizational goals – and decision premises in general – contribute to the regulation of the autopoiesis of organizations?

In a nutshell, this contribution can be characterized as follows: *new decisions are produced taking into account existing decision premises*. Applying this to the example of the goal to produce and sell solar energy, this means that new decisions are produced taking into account this goal. According to Luhmann, this “taking into account” involves three aspects of decision premises.

First, it means that decision premises function as normative points of reference, providing a focus for the production of follow-up decisions. For example, once it is decided that producing and selling solar energy is the goal, other decisions may be produced taking this goal into account. As a normative point of reference for the production of other decisions, the goal opens up a logical space of possible decisions that take this goal into account.[5] For example, in line with the goal of producing and selling solar energy, it may be decided to explore more efficient means of storing this type of energy. New decisions are produced taking into account existing decision premises means that these premises provide a normative point of reference opening up the logical space for the generation of operations that can be recognized and connected as decisions to prior decisions.

Second, as normative points of reference, decision premises contribute to the regulation of the production of decisions by decisions by “marking” decisions deviating from them. For example, suppose, that someone decides to plan an advertising campaign for *eco-energy*. Given the goal of producing and selling

[5] That a goal is a normative point of reference does not say anything about its instrumental rationality, goal rationality, or its moral value. As we all know, the most instrumentally stupid or ethically abject goals may be adopted in organizations. It just states that it specifies a factual organization specific standard or norm opening up the logical space for the production of new decisions in the organization.

solar energy, this plan may be marked as too broad, and its maker may be instructed to redesign the campaign. Such a "marking" of the advertising campaign as deviating from the goal, is an example of how decision premises "regulate" the "generation" of decisions in organizations. Because the goal functions as a normative point of reference for producing follow-up decisions, decisions that are at odds with it may be marked. Organizational members are aware of this; they know that deviating decisions may be marked *as* deviations. They know that these deviations may be attributed to them and that they may be held accountable and punished accordingly. As a result, they will think again before generating behavior they expect to be treated as "deviating." For this reason, decision premises, as normative points of reference, have a normalizing influence on the generation of new decisions. New decisions are produced taking into account existing decision premises because these premises allow for "marking" deviations from them *as* deviations that may be attributed to organizational members who may be held accountable for them.

Third, new decisions take into account existing decision premises as presuppositions. According to Luhmann, this means that, relative to new decisions, they – most of the time – are treated as "given." As presuppositions they release the organization of the burden of testing their truth each time a new decision is produced. What counts is, that they are "relevant" in a given situation (Luhmann 2000, p. 222). For instance, the decision to hire a new marketing manager may be made taking into account the goal of producing and selling solar energy. However, this decision does not require a re-assessment of this goal. Relative to the decision to hire a new marketing manager, producing and selling solar energy is treated as a presupposition.[6] This point that relative to a decision, existing decision premises function as presuppositions increases the probability of the "acceptance" of decisions. In the subsection on the inner structure of decisions we argued that decisions because of their "form" trigger questions that, in the end, cannot be answered by rational argumentation alone. As argued, this might endanger the production of follow-up decisions. Deciding in reference to "presupposed" decision premises increases the probability of acceptance, putting a stop to questions that, in the end, endanger the production of decisions by decisions.

In general, then, we can say that decision premises contribute to the production of decisions by decisions by providing normative points of reference. As normative points of reference, decision premises open up a logical space of possible follow-up decisions, allow for marking decisions deviating from them, thus normalizing the production of new decisions, and provide a given point of reference that does not need to be questioned every time a new decision is produced, thus reducing the complexity of the production of decisions by decisions. In Luhmann's own words: decision premises, "[...] focus communication on the distinctions laid-down in the

[6]Of course, this does not imply that the goal can no longer be discussed or changed. In some cases, it will be decided to re-assess the goal. However, for many decisions, the goal functions as a 'given'; pre-structuring the logical space allowing for their production.

premises. This makes it probable that future decisions taken under reference to these given premises will be observed from the point of view of their taking into account or not taking into account and their conformity with or deviation from these distinctions, instead of “rolling out” each time the full complexity of situations” (Luhmann 2000, p. 224; our translation).[7]

4.3.2.2 Types of Decision Premises

Above we discussed an organizational goal as an example to explain the function of decision premises in organizations. In his theory, Luhmann distinguishes nine different types of decision premises. To continue our explanation of how decision premises “work” in organizations, it is useful to discuss these different types of premises.

Membership

The first type of premise is *membership*. Membership is an important decision premise in organizations. To understand why, take the following example. Suppose you work for an organization. One evening you go home and you meet a stranger on the subway. This stranger “asks” you to erase all confidential files from the hard disk of your office computer. Probably, you do not interpret this “request” as an organizational decision that should connect to follow-up decisions. As a rule, “strangers one meets on the subway” are not in a *position* to issue “orders,” “counting” as organizational decisions. Now, suppose the next day, you sit at your desk and the new manager enters the room. To your surprise, this new manager is the same person as the stranger you met on the subway. Again, she orders you to erase the requested files. This time, you probably treat the “operation” as a decision that can either be accepted or rejected, and as such, should be connected to follow-up decisions.

From this example, we can learn that if operations are to be generated, recognized and connected *as* decisions, the question of membership is relevant. It is, as it were, a “test” that has to be passed if operations are to “count” or “treated” as decisions in organizations. In organizations, only members can contribute to the generation of decisions and – this is the other side of the coin – decisions can be attributed to organizational members who can be held accountable for them. The moment you become an organizational member, you have to take into account that your behavior can be “treated” as a decision that can be “attributed” to you and for which you can be held “accountable.” So, according to Luhmann, membership is

[7] “… fokussieren die Kommunikation auf die in den Prämissen festgelegten Unterscheidungen, und das macht es wahrscheinlich, dass man künftige Entscheidungen mit Bezug auf die vorgegebenen Prämissen unter dem Gesichtspunkt der Beachtung oder Nichtbeachtung und der Konformität oder Abweichung beobachten wird, statt die volle Komplexität der Situationen jeweils neu aufzurollen” (Luhmann 2000, p. 224).

fundamental to the autopoiesis of organizations. Without membership there is no way to distinguish mere operations from decisions, and without decisions there are no organizations (as a particular type of social systems). To quote Luhmann, "That membership relations can be established and ended by means of decisions is constitutive for the development of organizations as such" (Luhmann 1993, p. 364; our translation).[8] Of course, as all decisions in organizations, decisions about membership require decision premises, regulating the production of these decisions. In this sense, membership and membership rules are important to establish the "boundaries" between organizations as a particular type of social system and their social and non-social environment.

However, membership is relevant for the regulation of the self-production of organizations for yet another reason. By becoming a member of an organization, we, in principle, agree to follow instructions and to act in accordance with relevant decision premises. So, if my boss instructs me to plan a course on social systems theory, I know (expect) that my boss expects me to follow this instruction. For by becoming a member of the university, I, in principle, agreed to follow the instructions of my superiors. Of course, I may question the instruction: I may use my right to *voice* disagreement. However, if my boss insists, I will have to choose between staying a member of the organization and carry out the instruction or asking for a different job (remaining a *loyal* member of the organization), or resigning, or being fired (*exiting* the organization). Moreover, I can sabotage the instruction, but I know that if this comes out, it may have serious consequences. For, as an organizational member, I know (expect) that I am expected to loyally carry out the instruction. Finally, due to circumstances, I may be unable to realize the goal set by my boss. If this happens, I expect that tough questions will be asked. And I know (expect) that satisfactory answers are expected of me.

So, membership, in principle, "binds" organizational members to relevant decisions and decision premises: they are expected to choose between loyalty, voice, and exit. For this reason, membership is a powerful mechanism to deal with the problem of acceptance.

As we mentioned earlier, all decisions involve a "jump" that cannot be calculated or deduced. All decisions, in principle, can be contested. One can always ask "Why?" Because of this, the acceptance of decisions (e.g., instructions) in the sense that decisions allow for the production of new decisions that are in line with it, becomes a problem. Membership, binding organizational members to loyalty, voice, or exit, reduces this problem. By becoming an organizational member, one, in principle, agrees to loyally execute instructions. This becomes the standard and not the exception. In case of disagreement, organizational members may, or are even expected to, voice their critique. But in the end, one has to choose between loyalty and exit. Note that this does not mean that decisions and decision premises never change in organizations. On the contrary, in organizations there are plenty of

[8] "Dass Mitgliedschaftsverhältnisse durch Entscheidung begründet und aufgelöst werden können, ist konstitutiv für Organisationsbildung schlechthin" (Luhmann 1993, p. 364).

members “bound” to critically assess old and invent new goals, plans, instructions, producing a continuous stream of new decisions and decision premises.

Membership, in sum, is an important decision premise. It contributes to the regulation of the production of decisions by decisions by (1) allowing for the distinction between operations that can count as decisions and operations that cannot and (2) “binding” organizational members to loyalty, voice, or exit thereby increasing the probability of the acceptance of decisions needed for the production of follow-up decisions.

Communication Pathways

Communication pathways are decision premises prescribing the “route” that should be followed if operations are to “count” as decisions in an organization. “By means of decision premises it is also possible to prescribe communication pathways that must be observed if the decision is to be acknowledged as an organizational decision” (Luhmann, 2000, p. 225; our translation).[9] If an operation is to “count” as an organizational decision, it is often not sufficient that it can be attributed to an organizational member. Most decisions – if they are to count as such – require the involvement of organizational members holding particular “positions” (see below). “Rules” prescribing which “positions” need to be involved in what order if an operation is to count as an organizational decision are called “communication pathways.” For instance, the decision to hire a new team member in a business unit may require authorization by first the director of the business unit and then the team’s leader. Without this authorization operations do not count as a decision. So, if one of the team members says to a candidate, “You are hired”, this does not count as a decision to hire a new team member. Still, the behavior of this team member can be treated as a decision; that is to say as a decision to deviate from organizational guidelines concerning the hiring of new personnel. As such, the utterance “You are hired” can be attributed as a decision to that team member, who can be held accountable for it, and sanctioned accordingly. So, just as membership, communication pathways are important decision premises supporting the constitution of the decisions constituting the organization. According to Luhmann, communication pathways do not necessarily involve “hierarchy.” There are also lateral communication pathways. Moreover, one and the same pathway may be used for different (types of) decisions. So, not every new (type of) decision needs the definition of a new decision pathway. Decision pathways may even be used for the purpose of the production of decisions in cases where it is unclear what the actual decisions are going to be. For instance, an organization may decide to define a “crisis organization” specifying the communication pathways for decisions that

[9] “Über Entscheidungsprämissen können auch Kommunikationswege vorgeschrieben werden, die eingehalten werden müssen, wenn die Entscheidung als eine solche der Organisation Anerkennung finden soll” (Luhmann 2000, p. 225).

may have to be made in times of crisis, without exactly knowing in advance, what the crisis and the type of decisions are going to be.

Decision Programs

Decision programs are regulative conditions for correct or incorrect decision behavior. Luhmann distinguishes two types of decision programs: *goal programs* and *conditional programs*.

Goal programs specify goals (desired effects) that should be pursued. These goals, then, function as decision premises for decisions about means (causes) to realize them. As a goal or desired effect can be realized by a variety of means/causes, goal programs allow for variety. Dependent on the circumstances or expected side-effects, different means, realizing the same goal, may be selected under different circumstances. For instance, under condition of scarcity of particular raw materials, it may be possible to substitute them for others, and still realize the goal. Or, dependent on the probability of the actual occurrence of a negative side effect of fossil fuels, it may be possible to select other energy sources to both realize the goal and avoid the negative side effects. According to Luhmann, goal programs can constitute hierarchies, allowing for a recursive unfolding of goals over different organizational levels: goals of the organization as a whole, of its business units, of departments within units, and teams within departments. In such a recursive unfolding, the selection of means may then be left to each level of recursion.

Conditional programs have as a general form: "if a particular condition is fulfilled, then a particular action or consequence should follow." This form allows for different levels of specificity of the condition and/or the consequence. In the case of a precise specification, it is possible to add all kinds of rules to deal with exceptions. In the case of a less precise specification, the repeated use of conditional programs probably yields additional cultural "norms" making the description of either the condition or the consequence more definite. It is possible to connect conditional programs into chains. In these chains, the actual occurrence of a consequence of conditional program A is the condition of behavioral program B, and so on. Moreover, it is possible to connect conditional programs to goal programs, e.g., if a particular goal is realized, particular consequences should follow. In this way, complex "webs" of decision programs can be designed, regulating the performance of production and control tasks. For instance, "tasks" of organizational members, teams, departments, business units can be described in terms of operations that should be performed under particular circumstances or as goals that should be realized.

Personnel

Luhmann considers *personnel* as a decision premise, because the training, indoctrination, professional competencies, and experience of organizational members function as a structure in organizations. These factors, guide our expectations about the expectations others have concerning our decision behavior (and the other way around). Knowing the training, competencies, or experiences of *who* makes the decision, in a number of cases also means knowing *what* the decision probably is about or what the decision probably will be. In fact, persons, with their particular training, competencies, or experiences may "embody" a huge variety of decision premises structuring an even larger variety of decisions. For this reason, the recruitment, allocation, training, and promotion of personnel has a structuring function in organizations. By means of personnel, organizations equip themselves to produce decisions on a variety of foreseen and unforeseen issues. For instance, if it is hard to specify decision programs, it may be convenient to leave decisions to the competencies and experience of personnel.

Position

The decision premises mentioned above come together in what Luhmann calls the "*position*" (German: "die Stelle"). A position is defined in terms of a *task* (in terms of goal and conditional programs), has a place in the network of *communication pathways*, and is occupied by a *member* of the organization's *personnel* (Luhmann, 1988a, b, p. 178). At a "position," therefore, multiple decision premises come together, structuring the set of "permissible" and "probable" decisions. The position can be called a "junction" of decision premises. If we know someone's position, we know – in general terms – what this person is supposed to do in the organization, we know what his or her place in terms of communication pathways is, and if we know the person holding the position, we may also know his or her experience and competencies. Based on this knowledge, all kinds of expectations about expectations may follow, functioning as decision premises within the organization. Because, multiple decision premises come together at positions, positions have an important *integrative* function in organizations.

Planning

Luhmann calls decisions that coordinate multiple decision premises "*planning*." Planning is necessary because of the "event-like" character of decisions. In our discussion of "meaning processing systems" we explained that the elements of social systems are "event-like." The moment they appear, they (start to) disappear. This has as a consequence that decisions at a particular moment in time only connect to prior and future decisions. Presupposing that organizations are not "sequential machines," producing one decision after the other, the question becomes how *multiple coherent decisions* can be produced at the *same* moment in time and at *different* moments in time. To decrease the probability of chaos, decisions are needed to coordinate decision premises regulating the production of both "synchronic" and "diachronic" decisions. Luhmann calls these decisions (that function as decision premises), planning.

Self-descriptions

According to Luhmann, it is impossible to both oversee all decision premises and decisions and consistently integrate them into an overall planning. This, however, need not be a problem. Organizations have other means of integrating decision premises. *Self-descriptions* function as a means to bring the multitude of decision premises to a unity. Self-descriptions are descriptions of the organization that use the distinction system/environment as a primary distinction. Examples are descriptions of the identity of the organization, its (desired or actual) image, organizational strategies, or codes of conduct. Such descriptions function as a point of reference for a multitude of decision premises, without exactly prescribing beforehand how each of these premises fits precisely into the larger whole.

Positions, planning, and self-descriptions are decision premises needed to *coordinate* multiple decisions occurring at the same and in the course of time. According to Luhmann, organizations also have decision premises needed to *change* decision premises. These are decision programs, communication pathways, personnel, planning, and self-descriptions relevant in the context of "planned" organizational change (of course, this does not mean that according to Luhmann all change in organization is planned).

The decision premises we discussed until now are all a result of decisions. For instance, it is decided that someone becomes a member of an organization. It is decided that this particular communication pathway is to be followed if some operation is to count as a decision. It is decided that producing and selling solar-energy is the goal of the organization. In general, we can say that decision premises in organizations are instituted or changed by decisions. However, according to Luhmann, there are also "undecided" decision premises in organizations: decision premises that do not directly result from a decision. These decision premises, for instance emerge as side effects of decisions. Luhmann distinguishes two types of "undecided decision premises": organizational culture and cognitive routines.

Organizational culture

Organizational culture refers to "undecided decision programs" structuring the way the organization handles its *own* decision-making (Seidl 2005, p. 44). For instance, suppose that a person who should be involved in a decision is always by-passed, this by-pass may become an undecided "way of doing things" in the organization. Or, if particular organizational members are always approached with particular questions, this may condense into an "undecided decision premise" co-structuring their position. According to Luhmann, basic societal values woven into the fabric of the organization as well as undecided "rules" for decent behavior or for the attribution of blame or praise are also examples of cultural decision premises. It is their function to structure decisions in cases where structuring by means of explicit premises is either impossible or undesirable. Cultural decision premises may both affect what "counts" as a decision in an organization and may function as "points of reference" for future decisions.

Cognitive routines

Cognitive routines are "undecided decision premises" regulating the production of decisions regarding the organization's environment. Luhmann gives the example of the ways to reach frequently contacted customers, the time usually required to get particular deliveries from suppliers, or presuppositions about the "normal" quality of production factors like machines or raw materials.[10] According to Luhmann, these cognitive routines are a by-product of ongoing practices in the organization's relation to its social and non-social environment. As a rule, cognitive routines appear and disappear together with the appearance and disappearance of these practices.

Because it is impossible to pinpoint either the "locus" or the "moment" of their emergence, "undecided decision premises" appear as "simply there" and as "ways of doing or seeing things." Although they are fundamentally contingent and changeable, in the organization where these "undecided decision premises" are active, this contingency and changeability are implicit. These premises are "there" as unquestioned, implicitly accepted ways of doing things.

It is possible to describe "undecided decision premises" in an act of reflection. For instance, by means of comparison to other organizations, an organization may become "aware" of (some of) its cultural decision premises. As a result, an organizational awareness may dawn of the contingency and changeability of these decision premises. However, this awareness does not make that cultural decision premises can be changed by a stroke of pen. On the contrary, in the context of planned change, cultural decision premises may prove to be very tenacious. In spite of their possible resistance to *planned* change, Luhmann does not foreclose that cultural decision premises and cognitive routines evolve over time. Changing societal values and norms and charismatic organizational leaders, ostentatiously breaking with existing undecided decision premises and investing new ones, not so much by means of decisions, but rather by the ways they act and the example they

[10]"Unter 'Kognitiven Routinen' wollen wir vielmehr Identifikationen verstehen, die für mehrfachen Gebrauch in Kommunikationen gespeichert sind und bei Bedarf abgerufen werden können. Zum Beispiel die Namen und Adressen und die Bedingungen der Erreichbarkeit von Kunden, mit denen man regelmäßigen Kontakt pflegt; oder die normale Zeit, die es kostet, um Lieferungen aus der Umwelt zu beziehen oder an sie ab zu geben; oder Annahmen über die typische Qualität von Dingen (etwa Maschinen, Rohstoffen, Verkehrsmitteln), mit denen die Organisation es regelmäßig zu tun hat" (Luhmann 2000, p. 250).

set, are, according to Luhmann, important drives for the evolution of organizational culture and cognitive routines.

What we have now is a listing of decision premises that, according to Luhmann, support the production of decisions by decisions in organizations. "Decisions" and the aforementioned "decision premises" define the class of organizations as a particular type of social systems. They describe, as it were, the "form" of all organizations. Note well, that Luhmann is a sociologist. He is not interested in the "rationality," "justice," or "meaningfulness" of decisions and decision premises. He just wants to reflect on the conditions of the autopoiesis of organizations and establish what decisions and decision premises develop in different (types) of organizations, under different circumstances, and at different times. Unlike the practitioners of cybernetics who are interested in changing and if possible improving (parts or aspects of) the world, for instance, by prescribing better designs for organizations, Luhmann contents himself with understanding and describing the processes of self-production of organizations as a particular type of social systems.

4.3.2.3 Some Remarks on the Function and Types of Decision Premises

In order to fine-tune our understanding of the function and types of decision premises, we need to make some remarks.

(1) On reading this overview of the different types of decision premises, one might be tempted to think that there is a transitive-hierarchical relation between them. For instance, one might think that organizational goals are first and that the rest follows. Luhmann emphatically denies this (Luhmann 2000, p. 226). More in particular, he argues that different types of decision premises presuppose each other. For instance, to specify a goal, one needs to follow a particular communication pathway. So, one could argue that the pathway is first. However, to specify pathways, one needs tasks, which in turn requires goals. According to Luhmann, it is impossible to find a type of decision premise that is first in a hierarchy of premises. Moreover, he argues that it is possible to use decision premises of different types as *substitutes* for each other. For instance, if a job is too complex to program it in detail, it is possible to fall back on the experience and competencies of personnel. In sum, according to Luhmann, decision premises contribute to the autopoiesis of organizations as web of mutually conditioning conditions.

(2) We argued that decision premises function as *normative points of reference* for the production of decisions. As such, they increase the chance of the production of decisions that are likely to contribute to the autopoiesis of the organization. Luhmann emphasizes that functioning as a normative point of reference does not imply that decision premises *cause* decisions or that decisions can be *deduced* with necessity from decision premises.

Given a particular decision premise, decisions are not its "causal effects" – it serves as a guideline for generating decisions. For instance, the conditional program, "If stocks decrease below 1250, order 200 new items" does not cause an order of 200 items if the stock actually decreases below 1250 – it is merely a heuristic.

That there is no necessary causal connection between a decision premise and decisions implies that operations may be generated that are recognized and connected as decisions deviating from that decision premise.

Moreover, that decision premises function as normative points of reference, according to Luhmann, means that decisions cannot be logically deduced from decision premises. In the first place, if such a deduction were possible, the decision would not be required. Deductive reasoning would be sufficient. In the second place, each new situation requires a new application of the decision premise to that situation. For instance, suppose that stock levels decrease below 1250, it may be warranted to deviate from the conditional program, for instance, if current demand is very low or the price of new items is temporarily high. That decision premises function as a normative point of reference means that they are taken into account in the process of generating, recognizing, and connecting operations as decisions to prior decisions.

(3) Given both the main function of the decision premises: regulating the production of decisions by decisions and their specific contribution to this main function, it is possible to subdivide sets of decision premises in two different ways.

First, we can classify them according to the way they are changed. As argued, some decision premises are changed by means of decisions ("decided decision premises") and others are changed implicitly in the wake of decisions ("undecided decision premises"). Because we already discussed the distinction between "decided" and "undecided decision premises," we pay no further attention to it.

Second, we can subdivide decision premises according to the type of contribution they make to the regulation of the autopoiesis of organizations. More in particular, some decision premises condition *THAT* operations can be generated, recognized, and connected *as* decisions to prior decisions (category 1) and other decision premises only condition *WHAT* decisions are likely to link up with prior decisions (category 2)

To explain the first sub-category of decision premises, we need to remind ourselves that elements of autopoietic systems are not "given." As indicated, "units" need to be generated, recognized, and connected *as* elements to other elements. Decisions are the elements of autopoietic systems called "organizations." As such, they are not "given" either. They are operations that have to be generated in such a way that they can be recognized and connected *as* decisions to decisions. Category 1 decision premises "ensure" that operations can be generated, which can be recognized and connected *as* decisions to prior decisions. Without them, mere operations could not be distinguished from decisions. As such, they are an absolute requirement for the existence of decisions constituting and producing the organization as a particular type of social system. Examples of this class of decision premises are "membership" and "communication pathways."

To explain what the second category of decision premises entails, reconsider the example of organizational goals provided above. In the example, the goal "producing and selling solar energy" functions as a decision premise for the decision to explore efficient means for storing solar energy. It provides a normative point of reference for what follow-up decisions should be about. Decision premises belonging to this second category only constrain the variety of the *contents* of decisions. By providing "points of reference" for decisions, they both increase the probability that decisions will be about particular issues and that deviations from the indicated points of reference are "marked." Examples of this category of decision premises are "goals" and "conditional programs."

4.3.3 Processes of Self-Production: The Autopoiesis of Organizations

Based on the discussion of decisions as elements and decision premises as the structure of organizations, we can now examine the process of production of decisions by decisions. At its core, this process is quite similar to that of meaning processing systems and social systems. For this reason, the lay-out of Fig. 4.9 is quite similar to that of Figs 4.6 and 4.7.

In an organization, we find at a particular moment in time (time = t) a set of "simultaneously" existing decisions "connecting" to prior decisions (that existed at time = t−1). For the sake of clarity, we depict only one of these decisions in Fig. 4.9. At that moment (time = t) the system has a structure, a "web" of decision premises regulating the production of decisions by decisions. As argued, these premises function as normative points of reference for the production of new decisions. They constrain the set of operations that can be produced to a subset of decisions, likely to contribute to the organization's autopoiesis.

Given prior decisions and decision premises, "new" operations can be generated (at time = t+1) that can be recognized and connected as "new" decisions (at

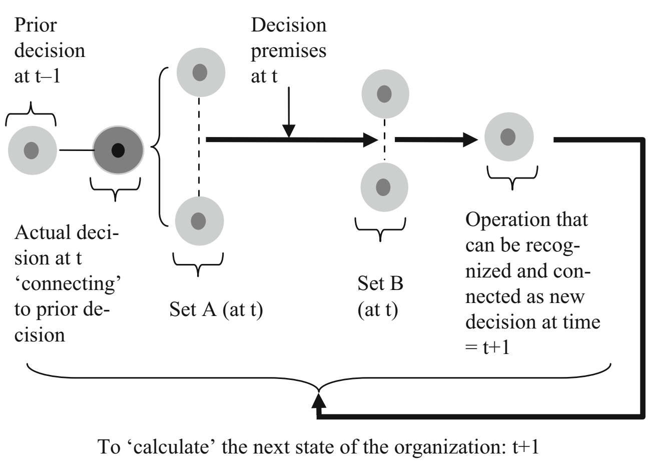


Fig. 4.9 The autopoiesis of decisions constituting organizations

time = t+1) to the now "prior" decisions (that existed at time = t). If such operations are actually generated, they can be treated (recognized) as decisions, and connected to prior decisions. These "new" decisions constitute the organization's elements at time = t+1. By means of the production of these new decisions, the organization's structure may change. For instance, a decision may be made to change a goal or to change a communication pathway. According to Luhmann, organizational change, in the end, is tied to change of the organization's structure, i.e., of its decided and undecided decision premises.

4.3.4 Questions Concerning Luhmann's Theory of Organizations

Luhmann's position is quite contra-intuitive and may lead to all kinds of questions. *We* sure had a lot of them when we read his texts! Below, we deal with three "frequently-asked-questions" regarding Luhmann's theory of organizations.

1. Luhmann asserts that organizations consist of decisions, but what about people, machines, buildings, money, communications, and non-communicative operations that are normally considered to be part of organizations?
2. Luhmann asserts that organizations consist of decisions. Does this mean that everything is decided in organizations?
3. Luhmann asserts that organizations consist of decisions. However, in other types of social systems, e.g., families, we also find decisions. Does this mean that families are organizations?

4.3.4.1 Question 1

The first question that may force itself on the "Luhmann reader" is, "What about "operations" like handling a machine or "normal" communications (a chat at the coffee machine), or what about "things" like people, machines, buildings, money, in sum, what about the "operations" and "things" we intuitively associate with organizations, are they not elements of organizations?"

Luhmann's answer to this question is "no." Organizations are a particular type of *social* systems. They consist of decisions producing decisions, and of nothing else. However, this does not mean that some of the things mentioned above (1) cannot contribute to the autopoietic production of organizations as "material conditions and processes," (2) cannot become the "object"/"content" of decisions, or (3) cannot be treated as decisions in organizations. Let us discuss the "contribution" of "operations" and "things" separately.

We begin with "*operations*" like handling a machine and "normal" communication in organizations. In the first place, such "operations" can "feed" the autopoietic production of organizations as "material conditions and processes" in its environment. As such, they are not "treated" as decisions. However, they may still be relevant. Organizational members may "do" or "fail to do" all kinds of things that are not organizationally treated as decisions (e.g., humming while working or forgetting to zip their fly) yet some of these things may still influence the conditions under which operations are generated, recognized, and connected as decisions to other decisions. For instance, my constant humming may irritate my colleague to such an extent that she refuses to share information, causing a delay in the execution of a conditional program. Although my humming may never be treated as a "decision" in the organization, it still changes the "material conditions and processes" contributing to the production of decisions by decisions.

In the second place, "operations" like handling a machine or "normal" communication may be the "object" of decisions. For instance, a decision may specify a particular conditional program for handling a machine. This program then, structures, follow-up "operations" that may be treated as decisions to either follow or deviate from the program (see below).

In the third place, "operations" like handling a machine or "normal" communication may be treated as decisions in the organization. Organizational members have expectations about the behavior expected from them. They "know" that complying with or deviating from these expectations can be organizationally treated as decisions. The things they do or fail to do may be and sometimes actually are treated in the organization as decisions that can be attributed to them and for which they can be held accountable. For instance, it is Peter's job to inspect the quality of the output of a machine. If the quality is sufficient, he is supposed to press a green button. If not, he has to press a red button, signaling a repair team. Peter "knows" that if he starts pressing buttons randomly, this behavior can be treated organizationally as a decision to deviate from the conditional program describing his "job." He "knows" that this can be attributed to him as a decision and that he can be held accountable for it (perhaps resulting in the decision to fire him). Moreover, he knows that pressing the right button may be treated as a decision to do his job well. In this way, "operations" like handling a machine and "normal" communication may be treated (recognized and connected) as decisions. In this sense, they are potential, and possibly, actual elements of organizations.

To sum up, "operations" like handling a machine or "normal" communication, can contribute to the organization's autopoiesis as "material conditions and processes" in its environment. They can be the "object" of decisions, and as such, may have a structural value for the production of follow-up decisions. And they can be treated as decisions constituting and contributing to the autopoiesis of the organization.

If we take a closer look at the role of "*things*" like people, machines, buildings, money, etc., we can in the first place say that they can function as "material conditions and processes" in the environment of the organization, contributing (as conditions) to the process of its self-production. For instance, people are needed as such a condition.[11] Or, suppose there is a conditional program specifying that "If

[11] In Luhmann's terminology 'people' have to be qualified as "interpenetrating organic and psychic systems."

machine X breaks down, its operator should call 71248". The machine actually breaking down, is an event in the world, that is a "material condition or process" contributing to the production of an operation that can be recognized and connected as a decision to prior decisions (e.g., actually calling 71248). In the second place, people, machines, etc. can become "objects" of decisions, and as such, they become organizationally relevant. For instance, personnel may be hired, developed, and fired; buildings may be built, maintained, and demolished; machines may be bought, repaired, and sold. All these things may be the "object" of decisions that, in turn, function as decision premises for the production of "connecting" decisions. In the third place, people, in the form of personnel, may function as a structure in organizations, regulating the production of decisions by decisions.

As it appears, Luhmann's social theory of organizations can vindicate our intuitive understanding of organizations. Decisions remain the elements of organizations. However, "operations" like handling machines and "normal" communications, do play a role in the process of self-production of decisions by decisions, and they may even themselves be treated as decisions in organizations. Something similar can be said about "things" like people, buildings, or machines. Although they are not operations that can be treated as decisions, they can still be the "object" of decisions and decision premises. Moreover, they can be "material conditions" needed for the production of decisions by decisions.

4.3.4.2 Question 2

The second question is whether Luhmann's assertion that decisions are the elements of organizations implies that everything in organizations is the result of decisions. Luhmann's answer to this question would be "no."

We argued that organizational members may "do" or "fail to do" all kinds of things. In principle, these things may be organizationally "treated" as decisions connecting to other decisions. The point is that this "treatment" as a decision can occur irrespective of the question whether the operation was or is a selection from a set of selectable options. An example may help. Suppose that I forget to attend a board meeting. In the organization, my "not showing up" can be treated as a decision, for instance, as a decision to deviate from established routines concerning board meetings. Although I had not decided to forget the meeting, my "not showing up" is still *treated* as a decision, possibly leading to follow-up decisions. Based on this example, we can draw the paradoxical conclusion that in organizations all kinds of "operations" (including omissions to do something) that are *not* decisions can still be treated *as* decisions, thereby *becoming* decisions (a communication of a selection as a selection). Or as Luhmann puts it, "Organizations are [...] *social systems* allowing themselves to treat *human behavior as if it were deciding*. As a consequence, the *social* reality of deciding in organizations is conceived of as a mere *assumption*, or a *supposition* or *suggestion* of those who take part in the system" (Luhmann 1993, p. 354; our translation).[12] Organizations are social systems that may treat human behavior *as if it were* a decision. Because members of organizations are "aware" of this, they "screen" their operations before performing ("generating") them: "What if this operation (e.g., coming too late, not attending the meeting, reporting ill for the third time in a month) were treated as a decision?" This kind of thought implicitly or explicitly accompanies all our operations in organizations (and certainly those verging on the edge of "breaking the rules"). It structures the "generation" of operations in organizations, amplifying the probability that they remain within the limits of established decision premises. To sum up our answer to the second question; that decisions are the elements of organizations does not imply that everything is decided in organizations. According to Luhmann, organizations can "*treat*" all kinds of operations (handling machines, gossiping, using the phone) *as if* they were decisions. Because of this, there is an increased probability that these operations remain within the range of what is "acceptable" given the organization's decision premises.

4.3.4.3 Question 3

Decisions are not unique to organizations. In other types of social systems, communications communicating a selection as a selection may be produced as well. So, the question may arise whether these types of social systems are organizations, too?

Take the example of a family planning a summer vacation. "Spending the summer vacation" is the "common denominator" functioning as a background for the generation of several options (Rome, Mantua, Naples). One of these options is selected and the selected option is communicated: "This year we're neither going to Rome nor to Mantua, but to Naples". In this example, we have a communication, communicating a selection as a selection. Does this make families into organizations?

Once again, Luhmann's answer is "no." To understand why, we have to take into account the *whole* of Luhmann's treatment of organizations as a particular type of social systems. Organizations are autopoietic systems, i.e., systems producing the elements of which they consist by means of the elements of which they consist, regulated by a structure. In order to characterize a social system as an organization it is insufficient to take into account its elements only. One also has to consider its structure and the process of the production of elements by elements. If we consider the structure of organizations, we see that it differs considerably from the structure of families.

Most importantly, organizations have "members." These members can be "hired" or "fired." "Hiring" and "firing" (just as "resigning") requires

[12]"Organisationen sind [...] *soziale Systeme*, die sich erlauben *menschliches Verhalten so zu behandeln*, als ob es ein Entscheiden wäre. Die *soziale* Realität des Entscheidens in Organisationen wird somit als eine bloße *Annahme* oder Unterstellung oder Suggestion der am System Beteiligten aufgefasst" (Luhmann 1993, p. 354).

organizational decisions about membership and these decisions are made against the background of (among other decision premises) membership rules. Moreover, organizational members hold "positions." At these positions, goal programs, conditional programs, communication pathways come together, prescribing what is expected of organizational members holding these positions (personnel). Without all these decision premises, the organization would not be an organization; these decision premises belong to the "form" of organizations as a particular type of social system.

In comparison, families have a quite different structure. Families also have members. But membership is not a matter of a family decision (you do not "hire" a child, even if you adopt it). It is impossible to "hire" and "fire" your family. You cannot "resign" from your family, as you can as a member of an organization. Families may formulate goals, conditional programs, or even communication pathways and positions. If they do, they become "awkward" simulacra of organizations. However, they do not become an organization. And what is more, unlike an organization that *must* formulate goals, conditional programs, etc. to be an organization, families do not *need* to develop such decision premises to be a family.

So, there are social systems – like families – that sometimes produce "operations" that have an inner structure that is the same as that of a decision. However, this does not necessarily imply that these social systems are organizations. To assess whether such social systems belong to the class of organizations, you have to critically examine the whole systemic "package"; elements, structures, and processes of self-production. If we do this in the case of families, we can see that their structure differs from that of organizations. And because a system's structure regulates the production of elements by elements, this difference will have an impact on both the elements of the family (as a particular type of social systems) and the process of the production of these elements.

4.4 Organizations as social systems reflecting complexity

Organizations are a particular type of social systems. This insight both retains and completely changes the intuition about organizations described in the Introduction to this book. There, we argued that organizations and communication are intrinsically connected. It is impossible to strip communication from an organization and still retain that organization. The social "arche" is not something that organizations "have too." Organizations are social through and through. In this chapter, we explained in Luhmann's theoretical terms what this means. We explained that organizations *are* a particular type of social systems. They consist of their "own" elements, have their "own" structure, and produce themselves by means of their "own" production processes. Organizations, in short, are a system-type of their *own*. They consist of decisions producing decisions. Decision premises both regulate the process by means of which decisions produce decisions and are (implicitly or explicitly) reinforced or changed in the course of this process. By

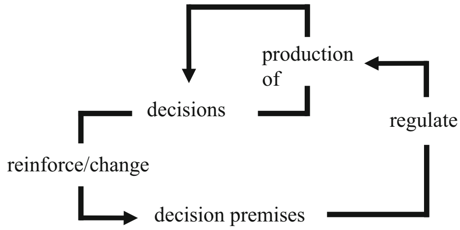


Fig. 4.10 Organizations, decisions, and decision premises

means of these processes, organizations produce themselves as a particular social system (see Fig. 4.10).

We also explained that organizations as a system-type of their own cannot be "reduced" to either allopoietic systems (like "machines" or "production systems") or other autopoietic systems (like "organic systems," "psychic systems," or other "social systems"). These types of systems can contribute to the self-production of organizations, and if they do, they exist in the organization's environment. However, they should not be confused with organizations.

For instance, "people" – "interpenetrating organic-psychic systems," in Luhmann's terms – are not elements of organizations. "We" as organic-psychic system exist in the organization's environment as one of the "material conditions" contributing to its self-production. Both "our" organic and psychic systems have their own elements, structures, and processes of self-production, and these should not be confused with the elements, structure, and processes of self-production of organizations. If a psychic system "influences" an organization, this "influence" is processed in terms of the *organization's* elements and structure: i.e., in terms of decisions and decision premises. And if an organization "influences" a psychic system, this influence is processed in terms of the *psychic system's* elements and structure. Although organic-psychic systems and organizations are related, they are interpenetrating systems, they should not be confused. Organizations and organic-psychic systems produce themselves and process knowledge in their own characteristic modes.

Because organizations are a system-type of their own, they have their own particular characteristics. These characteristics are linked to the decisions out of which they consist, the decision premises that regulate the production of decisions by decisions, and the processes by virtue of which organizations produce themselves.

In this final section, we want to highlight one characteristic feature of the social "arche" of organizations, which is especially important for the project of this book: thinking organizations as social systems conducting experiments with their own survival. Here, we just want to explain what this feature is. In the next chapter, we connect it to the experimental "arche" of organizations.

The characteristic feature we refer to is the way organizations, because of the inner structure of their elements, deal with variety, or "process complexity" as Luhmann would call it. To understand what "processing complexity" means and to understand how organizations "process complexity," we need to go back to the topic of "meaning processing systems" and to the distinction between decisions and "normal" communication.

According to Luhmann, all social systems (including organizations) are "meaning processing systems." This means that the elements of these systems have a focal content and marginally refer to the rest of the world. Producing a follow-up element involves *selecting* a "new" focal content out of all possible focal contents. Each time a meaning processing system produces a new element, it faces the problem of selecting a focal content out of a variety of possible (focal) contents. Each time such a system produces a new element it must deal with this variety of possible contents; it must process complexity.

As a particular type of social system, organizations are meaning processing systems. They too must, each time they produce a decision, select a focal content out of the set of all possible contents. However, as a system-type of their own, organizations are not just meaning processing or social systems. They are a *particular type* of social systems, characterized by their *own* elements that have their *own* characteristic inner structure. And it is because of this inner structure that organizations process complexity in a particular way.

In the previous section we argued that it is the inner structure of decisions to communicate a selection *as* a selection. Decisions communicate that:

1. there is a set of *contingent* selectable options (there could have been other sets as well)
2. out of this set a particular option has been *selected*
3. the selected option is *contingent* (other selectable options could have been selected as well).

What stands out in this "inventory" of the inner structure of decisions is that decisions "emphasize" "*selection*." Because they communicate a selection *as* a selection, they communicate that out of a variety of selectable options this particular option has been selected. Moreover, both the selected option and the variety of selectable options could have been different. Both the set of selectable options and the selected option are "*contingent*." Not only selection is emphasized by a decision, contingency is too. This emphasis on selection and contingency may amplify "*uncertainty*." "Have we generated the "right" set of selectable options?" "Have we selected the "right" option?" These are questions that, in principle, accompany all decisions constituting organizations. Amplified uncertainty, in turn, introduces the possibility to treat selections in organizations as a "*risk*." By "risk" we mean the awareness that a particular selection may have adverse side effects that perhaps could be avoided if another selection were made (of course, this "other" selection may have its own adverse side effects). Dealing with "risk" then means: weighing probable positive effects and adverse side effects of contingent selectable options, under conditions of uncertainty, and in a situation in which a selection must be made. In principle, all decisions in organizations can be treated as a "risk."

To sum up, because organizations are social systems that consist of decisions, and because decisions communicate a selection *as* a selection, "selection" and the contingency, uncertainty, and risk inherent in it tend to become an explicit issue in organizations. Because of the internal structure of their elements, organizations have the tendency to process complexity explicitly in terms of contingency, uncertainty, and risk. This makes organizations different from social systems consisting of “normal” communications. The elements of these systems just communicate a selection; they do not communicate a selection *as* a selection. For this reason, they do not tend to process complexity explicitly and in terms of contingency, uncertainty, and risk (although, from time to time, for instance, if things go wrong, they may happen to do so).

In addition to all the “other” characteristic features of organizations, their elements, their structure, their processes of self-production, organization have a tendency to process complexity explicitly in terms of contingency, uncertainty, and risk. Besides all kinds of “implicit” processes going on in organizations (for instance, the emergence of “undecided” decision premises), organizations are a type of social system “reflecting” in the elements out of which they consist the complexity that all meaning processing systems have to deal with. As indicated, this feature is important for connecting the social “arche” of organizations to their experimental “arche.” What this connection looks like is the topic of the next chapter.

References

Greshoff, R. (1999). *Die theoretischen Konzeptionen des Sozialen von Max Weber und Niklas Luhmann im Vergleich*. Opladen: Westdeutscher Verlag.

Kieserling, A. (1999). *Kommunikation unter Anwesenden. Studien über Interaktionssysteme*. Frankfurt am Main: Suhrkamp.

Luhmann, N. (1984). *Soziale systeme: Grundriß einer allgemeinen Theorie*. Frankfurt am Main: Suhrkamp.

Luhmann, N. (1988a). *Die Wirtschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.

Luhmann, N. (1988b). Organisation. In W. Küpper & G. Ortmann (Eds.), *Mikropolitik: Rationalität, Macht und Spiele in Organisationen*. Opladen: Westdeutscher Verlag.

Luhmann, N. (1993). *Soziologische Aufklärung 3*. Opladen: Westdeutscher Verlag.

Luhmann, N. (1995). *Social systems*. J. Bednarz, Jr. & D. Baecker (Trans.). Palo Alto, CA: Stanford University Press

Luhmann, N. (1998). *Die Gesellschaft der Gesellschaft, Bd. 1*. Frankfurt am Main: Suhrkamp.

Luhmann, N. (2000). *Organisation und Entscheidung*. Opladen: Westdeutscher Verlag.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition*. Dordrecht: Reidel Publishing Company.

Seidl, D. (2005). The basic concepts of Luhmann’s theory of social systems. In D. Seidl & K. H. Becker (Eds.), *Niklas Luhmann and organization studies*. Copenhagen: Liber & Copenhagen Business School Press.

von Foerster, H. (1993). *KybernEthik*. Berlin: Merve Verlag.

Chapter 5
Epilogue to Part I: The Two "Archai" Combined

5.1 Introduction

In Chap. 1 we advanced the position that organizations have an experimental and a social "arche." These "archai" are features of organizations that cannot be negated without negating organizations altogether. They are unavoidable characteristics of the "phenomena" we call organizations. Following Aristotle's "method" of starting with the phenomena as we experience them, we introduced the "archai" referring to everyday experiences with organizations.

In the chapters that followed, we explored the experimental and social "arche" separately. We formulated them in terms of the "languages" of (first and second-order) cybernetics and social systems theory. This resulted in a "theoretical" understanding of our everyday experiences with each of the two "archai." However, this leaves us with a *separate* understanding of the "archai" which is still insufficient to theoretically understand organizations as social systems conducting risky experiments. For this reason, we need to take one final step in which the two "archai" are *combined*.

Taking this final step is the objective of this epilogue. In it, we resume the position advanced in Chap. 1. However, now, we state it in terms of cybernetics and social systems theory, combining them to provide a theoretical understanding of organizations as social systems conducting risky experiments with meaningful survival.

In order to realize this objective, we start, in Sect. 5.2, by summarizing the core concepts of the theoretical languages of cybernetics and social systems theory. We do this to characterize these languages and to show how they can be combined into one theory. In Sect. 5.3 we actually combine cybernetics and social systems theory to theoretically describe organizations as social systems conducting experiments with their own survival. Finally, in Sect. 5.4 we make the transition to Part II of the book. In that part, we unfold functional and specific principles supporting the design of organizational infrastructures enabling organizations to conduct their experiments with meaningful survival.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_5, © Springer-Verlag Berlin Heidelberg 2010

5.2 Cybernetics and Social Systems Theory; an Exploratory Comparison

As indicated, it is the purpose of this chapter to combine the separate theoretical descriptions of the experimental and social “arche” provided in the previous chapters into a theoretical understanding that can make sense of our experiences with organizations. In order to become sensitive to the opportunities they offer for the formulation of a combined theory of organizations as social systems conducting risky experiments, we need to devote some attention to the languages, the conceptual devices, we made use of to describe these two “archai.”

To explore them, we start with a summary of the goal and core concepts of cybernetics and social systems theory. Next, we explore how they can be combined fruitfully. Based on this exploration, we combine the cybernetic description of the experimental “arche” and the social systemic description of the social “arche” into a theoretical understanding of organizations (Sect. 5.3).

5.2.1 The Goal and Core Concepts of Respectively Cybernetics and Social Systems Theory

Ashby formulates the goal of cybernetics right at the start of his introduction to cybernetics: cybernetics hopes to provide “effective methods for the study, and control, of systems that are intrinsically extremely complex” (Ashby 1958, p. 6). However, if we ask, “What is the ultimate goal of cybernetics?” Ashby would probably say, “Providing methods for the control of systems that are intrinsically extremely complex”. Studying systemic behavior is only a prerequisite for controlling it. Ultimately, cybernetics wants to provide methods to effectively control complex systems.

It realizes its goal by focusing not on “things” but on behavior. To quote Ashby, “Cybernetics, [...] treats, not things but *ways of behaving*”. It, “...deals with all forms of behavior in so far as they are regular, or determinate, or reproducible. The materiality is irrelevant...” (Ashby 1958, p. 1). In this sense, cybernetics treats “all possible machines,” irrespective of their actual (past, present, or future) realization. What it has to offer is, “...a framework on which all individual machines may be ordered, related, and understood” (Ashby 1958, p. 2).

As it appears, cybernetics wants to prescribe methods to effectively study and control complex systems by focusing on behavior. Moreover, it does not focus on this or that regular behavior, describing, for instance, the specific regular behavior of social systems, or mechanical systems, or psychic systems. It rather wants to provide a general framework allowing for the study and control of all possible regular behavior. Now, what are the core concepts Ashby uses to realize this objective?

A short survey of Chap. 2 suggests that in order to *study* behavior cybernetics needs to be able to:

1. *Describe* behavior: it does so in terms of "variables" combined into a "system"; "transitions" and "transformations," allowing for the description of changes in the value of variables; "parameters," i.e., factors influencing transformations
2. *Assess* the regularity of behavior*:* to this purpose Ashby introduces the concept of the "machine" (determinate or Markovian, without or with input)
3. *Find* regular behavior: to this purpose it uses the Black Box Method, combining the concepts mentioned before.

Controlling behavior is about:

1. *Setting goals*, i.e., selecting "essential variables" and specifying their "desired states"
2. *Designing* a mechanism to realize these goals by attenuating the "variety" of "disturbances" as much as possible, and amplifying the "variety" of "regulatory actions" as much as needed
3. *Operationally regulating* the realization of the desired states of the essential variable by blocking the flow of "variety" from the disturbances to the essential variables.

In sum, the conceptual core of cybernetics can be characterized as in Table 5.1.

Table 5.1 (Ashby's) cybernetics, goal and core concepts

	Cybernetics
Goal	Cybernetics hopes to provide "effective methods for the study, and control, of systems that are intrinsically extremely complex" (Ashby 1958, p. 6)
	It wants to offer a general and prescriptive framework enabling: the ordering, relating, and understanding of all regular behavior as well as the adaptation and realization of goals
Core concepts	System, variable, parameter, transformation, machine
	Essential variable, norm, disturbance, regulatory action
	Variety, control, design, operational regulation

If we turn our attention to Luhmann's theory of organizations (Luhmann 2000), we explained that its goal is to clarify their "Eigenlogik." By this he means: (1) understanding how organizations produce themselves, i.e., distinguish themselves from their environment, and (2) describing the distinctions they produce to support this process of self-production. As argued in Chap. 4, Luhmann distinguishes this goal from attempts to find the "essence" of organizations, normative theories of organizations, and empirical descriptions of all kinds of features of organizations.

Luhmann conceptualizes an organization as a particular type of social system, consisting of decisions producing new decisions against the background of decision premises. To understand what this means we need concepts from Luhmann's general theory of social systems and his specific theory of organizations.

In Luhmann's general theory a social system is conceptualized as an autopoietic, meaning processing system, consisting of communications as elements that produce, generate, recognize and connect to new communications. Expectations about expectations function as structures supporting the production of new communications by reducing the complexity, contingency, uncertainty and risk involved in this production process.

In his theory of organizations, Luhmann conceptualizes organizations as a particular type of social systems consisting of "decisions" as elements. Decisions are "communications communicating a selection as a selection." "Decisions premises" are the structures providing guidance for the production of decisions by decisions. Because decisions communicate selections as selections, complexity, contingency, uncertainty, and risk may be and usually are reflected in organizations, i.e., they become an explicit issue requiring explicit control.

Based on his general theory of social systems and his specific theory of organizations, we can summarize the goal and conceptual core of his organization theory in the table below (Table 5.2).

Table 5.2 Luhmann's organization theory, goal and core concepts

	Luhmann's organization theory
Goal	Clarifying the processes of self-production of organizations as a particular type of social systems
Core concepts	Decision (communication communicating a selection as a selection), decision premise, complexity, contingency, uncertainty, risk

5.2.2 Cybernetics and Social Systems Theory Compared

Complexity seems to be a problem or theme shared by both (first and second-order) cybernetics and Luhmann's social theory of organizations. However, the two conceptual devices "process" this problem in terms of their own goals and concepts.

For Ashby, the ultimate goal of cybernetics is providing effective methods supporting the control of complex behavior. The behavior he is talking about, is not this or that behavior, e.g., the behavior of frogs or rats, or behavior given the known laws of physics, it is about all possible behavior independent of the specificity of "things" or laws of physics known to us. Ashby's cybernetics aims at full *generality*, searching for effective methods that are valid in all possible worlds.

Moreover, by means of the methods it develops, cybernetics wants to support adapting and realizing goals. In this sense, its purpose is to support control. If one wants to effectively control systemic behavior, one should use the methods developed in cybernetics; cybernetics thus aims at *prescription*.

Luhmann's social systemic theory of organizations is devoted to a quite different purpose. It aims at *describing* the "*Eigenlogik*" of the autopoiesis of a particular type of social systems: organizations. Neither intervention nor prescription is a purpose of Luhmann's theory. He wants to *understand* how the process of the autopoiesis of this type of systems works.

Because of the specificity of its object – organizations – Luhmann's theory does not aim at generality in Ashby's sense. On the contrary, he uses his general theory of social systems to describe the specifics of the autopoiesis of organizations. Instead of aiming at generality, Luhmann aims at *specificity*.

Not only the goals of Ashby's cybernetics and Luhmann's social systemic theory of organizations are different, the conceptual systems that realize these goals differ too. Ashby's "systems" are variable systems. Luhmann treats "systems" as element-systems. In Ashby's conceptual system, the behavior of variables is affected by parameters. Luhmann treats the production of elements by elements, which is regulated by the system's structure. Table 5.3 provides an overview of the shared basics of and differences between cybernetics and Luhmann's social systemic theory of organizations.

Table 5.3 Cybernetics and Luhmann's social systemic theory of organizations

	Ashby's cybernetics	Luhmann's social systemic theory of organizations
Shared basic problem	Complexity	Complexity
Goals	Providing methods for effectively studying and controlling intrinsically complex systems	Clarifying the "Eigenlogik" of the self-production of organizations as a particular type of social systems
	Supporting prescription	Supporting understanding
	Aiming at generality	Aiming at specificity
Core concepts	Systems are described in terms of: – Essential variables – Parameters – Disturbances – Regulatory actions	Systems are described in terms of: – Elements: decisions – Structures: decision premises
Important systemic activities	Control, design, operational regulation, transformation	Autopoietic production: generation, recognition, connection

From their comparison we can learn that cybernetics and Luhmann's organization theory have characteristics that are complementary. Cybernetics aims at the improvement of methods supporting control. It is prescriptive and aims at generality. It is about prescribing how to support the control of the behavior of whatever system. Luhmann's theory of organizations is descriptive and aims at both clarification and specificity. It is about understanding the self-production of organizations. These complementary features may be an advantage.

To begin with, cybernetics may be complemented by Luhmann's theory of organizations. For if we think about controlling an organization, cybernetics, because of its generality, does not tell us what it means to control this particular type of social system. It provides methods supporting the control of whatever system, irrespective of its specific mechanical, biological, social, or organizational characteristics. Still, a pre-understanding of the specific characteristics of the system that is the "object" of control is important, for controlling social systems or organizations is quite different from controlling mechanical or biological systems. Luhmann's social theory of organizations can deliver such a pre-understanding. It provides a theoretical understanding both of the process of self-production of organizations as a particular type of social systems and of the type of social system that emerges in the course of this process of self-production. If we want to support organizational experiments with meaningful survival, it is important to understand the particular characteristics of organizations. Luhmann's social theory of organizations is an advanced theory providing this type of understanding. In this way, the generality of cybernetics may be complemented by the specificity of Luhmann's organization theory.

Conversely, Luhmann's organization theory may be complemented by cybernetics. As a purely descriptive theory, Luhmann's theory does not tell us a thing about how to effectively support organizational experiments with meaningful survival. This type of knowledge can be provided by cybernetics. Cybernetics is about supporting the control of whatever behavior. As such, it may deliver important insights into functional and infrastructural requirements needed to conduct these experiments. In this way, the prescriptive potentials of cybernetics complement the descriptive character of Luhmann's theory of organizations.

Viewed in this way, combining cybernetics and Luhmann's theory of organizations into a theory describing organizations as social systems conducting experiments offers advantages each of the theories in isolation cannot offer.

5.3 Combining the Experimental and the Social "Arche"

In this section, we address the question how the cybernetic description of the experimental "arche" and the systemic description of the social "arche" can be combined to provide a more theoretically grounded description of the phenomena we discussed in Chap. 1 of this book.

In that chapter (Sect. 1.2), we stated that organizational communication is about or refers to objects figuring in the experiment. Moreover, we argued that selections concerning these objects, e.g., the selection of a goal or an infrastructural arrangement, "condition" organizational communication.

Based on cybernetics and social systems theory, we now have a theoretical understanding of the experimental and the social "arche." First-order cybernetics provides a more precise definition of regulation and its different types. Given these types of regulation, their function in the experiment can be described, providing an account of the activities and objects involved in it. More in particular, experiments with meaningful survival involve: (1) selections with regard to goals: strategic regulation; (2) selections with regard to organizational infrastructures: regulation by design, (3) selections with regard to the regulation of transformations: operational regulation and (4) selections with regard to the performance of transformations constituting the organization's "*raison d'être*." Second-order cybernetics highlights the contingent, and therefore risky, character of the models underpinning organizational experiments with meaningful survival.

Social systems theory argues that organizations are a particular type of social systems consisting of decisions. Decisions are a particular type of communications:

communications communicating a selection as a selection. Organizations are autopoietic systems; the decisions out of which they consist are produced by these decisions. According to social systems theory, these production processes involve: (1) the *generation* of operations that can be recognized as decisions, (2) the actual *recognition* of these operations as decisions, and (3) the *connection* of these operations as decisions to prior decisions. Generation, recognition, and connection, require a background of decision premises that structure these activities. According to Luhmann, different types of decision premises are constitutive for organizations as social systems. More in particular, he mentions: membership, goal programs, conditional programs, communication pathways, positions, personnel, planning, self-descriptions, culture, and cognitive routines.

Given the theoretical descriptions of the experimental and social "arche," we can restate their relation in more precise terms.

In general, we can say that (1) decisions may be about or refer to objects figuring in the experiment and (2) as decision premises, selections regarding these objects structure the production of decisions.

The selections communicated by a decision may be about or refer to objects figuring in the experiment. This means that they may be about goals, infrastructural arrangements, operational regulation, or performing primary transformations. For instance, it is decided that a particular goal will be pursued or a particular machine will be purchased, i.e., selection about an object in the experiment, is communicated as a selection.

Once a selection has been communicated as a selection, it may either become a new decision premise or reinforce or adapt existing ones. For instance, the decision to pursue a particular goal may reinforce or adapt a particular goal program. Communicated selections regarding objects in the experiment, therefore not only figure in the experiment, they also appear as decision premises in organizations as social systems. As decision premises selections regarding these objects structure the production of new decisions about these objects.

In Chap. 1 we discussed the relation between organizational communications and selections about objects figuring in the experiment in terms of "conditioning": selections regarding these objects condition organizational communication (see Fig. 1.5, Chap. 1).

Conditioning – in Luhmann's social systemic language – means that decisions involving objects in the experiment are generated, recognized, and connected under reference to decision premises retaining earlier decisions regarding these objects. These decision premises, *structure* the production of new decisions regarding these objects. In Luhmann's theory this structuring relation has three dimensions.

First, new decisions about "experimental objects" explicitly refer to or are about decision premises specifying selections made earlier. For instance, in an organization, a new position is defined replacing two old ones. In Chap. 1, we called this "direct conditioning": organizational communication is about or refers to earlier selections regarding experimental objects.

Second, new decisions about objects figuring in the experiment are produced against the background of existing decision premises, presupposing these premises as given. For instance, the new position of the previous example is specified given the existence of particular goal programs in the organization. In this example, these goal programs are not the object of the decision. They are presupposed as a "background" for the decision about the new position. In Chap. 1, we called this "indirect conditioning."

Third, decision premises are needed to ensure that operations can be generated and recognized *as* decisions in organizations. For instance, if the decision about the new position is to "count" as an organizational decision, it has to be taken by the manager of the department and not by one of the department's members. Decision premises such as membership, communication pathways, or positions allow for the generation and recognition of operations as decisions. In Chap. 1, we phrased this in terms of the organization's infrastructure conditioning the way organizational communication evolves in organizations; who is supposed to communicate with who about what subjects.

As it appears, the social "arche" and the experimental "arche" are intertwined in organizations. Decisions may be about or refer to objects in the experiment and decision premises structure the production of decisions about these objects (see Fig. 5.1).

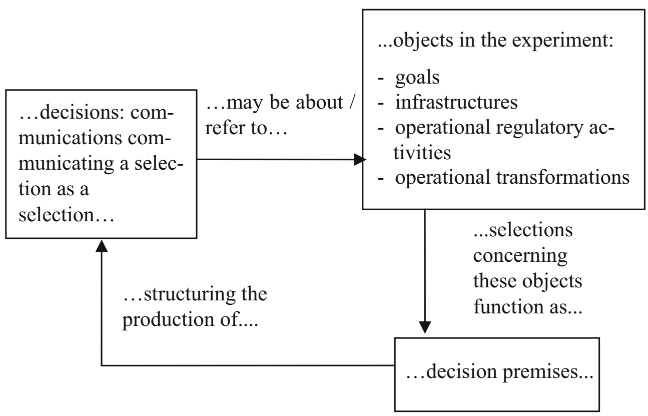


Fig. 5.1 Decisions, decision premises, and objects in the experiment

However, this intertwinement is not the whole story. In addition to being intertwined, the experimental and the social "arche" are also mutually conditioning.

Given the intertwinement of the experimental and social "arche," it is relatively easy to understand that the social "arche" is the condition for all possible organizational experiments with meaningful survival. Both the decisions constituting the organization and the decision premises structuring the production of decisions open up the social space for all possible selections regarding objects figuring in these experiments. Without them, these selections would not socially exist. Perhaps, someone may think of them or write them down. However, as long as they are not communicated as selections and/or structure the production of decisions, they have no organizational existence. In this way, the organization as a particular type of social system can be understood as the condition of the possibility, as the "platform" for all possible experiments with meaningful survival.

Although this may sound trivial, it is not. By explicitly stating that organizations as social systems are the platform for all possible experiments, the *social* complexity of these experiments in organizations can not be denied or overlooked. If cybernetic methods or models supporting such experiments are to be effective, they must take into account this social complexity. For instance, they must take into account that existing decision premises may stand in the way of the acceptance of methods or models that from a cybernetic perspective appear as highly desirable.

The social "arche" of organizations is not only the platform for all experiments it is also a condition for the reflection of the experimental character of these experiments in the organization. As indicated in Chap. 1 of the book, we take the term "experiment" not in its technical sense, referring to controlled scientific experiments. We rather use it to point at the contingent, uncertain, and risky character of selections regarding goals, infrastructures, operational regulation, and performing transformations. These selections can be seen as risky hypotheses of a kind about the relation between meaningful survival and actually selected goals, infrastructural arrangements, operational regulation, and operational transformations to realize them.

In organizations, being the particular type of social systems they are, the contingent, uncertain, and risky character of these selections is highlighted by the inner structure of the decisions that are their elements. As argued, a decision communicates a selection as a selection. Because of this, decisions either marginally or focally refer to contingency, uncertainty, and risk, involved in them, making them into (possible) issues in organizations. Therefore, given the inner structure of decisions, selections regarding the objects figuring in the experiment can be and often are treated in organizations as contingent, uncertain, risky, and "experimental." In this sense, the social "arche" is a condition of the possibility of the reflection of the experimental "arche" in organizations.

The experimental "arche," in a sense, also is a condition for the social "arche." By means of ongoing experiments, the organization realizes its *meaningful survival*. Without the experimental "arche," the self-production of the organization as a particular type of social system would become impossible or pointless Suppose that an organization is unable to make a meaningful contribution to its environment, its continued autopoiesis would become pointless, and in the end, it would stop.

Understood in this way, the relation between the social and the experimental "arche" can be compared to that between "being alive" and "living a more or less meaningful life." To live a meaningful life, we have to be alive, i.e., we need to be able to exercise the capacities characteristic of us as human beings. Given that we can exercise these capacities, we live a more or less meaningful life dependent on how we actually develop and use them. The meaningfulness of the life we actually lead can be said to be the point of our being alive.

The relation between the autopoiesis of the organization and its experiments with meaningful survival seem analogous to this relation between "being alive" and "living a meaningful human life." The organization's continued self-production as a social system, its autopoiesis, opens up the social space for all possible organizational experiments with meaningful survival. It endows the organization with the capacity to adapt and realize its contribution to its environment. However, its continued autopoiesis is not the point of being an organization. Its meaningful contribution to its environment is. Dependent on how the capacities for adaptation and realization of its contribution are developed and used, the organization's survival in its environment becomes more or less meaningful. In the end, more or less meaningful survival can be understood as being the point of its continued autopoietic production. Viewed in this way, the autopoietic production of the organization is only a beginning. It is a condition that must be fulfilled if organizational experiments with meaningful survival are to be conducted. However, whether the organization's survival is either actually meaningful or brought to an end depends on how the experiments are conducted given its autopoiesis.

To sum up, the social and the experimental "arche" are intertwined. The decisions constituting organizations are about or refer to objects figuring in the experiment. They communicate selections regarding these objects as selections. Once such a selection has been communicated, it may become a decision premise figuring in the autopoietic production of the organization. As decision premises, these selections structure the production of further decisions about or referring to objects figuring in the experiment. Moreover, the social and the experimental "arche" are intertwined in a mutually conditioning relation. The social "arche" is the condition for (1) all possible experiments with meaningful survival and (2) the reflection of the experimental character of these experiments in the organization. As long as this condition is fulfilled, i.e., as long as the autopoiesis of the organization continues, organizational experiments can take on all kinds of directions. Dependent on how organizations give shape to their capacity for the adaptation and realization of goals, their survival is more or less meaningful, which, in the end, is the point of and conditions their continued autopoiesis.

5.4 Transition to Part II: Designing organizations

From the previous sections it appeared that the cybernetics and social systems theory describe different "things" by different "means" with different "goals" in mind. Still, the experimental "arche" (described in cybernetic terms) and the social "arche" (described in social systemic terms) are intertwined as well as mutually conditioning. Given this intertwinement and mutually conditioning relation, the question becomes how organizations should be organized to continue their experiments with meaningful survival.

As argued in Chaps. 1 and 2, the potential to adapt and realize the organization's contribution to society crucially depends on the design of its infrastructure. Conditional on the design of the division of work, its human resources management systems, and its production and information technology, an organization is more or less in a position to perform activities related to control, design, operational regulation, and its primary transformations. So, the question becomes how this infrastructure should be designed to support these activities: what are the *principles* that should guide its design?

In Chap. 1, we argued that there are two classes of relevant design principles: functional design principles and specific design principles. These principles are the topic of Part II of the book. In Chap. 6, we unfold a set of functional design principles, using Beer's Viable System Model. These functional design principles specify the effects an organization's infrastructure should produce if that organization is to continue its experiments with meaningful survival.

Specific design principles provide rules for the design of specific "parts" of the infrastructure, i.e., for the division of work, for human resources management systems, or for the organization's production or information technology. Because we think that the division of work is crucial for an organization's potential to conduct experiments with meaningful survival, we specify specific design principles for this "part" in Chap. 7. To this purpose, we discuss De Sitter's sociotechnical theory concerning the design of organizational structures.

References

Ashby, W. R. (1958). *An Introduction to Cybernetics*. London: Chapman and Hall.
Luhmann, N. (1984). *Soziale Systeme*. Frankfurt am Main: Suhrkamp.
Luhmann, N. (2000). *Organisation und Entscheidung*. Opladen: Westdeutscher Verlag.

Part II
Designing Organizations as Social Systems Conducting Experiments

Chapter 6
Beer: Functional Design Principles for Viable Infrastructures

6.1 Introduction

In Part I of the book, we explored the two "archai" of organizations indicating that they are social systems conducting experiments. In the present part, we will give a systematic exposition of ways of organizing this experiment. Given the "logic" of the experiment, this means that we have to look for principles enabling the design of infrastructural conditions allowing organizations to experiment. These infrastructural conditions are important because an organization's potential to select and reselect goals, infrastructures, operational regulation, and transformation processes (and all other "objects" related to these "focal" objects), crucially depends on the design of its infrastructure.

Above, we distinguished two classes of design principles: functional design principles and specific design principles. Functional design principles specify what a system's infrastructure must be able to *do* if the system is to survive. Specific design principles specify rules and heuristics for the design of particular parts of the infrastructure (the division of work, human resources management, technology), given the set of functional design principles.

In this chapter, functional design principles are at issue. Ashby's model already provides a first and crude set of functional principles. In order to survive in its environment an organization must be able to:

1. Select goals: control
2. Select infrastructures: design
3. Select actions to regulate transformation processes: operational regulation
4. Perform transformation processes

However, it can be asked whether this model is *complete*. Does it specify *all* the functions needed to be able to survive? This is the question we want to answer in this chapter, "What functions are necessary and sufficient for an organization to be able to survive and what are the relations between these functions?"

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_6, © Springer-Verlag Berlin Heidelberg 2010

To answer this question, we introduce (a part of) the work of Stafford Beer. Building on Ashby's insights, Beer unfolded exactly the type of model we are looking for. In his books "Brain of the Firm" (Beer 1972), "The Heart of Enterprise" (1995), and "Diagnosing the system for organizations" (Beer 1996), Beer formulated a model specifying what a system must do in order to be able to survive. He called his model the Viable System Model, defining "viability" as "being able to survive."

To explain the Viable System Model, we selected "The Heart of Enterprise" (1995) as our central text. In this book Beer provides a rigorous treatment of the functions necessary and sufficient for viability. To discuss Beer's views on diagnosis and design, we selected "Diagnosing the system for organizations" (1996). In addition to these texts, we used texts by Espejo et al. who elaborated Beer's Viable System Model (Espejo et al. 1996).

The answer to the main question describes the functional principles needed to diagnose and design organizations conducting experiments. As such, the Viable System Model specifies the functional design principles for the first "arche" of organizations we want to discuss in this book.

However, based on Beer's description, it is not clear yet how the Viable System Model relates to the other "arche" of organizations: their social systemic "arche." So, we need to answer a second question in this chapter: "How can we relate the model describing the functional design principles of the experimental "arche" of organizations to the theory of organizations as a particular type of social systems?"

To introduce Beer's work on viable systems and to link it to organizations as social systems conducting experiments, we arranged the rest of this chapter into three sections.

In Sect. 6.2 we link viability to complexity. Viable organizations exist in a world that is indefinitely more complex than they can ever be. Yet, according to Ashby's theory (see Chap. 2), the success of a regulator in keeping essential variables within bounds depends on the relation between the variety of the disturbances and variety of the regulatory moves. If variables essential for viability are to be kept within limits, viable organizations somehow must cope with the huge variety of the world it is a part of. In Sect. 6.2, we discuss strategies available to viable organizations to deal with this complexity.

In Sect. 6.3 we describe the necessary and sufficient functions for organizational viability. Because the Viable System Model articulates the functions and the relations between them needed for viability, it provides a set of norms for diagnosing organizational viability. We also discuss these norms in Sect. 6.3.

Section 6.4 links the Viable System Model to the problem of organizing as a risky social experiment. It answers the second question about the relation between the functional model of the viable system and the theory of organizations as social systems.

6.2 Viability and Complexity

Beer's work on viable systems is firmly rooted in cybernetics. Right from the beginning, the problem of dealing with complexity dominates the discussion. To introduce this problem, Beer points at the complexity relations between an organization and its environment and an organization and its management. In this section, we discuss these complexity relations and the way organizations and their management deal with it.

6.2.1 Complexity Relations Between Environment, Organization, and Its Management

In our daily lives we take the existence of organizations for granted. However, Beer argues that it is actually more problematic than we think. To see why, it is useful to inspect the complexity relations between (1) an organization and its environment and (2) an organization and its management.

If we zoom in on the complexity relation between organization and environment, it is easy to see that the environment is more complex than the organization. The same holds for the complexity relation between organization and management. As a rule, organizations are more complex than their management. Of course, Beer never calculated the complexity of respectively all environments, organizations, and their management in terms of variety. What he argues is that just as we can tell that John is larger than Mary without measuring either John or Mary, we can tell that environments have a larger variety than organizations and organizations have a larger variety than their management. Fig. 6.1 depicts this relation between the varieties (the v's in the figures) of the environment, the organization, and its management. The curved arrows are "variety exchangers." The arrows pointing from the left to the right depict (disturbing) variety flowing to respectively the organization and its management. The arrows pointing from the right to the left depict (regulatory) variety flowing to respectively the environment and the organization.

Given the distribution of varieties and Ashby's law stating that only variety can absorb variety, we seem to face a problem. Organizations exist. We know this from our everyday experience. However, we also know that to remain viable, we need

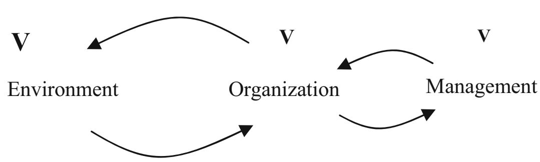


Fig. 6.1 Environment, organization, and management (adapted from Beer 1995, p. 95)

regulation, and to regulate, we need a variety of regulatory moves to block the flow of variety from the disturbances to the variables essential for viability.

Now, if the environment of an organization is more complex than the organization and if the organization is more complex than its management, how then is it possible that, given Ashby's theory, management can regulate the organization, enabling it to remain viable?

On the one hand, we have Ashby's law stating that if regulation is to be possible we need variety to absorb variety. On the other hand, we continuously encounter apparently viable organizations that are less complex than their environment and management that is less complex than the organizations it is supposed to manage. Somehow, organizations and their management must have found ways – *complexity strategies* – to deal with the complexity they face. Below, we discuss these strategies.

6.2.2 Three Strategies to Solve the Complexity Problem

Based on Ashby's theory, Beer formulates three strategies to deal with complexity, and thereby solves the apparent complexity problem: (1) defining goals, (2) attenuation and amplification, and (3) recursion.

6.2.2.1 Strategy 1: Define Goals

The first strategy builds upon the insight that it will not do to speak in general terms about a "mismatch" between the variety of "the environment" and the variety of "the organization" or about the "mismatch" between the variety of "the organization" and the variety of "its management" without knowing what is wanted.

Central to Ashby's description of regulation is that regulators block the flow of variety from disturbances to *essential variables*. What this means is that someone may say that relative to an organization, the environment is more complex and that relative to its management, the organization is more complex, but only a part of the complexity of the environment is relevant for the organization and only a part of the complexity of the organization is relevant for its management.

Which parts and *which* aspects are relevant depends on the selection of goals and the essential variables implied in these goals. Given a set of essential variables, only changes in some parts and aspects of the environment can displace the value of these essential variables. Changes in other parts and aspects will not have a detectable influence on these values.

In practice, this strategy means that organizations must specify their goals and thereby the variables they consider essential. Relative to these essential variables, only particular parts and aspects of the environment of an organization become a source of possible disturbances. These parts and aspects constitute the organization's relevant environment. Relative to the environment *tout court*, the relevant environment of an organization is less complex.

The same holds for the relation between management and the organization and its environment. Once management defines its goals, particular parts and aspects of the organization become relevant to management. Given the selection of these goals, the complexity of the organization is reduced to the complexity of the relevant organization.

6.2.2.2 Strategy 2: Attenuate Disturbances and Amplify Regulatory Potential

The second strategy Beer derives from Ashby's theory is that there are two ways of dealing with complexity once goals and essential variables are defined: *attenuation* and *amplification* (see Fig. 6.2).

Attenuation means the reduction of the variety of possible disturbances given the selection of essential variables. Beer distinguishes two domains of application of attenuation strategies.

Firstly, it is possible to apply attenuation strategies to the variety of disturbances coming from the environment to the organization. For instance, an insurance company may suffer all kinds of problems because many of its clients forget to fill in relevant items on the claim form. To attenuate this problem, instead of the old-fashioned paper form, it introduces a web page that only allows for the submission of claims if all the relevant items have been filled in. In general, attenuation of possible disturbances coming from the environment implies thinking of smart ways of designing interfaces with the organization's environment to make it work *for* you and not against you.

Secondly, it is possible to attenuate the variety of possible disturbances coming from the organization to its management. For instance, managers may seek to increase the problem-solving capacity of the people working for them. In this way, a part of the total variety of disturbances they had to face may be absorbed by their co-workers.

Thus, to keep an organization viable, management should design and implement attenuators to reduce as much as possible the disturbing variety the organization and its management have to cope with.

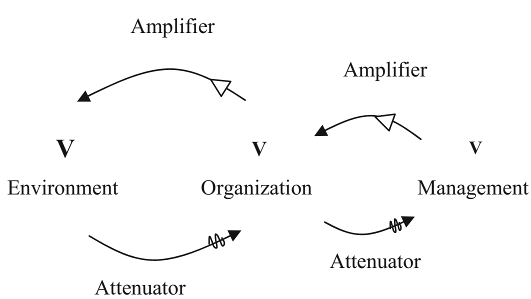


Fig. 6.2 Attenuation and amplification (adapted from Beer 1995, p. 96)

After attenuation of variety coming from respectively the environment and the organization, both the organization and its management may use amplification to cope with the residual variety of disturbances. Amplification means increasing the regulatory variety to a level needed to cope with the remaining disturbances, given the selection of essential variables and the design of relevant attenuators.

Just as in the case of attenuation, amplification has two domains of application. Firstly, it is possible to enhance the regulatory variety of the organization relative to its environment. An example of this type of amplification is the introduction of an IT application informing both a car manufacturer and his dealers on the stock-levels of spare-parts. By means of this application, the car-manufacturer amplifies his potential to deal with remaining problems with spare-parts after the attenuating measure of introducing the dealer system.

Secondly, it is possible to increase the regulatory variety of management relative to the organization. For instance, given the introduction of the attenuating program to enhance the problem-solving capacity of the people working for them, managers may amplify their own variety to deal with remaining disturbances by analyzing them and by discussing the best ways to deal with them. To cope with the residual variety coming from both the environment and the organization, one should design amplifiers increasing the regulatory variety of respectively the organization and its management.

Together, attenuation and amplification of variety further reduce the variety differences between respectively the environment and the organization and the organization and its management. In this way, the existence of viable organizations becomes possible instead of problematic. To quote Beer, "To design organizations that do not flout the canons of nature means designing appropriate attenuators of variety [. . .], but also it means designing appropriate amplifiers of variety too" (Beer 1995, 92).

Looking back on the first two strategies to solve the complexity problem, two remarks are opportune.

The first is that the two strategies are quite different in character. Attenuation and amplification are only possible *given* selected goals. Without these goals, it is impossible to speak about disturbances and regulatory variety to cope with these disturbances. Thus, specifying essential variables and goals ("control" in Ashby's terms) must not be confused with attenuation and amplification ("design" in Ashby's terms).

The second remark concerns the application of the two strategies: they stand in a relation of priority. The required order is as follows.

1. Specify organizational goals, i.e., essential variables and norms.
2. Design attenuators and amplifiers. To this purpose:
   – first think of possible disturbances and design attenuators to reduce the variety of possible disturbances as much as possible
   – next design the required amplifiers to deal with the remaining disturbances.

This order of priority is a heuristic. However, deviating from it may introduce different kinds of problems. For instance, implementing attenuators before specifying goals and essential variables may lead to ineffective situations. Implemented attenuators may deal with complexity irrelevant to essential variables that should have been selected. In a similar vein, implementing amplifiers before specifying relevant attenuators may lead to inefficient regulation. More amplifiers than strictly necessary may be installed.

6.2.2.3 Strategy 3: Recursion

Beer calls his third strategy for dealing with complexity: recursion. That is, modeling or designing a viable system as containing a set of viable systems, which in turn contain viable systems, etc.

To explain how "recursion" can be viewed as a strategy for dealing with complexity, we begin with giving a general description. Next, we show, by discussing an example, how recursion applies to viable systems and how it helps in dealing with complexity.

In order to understand the general concept of recursion, some concrete system S with particular characteristics – say, C1 and C2, is required. If within S, a subsystem exists that also has characteristics C1 and C2, we speak of recursion (with respect to these two characteristics).[1]

To understand what this means for organizations, it should be clear (1) which particular characteristics are used to identify recursion, and (2) what the nature of the sub-system is. In the case of Beer's theory, one relevant characteristic would be "the ability to maintain its viability," and the recursive sub-systems could be called "viable sub-systems." In his viable system model, he specifies this idea of recursion, by refining the particular characteristics (e.g., in terms of "defining goals" and "being able to attenuate and amplify"; or in terms of five interrelated functions). For the time being, it suffices to see that recursion with respect to some system with particular characteristics, has to do with identifying a subsystem within it with the same characteristics.

To explain this description and to show how it helps in dealing with complexity, suppose that we have an organization – our viable system in focus – named "Energeco." Energeco has its own primary activity defined in terms of goals and essential variables. In the case of Energeco, this primary activity is providing the market with eco-energy. Moreover, Energeco has its own management. Energeco's management has its own attenuators and amplifiers. Together, Energeco's primary activity and its management constitute the viable system Energeco.

[1] Recursion derives from "recurrere," which means "to run, walk back" in Latin. In mathematics or computer science, recursion is used to indicate that some function or routine calls itself. For instance, we can define F(n) (on $\mathbf{N}$) as (1) 1 if $n = 0$ and (2) $3^*F(n-1)+3$ if $n > 0$. This is a recursive definition. However, the general description given in text also applies to these functions: In the main function (the concrete system f(n)) we can find a subsystem (the recursive call f(n−1)) with the same characteristics (the defining aspects).

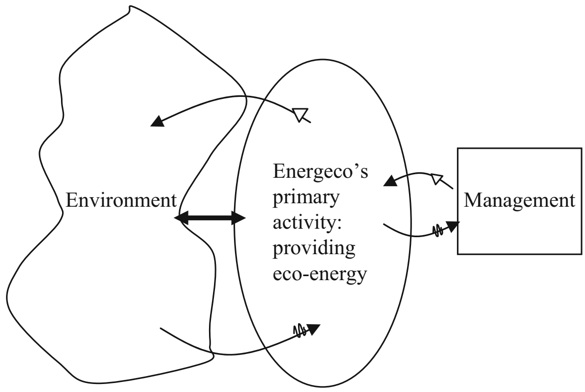


Fig. 6.3 Energeco's primary activity, its management and Energeco's environment

Given its primary activity and management, Energeco has its "own" relevant environment, input and output relations with this environment, and attenuators and amplifiers enabling Energeco to deal with its environmental complexity.

Figure 6.3 is a stylized depiction of Energeco as a viable system in its environment. The large oval shape in the center stands for Energeco's primary activity: providing eco-energy. Its management is represented by the box at the right hand side of Fig. 6.3. Taken together, the oval shape and the management box represent Energeco as a viable system. Energeco's environment is depicted as the large "shape" at the left hand side of the figure. The curved arrows are the attenuators and amplifiers between respectively Energeco's primary activity and the environment and Energeco's management and its primary activity. The straight bi-directional arrow represents input/output relations between Energeco's primary activity and its environment.

To further unfold the complexity Energeco has to deal with, it is useful to model Energeco as a viable system containing a set of viable systems. To this purpose, Energeco's is decomposed into three viable systems: (1) a viable system providing the market with wind energy, (2) a viable system providing the market with solar energy, and (3) a viable system providing the market with tidal energy.

Each of these viable systems has its own particular primary activity (fitting in and contributing to Energeco's primary activity), its own management, its own relevant environment, its own attenuators and amplifiers, and its own input/output relations. Thus the original viable system is modeled as a viable system containing a set of viable systems.

Figure 6.4 represents these viable systems contained within the larger viable system[2]. Once again, the (small) oval shapes represent primary activities. In the example: providing respectively wind, solar, and tidal energy.

[2] For purposes of presentational convenience, the new viable systems are drawn within the primary activity of the higher-level viable system.

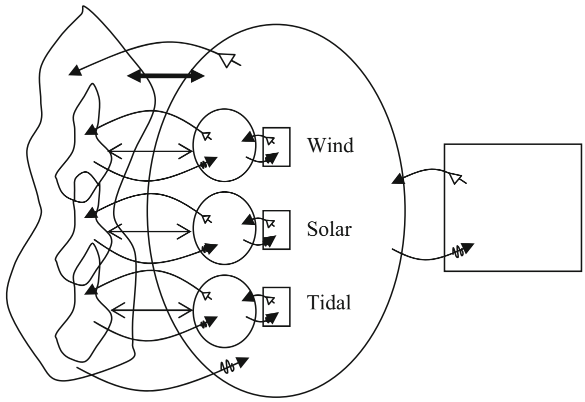


Fig. 6.4 Recursion: Energeco and viable systems contained in it

The small boxes represent the management of these primary activities. Just as in the case of Energeco as a whole, each of the management boxes has its own attenuators and amplifiers.

Combined, each of the oval shapes (primary activities) and boxes (management) constitutes a viable system within the larger viable system. Each of these viable systems has its own environment. These environments are represented by the smaller "shapes" within the environment of the larger viable system. Just like in Fig. 6.3, the curved arrows represent attenuators and amplifiers, and the double tipped arrows represent input/output relations between each viable system and its environment.

Now, if we take a closer look at Fig. 6.4 and zoom in on one viable system and its environment (for instance, on Wind energy and its environment), and if we compare the figure representing them with Fig. 6.3, we can see that the *structure* of both figures is the *same*. Energeco as a viable system in its environment has the same structure as Wind energy as a viable system in its environment. Below, we explain why this structural equivalence is important. Here we just point at it.

It is possible to continue decomposing Energeco into viable systems. To this purpose, we take one of the viable systems contained within Energeco, for instance, Tidal energy. This viable system can, in turn, be decomposed into a set of viable systems, for instance, one dealing with the northern and another with the southern part of the country. Once again, each of these viable systems has its own primary activity, its own management, its own relevant environment, its own attenuators and amplifiers, and its own input/output relations.

In this way, we see that the viable system "Energeco" contains a set of viable systems (Wind, Solar, and Tidal energy), each of which, in turn, contain a set of viable systems. Beer calls this unfolding of a viable system in focus into a set of viable systems, *recursion*. The different levels, "Energeco" as a whole, "Wind energy," "Region North," are called *levels of recursion*. By means of recursion, each viable system deals with its share of the total complexity Energeco has to deal with. To this purpose, each viable system has its own viable systems contained within it, its own essential variables, attenuators and amplifiers, and its own management. Within the larger whole, each viable system has its own bounded autonomy, enabling it to contribute its share to the viability of the larger whole. Given the Energeco example, we can now speak about recursion in more general terms. To understand Beer's conception of recursion two propositions are crucial:

1. A viable system is contained in a viable system and contains within itself a set of viable systems.
2. A viable system at a lower level of recursion has exactly the same type of "elements" and relations (primary activities, management, attenuators and amplifiers) between them as the viable system it is contained in.

What we, given these two propositions, get is something like a Russian doll. Each time you open it, you will find a similar but smaller version within it. In the case of a viable system, each time you open it, you will find a set of equally structured viable systems within it. Each of these viable systems contributes to the goals of the viable system it is a part of, has its own essential variables, and its own environment. Moreover, it has its own attenuators and amplifiers, and it contains its own set of viable systems and has its own management.

In the "Energeco" example we defined two levels of recursion. The first level is that of the viable system concerned with providing eco-energy. This system contains the set of viable systems: "Wind Energy," "Solar Energy," and "Tidal Energy." Then we went one recursion level down, focusing on "Tidal Energy." This viable system provides Tidal energy to its environment. It contains two viable systems "Region North" and "Region South." Relative to "Energeco" as a whole, this is the second level of recursion.

Recursion allows for designing organizations containing viable systems that both contribute to the goals of the organization and have viable systems, attenuators, amplifiers, and management to absorb a part of the total complexity the organization has to cope with. This means that these viable systems must have the freedom to manage their own affairs without interference of the organization as a whole. Without this freedom, these viable systems would not be able to absorb their share of the total complexity the organization has to deal with.

However, at the same time, the freedom of these viable systems must be constrained if they are to contribute to the goals of the organization as a whole. Without constraint, the viable systems contained in an organization would not be knit into a cohesive whole and would not contribute to the viability of that organization. As a rule for dealing with this delicate balance between freedom and constraint, Beer specifies that a viable system, "should make only that degree of intervention that is required to maintain cohesiveness in a viable system" (1995, p. 158).

Applied to the "Energeco" example, this means that the management of the organization as a whole should only interfere in the management of Wind, Solar, or Tidal Energy, if the activities or performance of these viable subsystems threaten the cohesion of Energeco as a whole.

Thus, to further reduce the complexity difference between respectively the environment and the organization and the organization and its management, organizations should build on the first two strategies and apply the *third strategy* of unfolding the viable system into a set of viable systems with their own goals, environment, attenuators and amplifiers, viable systems and management, absorbing their share of the total complexity the system they are a part of has to deal with.

Now, with regard to this third strategy of designing organizations as viable systems containing viable systems at least two questions may be asked. First, how should one go about it? Given an organization in focus, e.g., Energeco, how should the set of viable systems contained in it, be modeled or designed? And second, where does this process of modeling/designing viable systems within viable systems stop? If every viable system is contained in a viable system and contains a set of viable systems, then there seems to be neither an upper nor a lower limit to the "chain" of viable systems.

Beer's answer to the first question is that, in the case of organizations, one should look at the organization's contribution to its environment. This contribution is the organization's "*raison d'etre*." In the case of Energeco, its *raison d'etre* is: "providing its environment with eco-energy." Given, this contribution, viable systems contained in the organization should be designed according to two criteria.

Criterion 1:
Each viable system contained in an organization should by means of its primary activity actually participate in realizing the *raison d'etre* of that organization, and....
Criterion 2:
It must in principle be possible for each viable system contained in an organization to be "hived off" from the organization and exist as an independent organization.

Applied to Energeco this means that, for instance, "Providing the environment with Tidal energy" can be modeled as a viable system contained in Energeco. For this activity actually participates in Energeco's contribution to its environment (criterion 1). It realizes a part of "providing eco-energy": its output actually is a kind of eco-energy. Moreover, this activity in principle could exist as an independent viable system (criterion 2).

However, Energeco's accounting department, for instance, cannot be modeled as a viable system contained in Energeco. Of course, Energeco's accountants may decide to leave the company and start for themselves (criterion 2). If they decide to do so, they start an organization that has "providing accounting services" as its *raison d'etre*. However, this activity is quite different from "providing eco-energy." As such, Energeco's accounting department, in spite of contributing to Energeco as a whole, does not actually participate in realizing Energeco's *raison d'etre*. Its output is not a kind of eco-energy. For this reason, it does not meet criterion 1 and cannot be modeled as one of the viable systems contained in Energeco.

So, the answer to the first question is that one should design viable systems within a larger organization in such a way that their contribution reflects the larger organization's contribution to its environment. As such, these viable systems should, in principle, be able to exist as a separate organization.

Beer's answer to the second question regarding the upper and lower limits of the recursive unfolding of a viable system in focus is that there are no such limits. In short, the recursive logic on which this approach to the science of effective organization is constructed admits of no 'origin' and of no 'limits'. What counts as the origin, at any given moment, is the focus of our attention at that moment, which could be called Recursion level x. "The modeling of organization moves away from any [recursion level] x towards the microcosmic in one direction, and towards the macrocosmic in the other." (Beer, 1995, p. 312).

If one selects a viable system as the system in focus, it is by definition contained in a viable system and it contains a set of viable systems. Given the system in focus, it is therefore possible to model the viable system it is contained in and the viable systems it contains. Because, by definition, these viable systems are, once again, contained in a viable system and contain viable systems, it is *logically* possible to repeat this process ad infinitum both in the upward and in the downward direction.

Some critics of Beer's Viable System Model find it hard to accept this claim of "infinite recursion." At this point, we do not want to enter into discussions about its logical or practical possibility or impossibility. For the time being, we take the Viable System Model as a conceptual *tool* for designing organizations, and we ask ourselves how the idea of recursion can be used as a tool to the benefit organizational design.

First, designers should use to their benefit the possibility of "upward recursion." Organizations do not exist in a void. They depend on and contribute to a larger environment. The Viable System Model offers the opportunity to model (a part of) this larger environment as nested sets of viable systems to which the organization contributes (e.g., an organizational network, the economy, the nation, or the society it is a part of.). Modeling these viable systems and the organization's contribution to them is useful (not to say crucial). It forces designers to become aware of and reflect on, for instance, both the society the organization is a part of and its contribution to it. This awareness and reflection are needed to enhance their organizational design, amplifying the quality of the organization's contribution to the viability of the viable systems it is a part of. Later, in Chaps. 10 and 11, we return to this theme of improving the quality of the organization's contribution to society. Here we just point at the usefulness of the possibility of "upward recursion."

Second, designers should use the possibility of "downward recursion." By means of downward recursion, the organization can be modeled/designed as containing viable systems that contain viable systems. By modeling the organization in this way, designers are forced to become aware of and reflect on the way the organization as a whole deals with the complexity it faces. By a smart design of viable systems within viable systems, each level of recursion absorbs its own complexity, leaving only a complexity residue to the next higher level of recursion.

Regarding upward and downward recursion, Beer remarks that one should at least model the next higher and the next lower level of recursion of some given system in focus. The next higher level provides a context for the design problem. The next lower level is the object of design. We would like to add that each time an additional level is opened, one should ask whether the complexity added by opening that level adds to the insight it provides to the design problem.

Together, by these three strategies, systems can handle complexity. Although it may seem that the variety of environments is larger than that of organizations and that the variety of organizations is larger than that of its management, in fact organizations have three strategies at their disposal to "even out" these variety differences. In the first place, the selection of goals decreases the variety of the environment. The total environment – the rest of the world – decreases to the relevant environment. In the second place, attenuators and amplifiers respectively decrease the variety of disturbances and increase the variety of regulatory moves. In the third place, recursion allows for the definition of viable systems within viable systems, each absorbing its own share of the total variety.

6.3 A functional Model of Viability

In the introduction of this chapter, we mentioned that Beer provides a model of the functions necessary and sufficient for organizational viability. In this section, we unfold this model. First, we specify what it means to make functional models (Sect. 6.3.1). In Sect. 6.3.2, we describe the five functions necessary and sufficient for viability. Next, we describe how these functions should interact to actually achieve viability (Sect. 6.3.3). Together, the functions and their relations provide a set of norms for diagnosing and designing viable infrastructures. This is the topic of Sect. 6.3.4.

6.3.1 What Does it Mean to Make a Functional Model of a Viable System?

Before jumping into the functions required for viability, we need to pay some attention to the question what a function is and what it means to make a functional model. Understanding the answer to these questions brings us in a better position to know what to expect of a functional model of viability.

To describe what a function and a functional model are, we need to bring back to mind a distinction made in Chap. 2. In that chapter, we distinguished between control and design (see Sect. 2.3). Control refers to the activity of setting goals. Design refers to the activity of developing the "mechanism," the causes, allowing for the realization of the goal. This distinction between control and design revolves around the difference between goals (resulting from control) and the means to realize them (resulting from design). Now, goals are nothing but "desired effects" that can be brought about by "causes." In order to design whatever system, one has to think about what effects are desired and how these effects can be brought about.

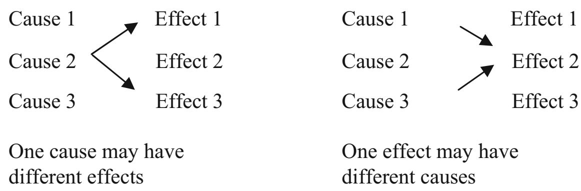


Fig. 6.5 Relations between causes and effects

To understand functions and functional models, it is useful to have a closer look at the relation between causes, effects, goals and means (see Fig. 6.5).

In the first place, it is important to see that one cause may have different effects and that one effect may have different causes. An example of a cause having different effects is a tent keeping the warmth in and the rain out. An example of an effect that has different causes is the sun and a stove causing a tent to warm up.

In the second place, goals and means are related to effects and causes. Relative to some value of a system or observer, some effects may be desired, others may be problematic, and still others may be indifferent. We call desired effects "goals," problematic effects "problems," and indifferent effects "neutral" (Christis 2002). "Things" or activities that cause desired effects we call means. Using these distinctions and relations we can now specify what a function is.

Generally speaking a function is a rule that allows for relating a "manifold" to a "unity" (Luhmann 1991, p. 14). Its general form is "...is X." Take the example of "... is blue." This unfinished proposition provides a rule for relating different objects, a manifold (e.g., the sky, the eyes of the babysitter, my car, this wall) to the property "being blue" (a unity). The rule is that the objects selected to finish the proposition are chosen in such a way that the proposition remains "true." We call the objects meeting this criterion *functionally equivalent*.

If we apply this general definition of function to the issue of design, i.e., to the issue of finding causes to realize desired effects, it can be argued that a function provides the rule to relate different things or activities (the manifold) as possible causes to a desired effect (the unity) or to relate different possible desired effects (the manifold) to something that is their cause (the unity).

We say for instance, "...is a cause of warmth." Given the desired effect (warmth), it becomes possible to define possible things that cause the desired effect (e.g., a fire, a stove, the human body). Or, we say "... is caused by fire." Given this particular something (fire), it becomes possible to define its possible desired effects (e.g., warmth, light).

If we design something, we constantly use such rules or functions. Either we select a desired effect and search for causes that may realize it. Or we select a cause and search for desired effects that may be realized by it. In both cases, we keep one side of the equation constant and look for instances that satisfy the rule. For instance, we define the desired effect and search for functionally equivalent causes or we define a cause and search for functionally equivalent desired effects.

Of course, in the process of designing we may smartly alternate the side we keep constant, now reasoning from desired effects, then from causes.

To sum up, applied to design, a function provides a rule for relating causes to a given desired effect or desired effects to a given cause. In this way they allow designers to select a desired effect and think about the different things or activities that cause the desired effect.

Given this definition of a function, it also becomes possible to specify what it means to make a functional model of a viable system (see Fig. 6.6). Making a functional model of a viable system means specifying the desired effects necessary and sufficient for the system to be viable. Once these desired effects are specified we have a rule that allows a designer to search for the many causes (activities and things) that may realize them.

In this sense, the functional model of the viable system is an "invariant" that can be realized in many different ways. Independent of the particulars of the material embodiment of the system, whether it is an organic, psychic, or social, if the system is viable, it should continuously produce the same set of desired effects necessary and sufficient for viability. Therefore, if we say that some specific system (e.g., an organism, an organization, or a state) is a viable system, we do not apply the term "viable" in a metaphoric sense to that system. We do not say that it is "*as if*" a viable system. What we do say is that the system in question *is* an instance, a functional equivalent, realizing the desired effects specified by the functional model.

That a functional model of the viable system is an invariant implies that once it is formulated, it is possible to describe different "embodiments" (e.g., particular organisms, social systems) as viable systems. More in particular, it becomes possible to describe and design organizations as a particular type of viable systems. In "The Heart of Enterprise", Beer (1995) accomplishes precisely these things. He unfolds a functional model of viable systems, describes organizations in terms of the model, and enables designers to improve organizational viability.

Given this specification of what it means to make a functional model of a viable system, we can also specify what to expect from it. A functional model of a viable system specifies the desired effects needed for viability. These desired effects function as a rule for identifying functionally equivalent causes that can bring about these desired effects.

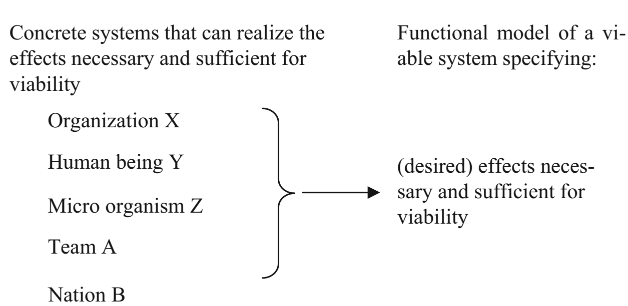


Fig. 6.6 Concrete viable systems and the functional model of viability

However, the functional model does not describe the specifics of these functional equivalents. For instance, in the case of organizations, a functional model of viability does not describe *how*, e.g., by what specific infrastructures, the desired effects are realized. It just describes effects possible infrastructures must realize for the organization to remain viable. In this sense, it provides a rule to determine whether a particular system belongs to the class of viable systems.

In all, a functional model lists the effects desired for viability. This list can then be used as a rule for the design of organizational infrastructures that can realize these effects. Given the functional model, we have a rule for designing concrete infrastructures that can realize the desired effects specified by the model. For this reason, a functional model of viable systems is a crucial ingredient in the process of designing infrastructures for viable organizations.

6.3.2 Beer's Functional Model of Viable Organizations

Now that we know what functional models are, we can start discussing Beer's functional model of viable systems and its application to organizations. The question Beer must answer is what functions are necessary and sufficient to enable an organization to maintain a separate existence in its environment. Beer's answer to this question is that it must be able to *realize* and *adapt* its goals. To this purpose, it needs exactly five functions. Functions one to three enable the organization to *realize* its goals. Functions three to five enable the organization to *adapt* its goals. Below, we discuss these functions and their relations.

6.3.2.1 Function One: Primary Activities

Function One is the collection of the primary activities of a viable system. These primary activities realize the goals constituting the viable system's *raison d'être* (Espejo *et al*. 1996, p. 110). In the example of "Energeco," its reason of existence is the supply of eco-energy. The primary activities of "Energeco" are: the supply of solar, tidal and wind energy.

In the Viable System Model, Beer models each of the primary activities collected in Function One as a viable system. This means that each of the primary activities has its particular goals, relevant environment, attenuators and amplifiers, and management. Thus, the viable system in focus "Energeco" consists of three viable systems, one aimed at supplying solar energy, one aimed at supplying tidal energy, and one aimed at supplying wind energy.

As a viable system, "Energeco" has its own relevant environment, i.e., the environment relevant for "delivering eco-energy." The three viable systems constituting "Energeco's" primary activities, in turn, have their own specific relevant environments. These are the environments relevant for delivering solar, tidal, and wind energy. The environment of "Energeco" as a whole is both wider than and

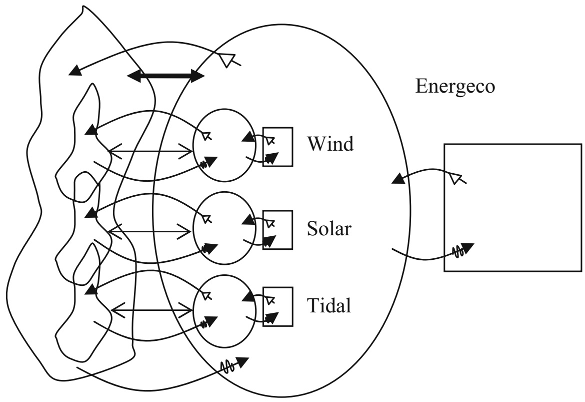


Fig. 6.7 Function one, primary activities (adapted from Beer 1995)

comprises the environments of the viable systems that are "Energeco's" primary activities.

Because each of the primary activities of a viable system is itself a viable system, it has all the functions necessary and sufficient for viability. For instance, the primary activity tidal energy of "Energeco" is a viable system that has a collection of primary activities (supplying tidal energy to region north and supplying tidal energy to region south). Moreover, the primary activity tidal energy has the other four functions required for viability. In this way, the Viable System Model not only describes necessary and sufficient functions for viability at one level of recursion, it describes the necessary and sufficient functions for viability at all levels of recursion.

In the rest of this chapter, we focus at the level of recursion of the organization ("Energeco") as a whole. The primary activities of "Energeco" are organized into three business units (Solar, Tidal, and Wind Energy). This collection of business units, realize Function One of "Energeco."

The existence of a collection of primary activities as such is insufficient to maintain the viability of an organization. In order to remain viable, it is necessary to forge the *manifold* of primary activities that have their own autonomy into the larger synergetic and cohesive *whole* of the organization. To this purpose, we need to "open" Energerco's "management box" (the large box at the right hand side of Fig. 6.7). In this box we find four additional functions needed to manage the identity of the viable system as a whole. Relative to the primary activities these functions are 'meta-systemic." Beer calls them: coordination, control, intelligence and policy.

6.3.2.2 Function Two: Coordination

The first of the functions directed at forging the primary activities together, has as its desired effect the coordination of interdependencies between these primary activities.

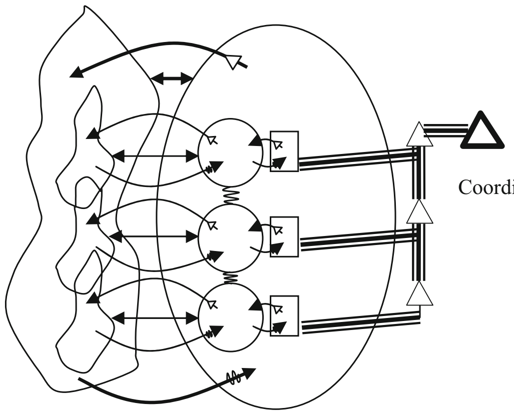


Fig. 6.8 Function two: Coordination (adapted from Beer 1995)

In organizations, primary activities may depend on shared resources and markets. For instance, primary activities may share a machine, or a team of specialists, or clients. Such dependencies make the primary activities interdependent. In Fig. 6.8, interdependencies between primary activities are depicted as twisted lines between these activities.

To give an example, specialists in high voltage energy are a shared resource between "Energeco's" business units. Now, suppose that there is no coordination between these business units. In this case, business units "Solar energy" and "Wind energy" may both "book" the same specialists for the same time frame for their "own" projects. This double booking may require the revision of the staffing of the projects. However, because there is no coordination between Solar and Wind energy, this revision may again cause problems, and so on. Without a function that supports the coordination of these interdependencies, the business units 'Solar Energy" and "Wind Energy" may become entangled in a process that oscillates between allocating and revising the allocation of high voltage specialists to projects. It is the task of Function Two to prevent these oscillations. Function Two facilitates the coordination of interdependencies between Function One activities, hence its name: coordination. In Beer's own words coordination, "has the highly specific function of damping oscillations, and nothing else. In fact, then, System[3] Two [the coordination function] is a service to System One [the primary activities]" (Beer 1995, p. 177).

In Fig. 6.8, the triangles depict coordination activities between the primary activities of function One. Examples of such coordination activities are; introducing

[3] Beer uses the word "System" instead of function. We have two reasons to depart from this usage. The first is that the Viable System Model is a functional model of viable systems, specifying functions required for viability. This terminology is consistent with other recent descriptions of the Viable System Model in literature (see for instance, Espejo, *et al*. 1996) The second reason is that we reserve the word system for the viable system as a whole.

quality standards, planning systems, introducing a logo or shared stationary to coordinate markets, developing a shared vocabulary facilitating the discussion of coordination problems between the primary activities. According to Beer, coordination is a "high variety" activity. This means that a lot of communication about a wide variety of topics is involved. To represent this "high variety" character of coordination activities, we used multiple lines between the primary activities and the coordination function (see Fig. 6.8).

6.3.2.3 Function Three: Control

The primary activities and coordination are necessary, but not sufficient for the viability of an organization. Each primary activity can still pursue its own goals without contributing to the realization of the identity and strategy of the viable system as a whole. For this reason, Beer argues that there must be a third function ensuring synergy and cohesion between primary activities. He calls this function: control. The control function, "is first of all typified by its metasystemic nature, and by the SYNOPTIC SYSTEMIC viewpoint from which it surveys the total activity of the operational elements [the primary activities constituting function One] of the enterprise. It is aware of all that is going on *inside* the firm and *now"* (Beer 1995, p. 202).

It is the task of control to translate the goals of the viable system (e.g., of the organization that supplies eco-energy) into goals for the primary activities (e.g., supplying wind, solar and tidal energy) and to monitor and regulate the realization of these goals. In this way, control takes care of the contribution of the primary activities ("inside") to the actual realization ("now") of the identity and mission of the viable system. To this purpose, control has to be a "synoptic" function, i.e., a function keeping an overview over the activities constituting the *raison d'etre* of the organization. Control has three "instruments" to discharge of this task.

First, it gives direct "*commands*" about goals that should be realized by each primary activities and it receives *management reports* about their realization. Moreover, it *bargains* with each primary activity about the resources to realize these goals. In Fig. 6.9, direct commands and reports and resource bargaining are depicted by the straight bi-directional arrows between control and the management boxes of the primary activities.

Second, control can *audit* the primary activities. Beer also calls these audits Function Three* (Function Three Star). Audits are required because the information control gets by means of reports from the management of the primary activities is general and may be flawed (e.g., management exaggerating production). Audits are aimed to keep in touch with what is really going on at the "shop floor" of the primary activities. By means of audits, control gets additional "first-hand" information about the operations of the primary activities and the way they are conducted. In Fig. 6.9, audits are depicted as a curved arrow pointing from the control function to the primary activities (for purposes of convenience only an arrow between control and the primary activity at the top is drawn).

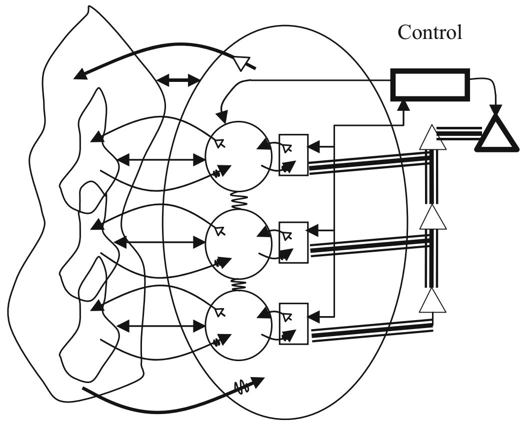


Fig. 6.9 Function three: Control (adapted from Beer 1995)

Third, control ensures the synergy of the primary activities by *controlling the coordination effort* by Function Two. This controlling relation is depicted in Fig. 6.9 as a curved arrow pointing from control to coordination. Below, as we discuss the relations between the functions of the VSM, we discuss these three "instruments" in more detail.

Control not only focuses on the realization of the identity and mission of the viable system by its primary activities, it is also involved in the adaptation of this identity and mission. Because of control's knowledge about the modus operandi and problems of the primary activities, its second task is to review proposals for innovation produced by the intelligence function (see below) and to assess whether these proposals can be realized given the potential for change of the primary activities. To this purpose, control is involved in a continuous discussion about the feasibility of proposed innovations.

6.3.2.4 Function Four: Intelligence

Functions One to Three are necessary for the realization of the identity and mission of viable organizations, yet they are not sufficient for organizational viability. To be viable also means, being able to adapt. To make adaptation possible, a fourth function is required. Beer calls this function intelligence. It is the task of intelligence to scan the environment of the organization for relevant developments and to initiate plans for innovation in such a way that the organization stays aligned with these developments (in Fig. 6.10 intelligence is depicted as the box above the control function). In Beer's own words, "the viable system that we are modeling must respond to this larger environment. System Three [the control function] cannot do this, because (under our definitions) it is concerned with inside and now. Thus we clearly need a new system, dedicated to the larger environment and to regulation in its regard, which environment I call the OUTSIDE and THEN. So this will be System Four [intelligence]" (Beer 1995, p. 227).

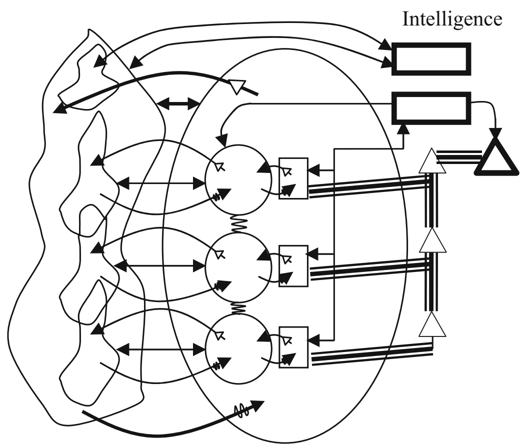


Fig. 6.10 Function four: Intelligence (adapted from Beer 1995)

To better understand the role of Function Four, we refer to the distinction between the relevant environment of the viable system in focus and the relevant environments of its primary activities. We stated that the relevant environment of the viable system in focus is both larger than and comprises the relevant environments of its primary activities. To paraphrase Beer (1995, p. 227), every viable system is involved in an environment that is wider than the sum of the environments of its primary activities.

In the "Energeco" example, this is the difference between the relevant environment of "Energeco" (the viable system in focus) and the relevant environments of solar, tidal, and wind energy (its primary activities). Because these primary activities themselves are viable systems, they have to deal with their own relevant environments. To this purpose, they have their own intelligence functions.

The intelligence function of "Energeco" as a whole focuses on its relevant environment. It deals with developments relevant for "delivering eco-energy" as such. Examples of these developments are the implementation of new legislation deregulating the energy market, the emergence of technology facilitating cost-effective, large-scale production of energy from bio-mass, or anti-globalists protesting against the dumping of nuclear waste in third world countries. Of course, these developments are also of relevance for "Energeco's" individual primary activities. For instance, the emergence of bio-mass production technology may be relevant to these primary activities as a threat. However, for "Energeco" as a whole, this technology may be relevant as an opportunity to expand its services portfolio with bio-energy.

Therefore, if Beer says that every viable system is involved in an environment that is wider than the sum of the environments of its primary activities, this does not only mean, in our view, that this wider environment contains more parties or relevant aspects. It does mean that these parties or aspects are relevant to the wider identity and mission of the viable system in focus and not just its primary activities.

Within this wider environment, Beer distinguishes between the *accepted* and the *problematic* environment (a distinction closely resembling that between the contextual and the transactional environment).

The accepted environment consists of those developments in the relevant environment of the viable system, happening to the viable system (in Fig. 6.10, the accepted environment is depicted as the large "shape" containing the environments of the primary activities). They are developments that given its identity and mission, a viable system cannot change. Managerial interest in these developments cannot be but largely reactive. The examples provided above of new legislation, a new production technology, and anti-globalist protests, are examples of developments in the accepted environment. Each of these developments constitute a given to the management of "Energeco," are of general interest, and may be apparent to any concerned citizen. Each one of them asks for a reaction from "Energeco." Intelligence should identify and follow these developments, assess them and develop plans to react on them.

The problematic environment is a different matter. It is not the relevant environment that is "happening to the viable system," but the environment that the viable system, given its identity and mission, should help shaping (in Fig. 6.10 the problematic environment is depicted as a "shape" with a question mark in it within the accepted environment). The problematic environment is "also outside the collection of System One environments; but especially concerned with a future that is the specific duty of the Energy Corporation to advance. That is to say that there is an area of concern in the Outside and Then that belongs – in a creative sense – to the viable system's responsibility" (Beer 1995, p. 227).

For this reason, managerial interest in this environment should be "wholly innovative – as deriving from special knowledge and dedication, from research and flair" (1995, p. 227). Given the identity and mission of the viable system, intelligence should pick up and filter out those developments in the wider environment of the viable system that can help to shape the future of both the organization and its problematic environment. In this sense, intelligence contributes to what Espejo et al. (1996) call, the *invention of the future*. Examples of this kind of contributions are finding and implementing innovative ways making the public conscious of the importance of using eco-energy or inventing technologies that will shape the competitive production of eco-energy.

6.3.2.5 Function Five: Policy

To complete the list of functions required for the realization and adaptation of a viable organization's identity and mission, Beer adds one more function: the policy function. This addition is motivated by the different perspectives of the control and intelligence function. Intelligence focuses on *adaptation* by initiating plans for innovation. It focuses on "outside" and "then". Control focuses on the *realization* of identity and mission of the organization. It focuses on the "inside" and "now." From this perspective, it assesses plans for innovation.

Given these different perspectives, it is important that control and intelligence debate plans shaping the organization's future. To develop plans both relevant "outside and then" and feasible "inside and now," intelligence and control should discuss them in-depth. It is important that these in-depth discussions are well coordinated. To show why this is so, let us consider three problems that can occur if the debate between intelligence and control would be left uncoordinated.

The first problem is a problem of balance that occurs if one of the two functions dominates the other. For instance, suppose that intelligence dominates control. In this case, the organization runs the risk of "innovatism." Innovative products or technologies are proposed and introduced that do not match the potential for change of the primary activities, resulting in failing projects of innovation. Now suppose that control dominates intelligence. In this case, the organization runs the risk of "conservatism." Control rejects the proposals for innovation intelligence comes up with because of an obsession with current markets, products, and production technologies. As a result, the organization will miss important opportunities for innovation.

The second problem is a problem of connectivity. Suppose that the communication between intelligence and control is not sufficiently intense. Proposals for innovations are not assessed in terms of the organization's potentials for change and existing potentials for change cannot be exploited by innovations. Both the realization and adaptation of the organization's identity and mission may come to a stop.

The third problem is a problem of complexity. Shaping the organization's future is a complex business, requiring the contribution of many experts. The communication between intelligence and control should allow for this complexity. Without it, plans for innovation become shallow and ill considered.

To counter these three problems, Beer introduces a fifth function that he calls policy. Policy has the tasks of coordinating the interaction between control and intelligence and consolidating its results by (re)defining the identity and mission of the organization in a way that fits both developments in its environment and its potentials for change. In Fig. 6.11, the ongoing discussion about innovation is depicted by the curved lines between intelligence and control. Policy's coordination of this discussion is depicted by the curved lines between policy and the ongoing discussion between intelligence and control. The arrows between policy, intelligence, and control stand for command and report lines (policy consolidating plans for adaptation). To coordinate the discussion between function three and four and to consolidate its results, function five does not need detailed knowledge about plans for innovation or how they affect the primary activities – this knowledge should reside within the intelligence and control function. It does, however, need a general insight in the organization's primary activity and its environment.

Policy gives closure to the cycle of adapting and realizing the organization's identity and mission. As such, it completes the list of necessary functions for viability. "We reach the final argument as to the logic of viability. There has to be System Five. That will do it. System Five will monitor the operation of the balancing operation between Three and Four. [...] Should anyone ask: "What

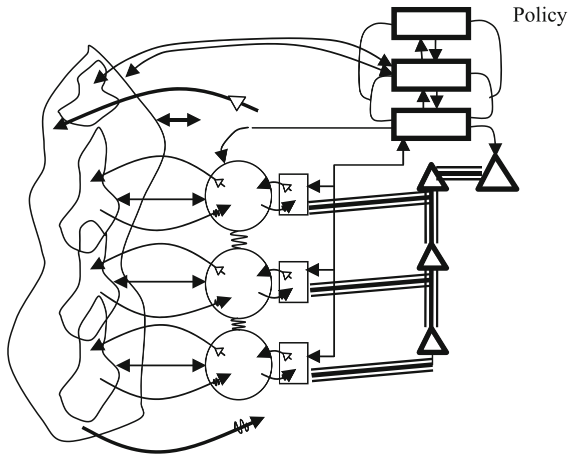


Fig. 6.11 Function five: Policy (adapted from Beer 1995)

happens beyond System Five? Is it System Six?," the answer is, no. What is beyond System Five is the next level of recursion, of which *this* fivefold viable system is an operational element" (Beer 1995, p. 259).

With the policy function, Beer completes his functional model of the viable system. In combination, the five functions not only are the necessary functions required for viability, they are also sufficient. No further functions are needed. Beyond Function Five is the viable system at the next higher level of recursion.

Before we proceed to explore the *relations* between the different functions specified in the model, we need to make one cautionary remark concerning functions and their instantiations.

Reading about "primary activities," "coordination," "control," "intelligence," or "policy" one may be tempted to forget that we are dealing with *functions* and *not* with departments, teams, or specific persons performing particular organizational roles (e.g., the CEO). For instance, one may be tempted to say, "Oh, policy is the organization's top management" or "Intelligence is the organization's R&D or Marketing department". However, using the VSM in this way would be a grave error.

The Viable System Model is a functional model of viable systems. As argued, this means in the first place that the Viable System Model specifies the functions necessary and sufficient for viability. These functions define desired effects that must be realized for a system to be viable. By specifying these desired effects, functions serve as a rule for *finding* things or activities that can realize these desired effects. In the second place, we argued that a desired effect may be caused by different causes and that different causes may realize the same desired effect.

Applying these insights to the relation between functions and organizational units (such as divisions, departments, teams, etc.), we can say that a function, e.g., "control," can be realized by different organizational units and that one organizational unit can realize different functions. For instance, in a consultancy firm, individual consultants may scan the environment of the firm for relevant developments. In addition to their contribution to the firm's primary activities, these consultants also perform an intelligence function. Or, one function, e.g., coordination may be distributed over different organizational units. So, it is impossible to make an a priori connection between functions and organizational units in the sense that a particular unit always performs a particular function or a particular function is always performed by a particular unit. In concrete cases, one must always be aware of what a particular unit does in order to interpret which functions it fulfills.

6.3.3 Relations Between Functions

The Viable System Model not only describes the functions needed for viability. It also describes the interactive relations that should exist between them and the characteristics these relations should have. To explain these relations, it is helpful to divide the five functions into two groups.

The first group consists of the primary activities, coordination, and control and *realizes* the organization's mission and goals. We call this group the "realization group." The second group consists of control, intelligence, and policy and *adapts* the organization's mission and goals. We call this group the "adaptation group."

The two groups are linked by control that is a part of both groups. In the group that is responsible for realizing the organization's mission and goals, control manages the synergy and cohesion of the primary activities. In the group that focuses on adapting the organization's mission and goals control concentrates on reviewing proposals for innovation. Below, we first discuss the relations between functions in the realization group. Then, we go into the relations between functions within the adaptation group.

6.3.3.1 Relations Between Functions Required to Realize the Organization's Mission and Goals

Within the realization group, Beer distinguishes four basic relations:

1. Direct commands and reports and resource bargaining between control and the primary activities (control-primary activities)
2. Audits of the management of the primary activities (control-primary activities)
3. Controlling the coordination function (control-coordination)
4. Coordinating interdependencies between primary activities (coordination-primary activities).

Direct Commands and Reports and Resource Bargaining Between Control and the Primary Activities

The first relation is the direct commands and reports relation between control and the primary activities. To maintain the cohesion and synergy of the primary activities, the identity and mission of the organization as a whole should be translated into goals for the primary activities and the performance of the primary activities should be controlled. By means of commands and resource bargaining, organizational goals and the means to realize them can be cascaded down into goals for the primary activities. By means of reports, the performance of the primary activities can be monitored. If reports show that the performance of the primary activities is not in line with the goals, control can intervene by means of direct commands.

Beer specifies that control should minimize the use of direct commands and reports. The reason for this is that the primary activities are viable systems that should deal with their own complexity as much as possible. In this way, each primary activity attenuates the variety control has to deal with. Control should only have to intervene if the synergy and cohesion of the viable system warrants this. This entails that management by means of commands and reports should be management by exception. Given general goals for the primary activities, control receives frequent (weekly, monthly), standardized, and aggregate performance reports from the primary activities. Only if these performance reports show significant deviations from the general objectives, control should intervene by direct commands. To attenuate the variety it has to deal with on the command and report axis, control provides the primary activities with the "freedom" to regulate their own activities.

Audits of the Primary Activities

The audit relation is the second relation between control and the primary activities. It is easy to see why audits are important. Suppose that aggregated performance reports from the primary activities are control's only source of information about what is happening in these primary activities. Now, without additional information control: (1) does not know whether the general picture described by these performance reports is correct, and (2) runs the risk of estranging itself from the actual operations and problems of the primary activities.

To solve these problems, Beer projects a channel between control and the primary activities. This channel by-passes the management of the primary activities and allows for relatively detailed scrutiny of and communication about their actual operations and performance. By means of the channel, control should be able to develop a more detailed picture of what is going on at the operational level of the primary activities, verifying or falsifying the management reports sent on the command and reports axis. Beer calls this channel the "*audit*" of the primary activities by control. To audit the primary activities, control may, for instance, track randomly selected "orders" from begin to end. In this way, control can find out how clients are actually treated, what the actual quality of order-processing is, where bottlenecks in the process are located, etc. The results of these audits are discussed with the management of the primary activities, helping it to improve the primary activities' contribution to the viable whole. Beer suggests that audits should occur at irregular intervals and should be executed with discretion. They involve detailed, in-depth examinations of the operations of the primary activities.

Controlling the Coordination Function (Control – Coordination)

The third relation is that between control and coordination. Above we stated that control focuses on the synergy and cohesion of the primary activities and that coordination is a service to the primary activities solving interdependency problems between them. In this sense, the aim of coordination lies well within the more general aim of the control function. Coordination is not only a service to the primary activities it also makes life easier for control.

The relation between control and coordination has a character quite similar to the commands and reports relation between control and the primary activities. To discharge of its task, coordination should identify domains of possible interdependence between the primary activities. To this purpose, control should communicate the goals of the primary activities to coordination. Moreover, control should set targets for coordination, for instance, to keep losses due to coordination problems below a specific level. It is also possible that reports by or audits of the management of the primary activities focus control on problems that should be the concern of coordination. Just as in the case of commands to the primary activities, commands to coordination should be aggregate and have a relatively low frequency.

Coordinating Interdependencies Between Primary Activities (Coordination – Primary Activities)

The fourth relation in the group of functions realizing the identity and mission of the organization is that between coordination and the primary activities. Above we argued that coordination is a service to the primary activities. It creates the platform to deal with problems stemming from interdependency relations. Moreover, coordination is a service to control. By dealing with problems related to interdependencies, coordination absorbs quite some complexity that otherwise would have been the concern of the control function.

To discharge of its task, the communication between coordination and the primary activities should be intense and continuous. Moreover, to deal with the variety of problems concerning interdependencies, communications between coordination and the primary activities should be detailed and flexible.

Table 6.1 summarizes the required relations between the functions in the group realizing the organization's identity and mission (columns 1 and 2). Furthermore, it lists the requirements to these relations in terms of periodicity, detail, standardization, and focus (columns 3 to 5).

Table 6.1 required relations between Group 1 functions focused on realizing the organization's identity and mission

Related functions	Relation	Periodicity	Detail	Standardization	Focus on
Control – Primary activities	Direct commands and reports and resource bargaining between control and the primary activities	Regular intervals	Low	High	Management by exception of synergy of primary activities
	Audits of the primary activities	Irregular intervals	High	-	Cross-checking the activities and reports of the management of the primary operations
Control – Coordination	Controlling the coordination function	Regular intervals	Low	-	Contribution to overall synergy by coordination
Coordination – Primary activities	Coordinating interdependencies between primary activities	Continuous	High	Low	Supporting interaction between primary activities

6.3.3.2 Relations Between Functions Required to Adapt the Organization's Identity and Mission

The second group of functions aims at adapting the organization's identity and mission. Its activities help shaping the organization's future (Espejo et al. 1996). This group consists of the functions control, intelligence, and policy. Within this group, three relations are of particular importance:

1. Consolidating plans for innovation: Policy-(Intelligence-Control).
2. Generating finalized proposals for innovation: Intelligence-Control
3. Facilitating communication between control and intelligence: Policy-(Intelligence-Control).

Consolidating Plans for Innovation (Policy – Intelligence – Control)

It is the task of the policy function to give "closure" to the process of adapting and realizing the organization's identity and mission by consolidating the plans for the organization's future resulting from the communication between intelligence and control.

Beer emphasizes that policy should not attempt to produce these plans by itself using the complex information generated by intelligence and control respectively and then force them unto the organization. Because this is such an important point, we give a somewhat longer quotation from "The Heart of Enterprise," "The autocrat supplies the closure by saying: "that's enough; I have heard all the arguments; THIS is how it will be." Then high variety generated between Systems Three and Four, and not absorbed between them, is CLOSED by System Five. [. . .]. Such an arrangement may be thought (ethically) to be immoral. Certainly (cybernetically) it is extremely vulnerable, unless the boss (Five) is God Himself. Sooner or later, five will make a destabilizing mistake. That is because his requisite variety (in administering closure) depends on a much attenuated input. It must so depend, because Five cannot deploy sufficient variety to absorb the variety of Three-plus-Four, which is multiplicative, without that enormous degree of filtration" (Beer 1995, p. 261).

Thus, it is not only ethically, but also cybernetically good practice, to keep the policy function a low variety function. Policy should not strive to fuse the information provided by intelligence and the information provided by control into plans that shape the organization's future. This task is too complex. The relevant information and its processing are so overwhelming and complex that policy cannot but deal with it by means of negation, i.e., taking some information into consideration, while neglecting other information. This may imply that policy neglects vital information, possibly leading to vital mistakes.

To prevent this information processing overload, policy must search for other ways of attenuating the information it receives from intelligence and control.

Beer argues that to do this, policy should delegate the planning of the organization's future to the high variety functions that can cope with the complexity involved in it: intelligence and control. Together, intelligence and control prepare the plans shaping the organization's future and then present them as proposals for consolidation to policy. In this way, intelligence and control not only generate, but also absorb the variety needed for adapting the organization's identity and mission.

To consolidate the plans shaping the organization's future, the relation between policy, on the one hand, and intelligence and control, on the other, should be a low variety relation. Communication should focus on essentials and only take place if considered opportune by intelligence and control. A highly interconnected intelligence and control function should do the real work. "The power to balance the Three-Four investment resides in the equation of variety between Three and Four. That can be a delegated power. [. . .]. Variety absorbs variety. All that remains for System Five is to monitor the regulatory machinery – to ensure that it does not embark on an uncontrolled oscillation. [. . .] The boss is still supplying closure – in a logical sense. He is not applying it by massive variety inhibition: he is leaving variety inhibition where it belongs, between Three and Four. His role is metasystemic" (Beer 1995, p. 261–262).

To summarize, policy should not make the plans shaping the organization's future. It should leave this complicated job to the functions equipped for it. It rather should consolidate finalized plans for innovation.

As argued above, policy has a second task that has to do with providing "closure" to the whole process of adaptation and realization. To relieve itself from making the plans for innovation (i.e., to attenuate the complexity of deciding about the organization's future), policy should monitor and facilitate the debate between the intelligence and control functions. The quality of plans for innovation depends on the quality of this debate. So, policy's second task is to ensure that this debate meets the criteria for inventing the organization's future. To see what these criteria are, we have to discuss the relation between intelligence and control.

Generating Finalized Proposals for Innovation (Intelligence – Control)

There are three basic requirements to the communication between intelligence and control about the organization's future: (1) It should be *balanced*, (2) intelligence and control should be highly *interconnected*, and (3) the debate should allow for sufficient *complexity*. By facilitating a debate meeting these criteria, policy creates the conditions for intelligence and control to make plans for innovation that are both relevant given developments in the organization's environment and realizable given the organization's potentials for innovation. Moreover, it relieves itself from the high variety task of making plans for innovation. Thus policy attenuates the complexity it otherwise would face.

(1) *Balance*. The first requirement is that the contributions of intelligence and control to the debate of the organization's future are balanced. To be able to "do the right things" and "to do these right things right," plans for the future not only should link up with or shape developments in the organization's relevant environment, they also should build on internal potentials for innovation. To generate these plans, it is necessary to balance the contributions of intelligence and control to the finalized plans for the organization's future. Above we stated that the inability to meet this requirement can lead to either "innovatism" or "conservatism."

Innovatism can occur if intelligence's contribution is stronger than control's. In this case, intelligence remains unchecked in generating plans for innovation. Moreover, control does not sufficiently scrutinize these plans on their feasibility. As a result, the organization faces an avalanche of possibly unrealizable plans. This not only implies a waste of resources, it also may undermine the willingness of the primary activities to participate in the implementation of yet another wild new plan.

Conservatism can occur if control's contribution is stronger than intelligence's. Here, intelligence is unable to defend the plans it can bring into the discussion about the organization's future against the scrutiny of control. As a result, it is possible that plans necessary for the survival of the organization are abandoned using the argument that they do not fit into its potentials for change. It is possible to prevent both innovatism and conservatism by balancing the contributions of intelligence and control.

(2) *Interconnection*. The second requirement to the debate between intelligence and control is that intelligence and control are closely interconnected. To prepare plans for innovation, intelligence and control should be involved in an intense and

continuous debate about the organization's future. In this way, it becomes possible that variety generated by intelligence and control, e.g., in the form of plans for adaptation or reviews of potentials for organizational change, is also absorbed by intelligence and control. This requirement of strong interconnection between intelligence and control to absorb the complexity generated by them also points at the third requirement.

(3) *Complexity*. This third requirement is that the debate between intelligence and control should be sufficiently complex. Planning the organization's future is a complex business. In the first place, it involves bringing into the debate perspectives on both the relevant environment (intelligence) and the organization (control). This requires the contribution of many specialists each one of them highlighting some aspect of either the environment or the organization. Communication between intelligence and control should allow for bringing in the contribution of these specialists. In the second place, the multiplicity of perspectives on both the environment and the organization also needs integration. Many-faceted-plans for innovation must first be projected on many-faceted-potentials for organizational change and then be integrated to become the balanced finalized plans that can be consolidated by policy. Both projection and integration need amplified communication between intelligence and control.

To meet these requirements, the discussion between intelligence and control needs to be regulated. As argued above, this is a task of the policy function. This brings us to the third relation between functions in the adaptation group.

Facilitating Communication Between Control And Intelligence (Policy – Intelligence – Control)

We explained that to consolidate the plans for the organization's future, policy should attenuate communication between itself and intelligence and control. To do this, it should monitor and facilitate the complex debate between intelligence and control, and reap its benefits, the finalized plans for adaptation. To this purpose it needs a general picture of the organization and its environment.

According to Beer, facilitating the debate between intelligence and control is a low variety activity. It does not focus on the complex task of generating and integrating the content of the plans for the organization's future. It rather focuses on the less complex task of creating and safeguarding conditions for the communicative processes required for making these plans.

To realize this task, policy should in the first place contribute to identifying and consolidating the main topics of the debate between intelligence and control. In this way, policy contributes to attenuating the vast domain of possible topics. From this domain a subset is chosen that opens and structures the debate. Given this creative act of attenuation, policy should balance the debate between intelligence and control. Moreover, it should monitor the tightness of the interconnection between intelligence and control. Finally, policy should monitor the capacity of the communication channels used by intelligence and control to generate and absorb the

Table 6.2 Required relations between Group 2 functions focused on adaptation

Related functions	Relation	Periodicity	Detail	Standardization	Focus on
Intelligence–Control	Generating finalized proposals for innovation	Continuous	High	Low	Balancing and integrating proposals for innovation and potentials for change into plans shaping the organization's future
Policy – Intelligence and Control	Facilitating communication between intelligence and control	Continuous	Low	Low	Balancing, interconnecting, and amplifying intelligence and control
	Consolidating proposals for innovation	Irregular intervals	Low	Low	(Re)defining the organization's identity and strategy

complexity needed for inventing the organization (for a similar list of requirements, see Espejo, et al. 1996, p. 114).

Table 6.2 summarizes the required relations between the functions in the group involved in adapting the organization's identity and mission. It also specifies requirements to their realization.

Tables 6.1 and 6.2 summarize the required relations between functions needed to ensure the viability of the organization. If we go over these functions and relations once more, a few additional remarks might be useful.

In the first place, the functions and their relations form a "closed" communicative network. This has as a consequence that complexity can "*resonate*" throughout the viable organization. For instance, problems in one of the primary activities, may keep control busy. Commands and reports are exchanged, new resource bargaining may be required, or additional audits may be summoned. All these activities may "overload" the control function. As a consequence, the quality of control's contribution to the process of inventing the organization's future may drop, endangering the viability of the organization. This, may, in turn, require policy to step in to enhance the quality of the discussion between intelligence and control. In this way, disturbances occurring in one function (in the example, in a primary activity) may "resonate" in other functions (in the example in control, intelligence, policy), in the end, affecting the whole system.

In the second place, because the functions and their relations constitute a closed communicative network, and because the functions act as they do, the whole viable system is also "built" to *attenuate* "resonating" complexity. First, each primary activity has its own potential to deal with its "own" complexity. Second, the remaining complexity is attenuated by the meta-system of the primary activities consisting of coordination, control, intelligence, and policy. Third, if we look within this meta-system, disturbances in the relation between intelligence and control are attenuated by their meta-system, the policy function. In this sense, the whole viable system can be regarded as a "complexity attenuator," maintaining the "calm" the system needs to survive (Beer 1995, p. 406). Because of closure, disturbances may resonate. However, because of closure, resonating disturbances will be attenuated, protecting the viable system from collapse.

Thirdly, the point of "systemic calm" triggered Beer to introduce one additional relation between functions within the viable system. As argued, if the functions work well, disturbances will be attenuated, introducing the calm needed for systemic viability. However, if this is the case, "Then at the level of System Five, nothing much will ever appear to happen. Therefore sleep supervenes. If we are not very careful, sleep turns into coma; and coma becomes death" (Beer 1995, p. 406).

So, according to Beer, "calm" is all right. However, too much of it can be lethal. To deal with this problem of possibly lethal calm, Beer introduces an alerting system orthogonal to the calming system. He calls this alerting system, the "algedonic" (pain and pleasure) mechanism. This algedonic mechanism is a line of communication cutting through all levels of recursion. If something is crucially wrong in a primary activity at whatever level of recursion, signals may be transmitted straight to function five at the highest level of recursion of the viable system. At that level, action may then be taken to deal with the problem.

The algedonic mechanism concludes our exposition of Beer's Viable System Model. It provides the answer to the first question posed at the beginning of this chapter; "What functions are required for organizational viability and what are the relations between them?"

Diagnosing and Designing Viable Infrastructures

Having discussed the strategies to deal with complexity and the functions needed for viability, it becomes possible to relate the Viable System Model to the experimental "arche" of organizations. More in particular, we can show that it provides the functional principles for the diagnosis and design of infrastructures enabling experiments with meaningful survival. To this purpose, we first explain in cybernetic terms what diagnosis and design entail. Second, we show that the Viable System Model provides the functional principles for diagnosing and designing organizational experiments. Third, we point at limitations of the Viable System Model.

Generally speaking, diagnosis entails (1) a "gap-analysis," establishing whether a system functions as it should, and (2) if malfunctions are found, a "cause-analysis," finding out why they occur.

For the purpose of the gap-analysis, we need norms. That is, we need essential variables and values specifying the desired behavior of the system. Given, these norms, the diagnostician inspects the actual behavior of the system and assesses whether there are malfunctions, i.e., difference between the desired and the actual behavior.

If malfunctions are found, the diagnostician needs to establish why they occur. Cause-analysis requires knowledge about the relation between the behavior of parameters and that of essential variables. Based on this knowledge, it is possible to establish the values parameters should have, if the system is to display the desired behavior. Given these desired parameter values, their actual values are established. Differences between desired and actual parameter values, then, provide an explanation of why the system malfunctions.

Designing a system entails constructing a "mechanism" that can realize the desired behavior of a concrete system. To this purpose, once again norms specifying this desired behavior are required. These norms provide the design's "functional specification." Given these norms, a "mechanism" should be constructed to both attenuate the influence of disturbing parameters as much as possible and amplify the influence of regulatory parameters as much as needed.

Applied to organizational experiments with meaningful survival this means that the aim of diagnosis is to establish whether some concrete organization produces the behavior needed to be able to survive, and in case of malfunctions, why these malfunctions occur. The aim of design is to construct an organizational infrastructure that can produce the behavior needed to be able to survive.

Apparently organizational diagnosis and design have two sets of prerequisites:

1. norms specifying desired organizational behavior, i.e., variables and desired values indicating what is needed if the organization is to be able to survive
2. a model specifying the relation between the design of the organizational infrastructure and the behavior of the essential variables.

As a functional model of viability, the Viable System Model provides the norms. It specifies desired behavior in terms of the complexity strategies and functions required for viability.

As such, the Viable System Model provides norms for both diagnosis and design. More in particular, because it provides the norms, it:

1. Can be used as a point of departure for gap-analyses, which are a part of diagnosis
2. Specifies the functional requirements, which are a prerequisite for design. It specifies what organizational infrastructures must be able to do, if the organization is to continue its experiments.

Because the Viable System Model is a recursive model, it might, in a sense, be used for the purpose of cause-analysis. For instance, gaps between desired and actual behavior at a particular level of recursion, may be "explained" by malfunctions at lower levels of recursion. Or, gaps with respect to a particular function (e.g., the malfunction of control), may be "explained" by the malfunction of other functions at the same level of recursion (e.g., the malfunction of coordination).

Note, however, that these "explanations" remain functional explanations. They only state that at either different or the same level of recursion, related difference exist between desired and actual organizational behavior. They do not explain these differences in terms of faulty infrastructures. Here, we touch on a limitation of the Viable System Model.

As a functional model, the Viable System Model specifies behavior that is wanted in terms of desired effects. It does not specify how the infrastructure should be designed in order to realize these desired effects. It does not specify a model stating the relation between the infrastructural design and desired organizational behavior.

For instance, the Viable System Model specifies that the discussion between control and intelligence should be balanced, intensive, and ongoing, but it does not specify how work should be distributed to enable such a discussion. By means of the Viable System Model, we may be able to diagnose that the interaction between control and intelligence is unbalanced, and we may be able to say that this is due to control's being too much involved in the regulation of malfunctioning primary activities. However, as a functional model, the Viable System Model does not prescribe how, for instance, work should be distributed to both enable the proper functioning of the primary activities and to support the discussion between control and intelligence. In general, it does not specify specific principles for the design of organizational infrastructures.

To conclude, the Viable System Model provides functional principles which can be used for the diagnosis and design of the experimental "arche" of organizations. As a functional model, it specifies both the norms required for the gap-analysis of organizational diagnoses and the functional requirements needed for organizational design. However, it does not provide specific design principles for (parts) of the infrastructure. In the next chapter, we go into the question of how to design distributions of work supporting organizational viability. However, before we do this, we still have to answer the second question formulated at the beginning of this chapter about the relation between the Viable System Model and the social systemic character of organizations.

6.4 The Viable System Model and Social Systems Conducting Experiments

Summarizing the content of this chapter, we can say that Beer's Viable System Model is a *recursive functional model of viable systems*, i.e., of systems that are able to survive in their environment. To explain this potential for survival, the Viable System Model provides insight into the strategies by means of which viable systems deal with the complexity they face and the functions needed for viability.

In the Viable System Model three complexity strategies are distinguished: (1) identity, (2) attenuation and amplification, and (3) recursion. The third strategy, recursion, multiplies the first two at all levels of the viable system.

The third strategy makes the Viable System Model into a *recursive* model of viability. At each level of recursion, the same strategies, functions, and relations reappear. Recursion allows designers to model organizations and the viable systems contained in them as systems that both have the autonomy to deal with their own complexity and are able to contribute to the viability of the viable systems they are contained in.

The model also specifies five functions necessary and sufficient for viability: the "primary activities" (collected in function one), "coordination," "control," "intelligence," and "policy." Each of the primary activities can be modeled as a viable system, comprising the five functions required for viability. Two groups of functions can be distinguished. One group is devoted to realizing the system's identity (primary activities, coordination, control). The other group is devoted to the system's adaptation (control, intelligence, policy). For each of the functions, the Viable System Model specifies what desired effects it should produce and how it is related to other functions if the system as a whole is to be a viable system.

This focus on desired effects makes it into a *functional model* of viability. As a functional model, it is perfectly general. It holds for all concrete embodiments of viability, for all concrete viable systems. As a functional model, the Viable System Model focuses on what the viable system should *do* to remain viable. It abstracts from the material substrate of which concrete viable systems are made, of their "flesh" and "metal" as Beer calls it. By specifying what a viable system should do, the Viable System Model provides a "rule" for recognizing, diagnosing and designing concrete viable systems.

As a functional model of viability, The Viable System Model also provides the functional design principles for organizational experiments with meaningful survival. By specifying desired effects that should be realized in order to survive, the Viable System Model provides a heuristic for the diagnosis and design of causes that can realize these desired effects. In the case of organizations, these causes are the organization's infrastructure. Therefore, as a functional model of viability, the Viable System Model specifies functional principles supporting the diagnosis and design of organizational infrastructures.

Given this summary, we can ask in this final section how Beer's functional model of viability fits within the larger project of this book: understanding organizations as a particular type of social systems conducting risky and ongoing experiments with their own survival. This amounts to the second question presented in the introduction to this chapter concerning the relation between the Viable System Model and the social systemic character of organizations.

Here, we want to rephrase this quite general question into the more specific one how the Viable System Model and social systems theory in combination can be of benefit to the design of infrastructures supporting social experiments with meaningful survival.

To deal with this question, we would like to split it into two parts. One: how can the Viable System Model help the particular type of social systems we call organizations to continue experiments with their survival? Two: why is it important to be aware that organizations are a particular type of social systems when applying the Viable System Model to/in them?

6.4.1 The contribution of the VSM to conducting experiments

How can the Viable System Model help the particular social systems we call organization to continue experiments with their survival?

If organizations are a particular type of social systems, and if they consist of decisions produced against a background of decision premises, everything that happens "in" and "to" organizations, in the end, involves decisions and decision premises.

This means that if we want to use the Viable System Model to diagnose or design organizations to enhance their potential for conducting the experiment with their survival, in the end, we need decisions and decision premises to do this. For instance, we *decide* to introduce a particular decision program for coordinating interdependencies between primary activities, or we *decide* to increase the number of communication pathways between intelligence and control to enhance the intensity of communication about the organization's future, or we use *organizational self-descriptions* for the purpose of selecting policy options.

If all applications of the Viable System Model in organizations, in the end, must involve decisions and decision premises, we can distinguish at least two contributions of the model. The first pertains to improving decisions and decision premises with regard to the *process* of adapting and realizing organizational goals, the second to improving decisions and decision premises pertaining to *the infrastructure* conditioning the adaptation and realization of organizational goals.

First, how can the Viable System Model help to improve decisions and decision premises with regard to the process of adaptation and realization of goals in organizations? Above, we argued that the Viable System Model is a functional model specifying desired effects, which can be used to search for infrastructures to realize these desired effects. Examples of the desired effects specified by the Viable System Model are: the organization should adapt its goals in the light of developments in its environment, it should scan its environment for relevant developments, the organization should control the cohesion of its primary activities, it should coordinate interdependencies between primary activities, etc.

By specifying these desired effects, the Viable System Model improves adaptation and realization processes by indicating about what kind of topics or issues decisions should be made in organizations and how these decisions should be related. In general, as a functional model specifying desired effects, the Viable System Model helps the particular social systems we call "organization" to continue experiments with their survival by prescribing a catalogue of: (1) "decision domains," and (2) relations between (decisions in) these decision domains. It prescribes which decision domains and which relations between decision domains are relevant if the organization is to be able to survive by means of adapting and realizing organizational goals.

To be more precise, by a "decision domain" we understand a particular topic/ issue about which decisions should be made if an organization is to remain viable

(Achterbergh and Vriens 2002). Examples of such topics related to, for instance, the control function, are decisions about:

- The identity of the organization
- Goals for primary activities
- Norms for admitted performance loss due to oscillations (goals for the coordination function)
- The feasibility of intelligence's proposals for innovation.

By a "relation between decision domains" we understand a dependency relation between a decision in one domain and a decision in another domain. Examples of such dependency relations, for instance, related to decisions about the feasibility of intelligence's proposals for innovation are:

- Proposals advanced by intelligence
- Estimations of the organizational changes required by these proposals
- Estimations of the impact of these changes on the primary activities
- Estimations of the required capacity for change in the primary activities
- Assessments of the available capacity for change in the primary activities
- Assessments of the gap between the required and the available capacity for change in the primary activities.

Based on the Viable System Model, it becomes possible to list decision domains and their relations and to model them in a "dependency-diagram." This diagram represents the (closed) network of function-related decision domains needed for an organization to remain viable. It can be used for diagnostic and design purposes.

The basic element of the decision dependency diagram is a tuple that ties a function for viability to a decision domain. This is denoted as <function, decision domain>. The tuple indicates that a particular function required for viability is "responsible" for decisions in a particular decision domain. In the diagram, tuples are linked by means of arrows. An arrow pointing from one tuple (T1) to another (T2) means that: decisions in the decision domain indicated by T1 are required as a premise for decisions in the decision domain indicated by T2.

Figure 6.12 is an example of a part of the decision dependency diagram. If we, for instance, focus on the tuple <policy, regulatory measures to reduce imbalances >, we can see that decisions about them, depend on quite a number of other (partly interrelated) decisions. For example, to decide about these regulatory measures, decisions about what are considered causes of actual imbalances, assessments of the success of prior applications of regulatory measures, or decisions about norms for imbalances may be needed as decision premises.

The complete decision dependency diagram provides an insight in the relation between decisions domains and the functions required for organizational viability. Moreover, it provides an insight into the dependency relations between decisions in these decision domains. It prescribes about what topics/issues decisions should be made in organizations and how these decisions should be related to each other if organizations are to conduct the adaptation and realization processes implied in experiments with their own survival. In this way, the Viable System Model

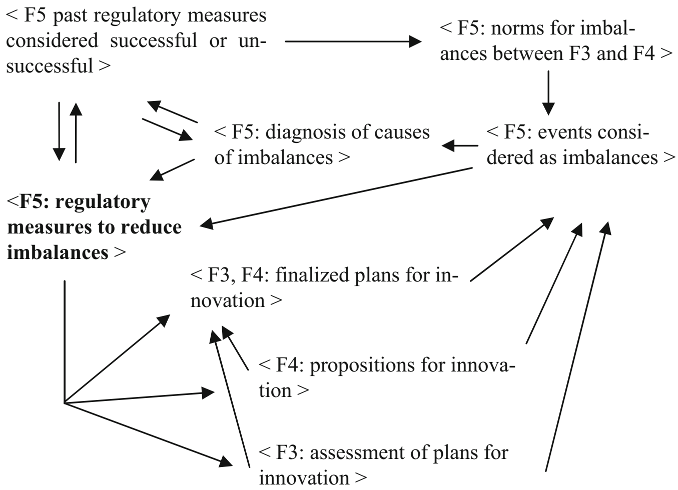


Fig. 6.12 Part of the decision dependency diagram (adapted from Achterbergh and Vriens 2002)

provides an important prescriptive heuristic for improving communication in organizations. It prescribes about what kind of topics/issues decisions should be made in organizations and how these decisions should be related if organizational experiments with their survival are to continue. As such, it can also be used as a more precise guideline, tying functional requirements to designing infrastructures.

The second question is how the Viable System Model can help to improve decisions and decision premises pertaining to the infrastructure conditioning processes of adaptation and realization of organizational goals.

Both in Chaps 1 and 2, we argued that processes of adapting and realizing goals require "mechanisms" allowing for these processes. These "mechanisms" are the object of the design activity. By means of design, it becomes possible to reduce the variety of disturbances a system has to deal with (attenuation) and/or increase the variety of its regulatory potential (amplification).

In organizations these mechanisms are their infrastructures. They consist of three important parts: the division of work, systems for human resources management, and the organization's technology. Dependent on the design of their infrastructure, organizations are more or less able to adapt and realize their goals.

Given the discussion of organizations as social systems in Chap. 4, we know that organizational infrastructures result from decisions. For instance, organizations select particular decision programs related to transformation processes, define communication pathways, define roles, select to hire particular personnel, or decide to buy particular equipment. Moreover, results of these decisions (e.g., goals, pathways, roles, personnel, machines) function as decision premises making particular follow-up decisions more or less likely. In the end, infrastructures (1) result from decisions and (2) function as decision premises.

The second contribution of the Viable System Model is that it specifies desired effects that should be produced by an organization's infrastructure if the organization is to remain viable. As such, it provides a prescriptive model specifying principles for decisions concerning the design of the organization's infrastructure. Each decision concerning design should be assessed in the light of its contribution to the realization of these desired effects.

All in all there are two ways in which the Viable System Model helps the particular social systems we call organization to continue their experiments with their survival.

First, it provides a prescriptive model of relevant topics/issues (decision domains) about which decisions should be made in organizations if they are to be able to continue their experiments with their own survival. Moreover, it prescribes how these decision domains should be related, i.e., which decisions should function as a decision premise for other decisions. Second, as a functional model, the Viable System Model provides a set of desired effects the infrastructure should be able to produce in order for the organization to remain viable. Each decision concerning the organization's infrastructure should be evaluated in terms of its contribution to these desired effects.

6.4.2 The VSM and Social Systems Theory

Why is it important to be aware that organizations are a particular type of social systems when applying the Viable System Model to/in them?

Until now, we looked at how knowledge of and experience with the Viable System Model can support "social experiments." We haven't yet addressed the question how knowledge of organizations as social systems can help the application of the Viable System Model in organizations.

To explain why such knowledge is useful, we need to take seriously that the Viable System Model is an invariant, a fully general model, holding for all instances of viability. In this sense, the Viable System Model is not about organisms, psychic systems, interaction systems, organizations, or societies *as such*; it is about *all* these systems *insofar* as their viability is concerned. However, if we want to use the model for the purpose of diagnosing or designing *actual* "mechanisms", which can realize the desired effects specified by the model, knowledge about the type of system we are diagnosing or designing is crucial, for dependent on the type of system, particular types of "mechanism" can or cannot be employed to realize these desired effects.

To give an example, think of yourself as suffering of some disease. Now you can choose between two experts. One is an expert in the Viable System Model, with only scarce knowledge about the functioning of the human organism. The other is a trained physician who does not know the Viable System Model. What both experts probably can do is look at your behavior to assess what your abilities/disabilities are. The first expert will do this in terms of the Viable System Model (combining it with a layman's knowledge of the human organism). The second in terms of variables and norm established within the medical profession (variables and norms that, although the physician per hypothesis does not know this, probably will be similar to those explicitly articulated by the Viable System Model).

However, diagnosis and design not only involve establishing *that* particular functions are badly realized, but also *why* they are badly realized, and *what* should be done to enhance the quality of their realization. To this purpose, the first expert is at a disadvantage. He or she knows (in general terms) the desired effects that should be realized, but knows little about the organic 'mechanism' that can realize them. The trained physician has got this knowledge. He or she knows the organic 'mechanism' of human beings needed to realize these desired effects.

If we want to use the Viable System Model for the diagnosis and design of concrete viable systems, we need to have knowledge about the way these concrete systems work, for otherwise we could neither spot their dysfunctions nor design system-specific infrastructures with the purpose of increasing their potential for survival.

Applied to organizations and their infrastructures this means that if we want to diagnose and design them by means of the Viable System Model, we need to have knowledge of what type of systems organizations are and what "mechanisms" can be used to realize the effects specified by the Viable System Model. So, it is important to realize that organizations are a particular type of social systems when applying the Viable System Model in them, because as a particular type of social system organizations will have system-specific "mechanisms" that can be used to design their infrastructures in accordance with the requirements set by the Viable System Model.

One way of acquiring this knowledge about system-specific "mechanisms" is by means of experiment. Given some basic understanding of organizations, a designer may select a complexity strategy or function that should be implemented to keep the organization viable. Relative to this strategy or function, the designer hypothesizes about the mix of organizational "mechanisms" that can realize it. Once this mix is installed in the organization, the designer checks whether it actually realizes the desired effects related to the selected strategy or function. If the mix fails to produce the desired effects, the designer can alter it, and check again. If the installed mix of mechanisms is successful, experience and knowledge about it can be added to the knowledge about system-specific "mechanisms" realizing a particular strategy or function. This knowledge is accepted as a "tested hypothesis" until it is refuted by experience.

Probably, all applications of the Viable System Model to organizations to some extent rely on this experimental process of hypothesis and refutation. However, if the process of hypothesis and refutation is used without taking into account prior, well established, knowledge about the system to which it is applied, it may be time consuming and risky. For instance, "installing a mix of mechanisms" in organizations and see whether they realize the desired effect required by the Viable System Model is a part of the process of hypothesis and refutation. However, "installing mechanisms" in organizations is more easily said than done. Because of the social character of organizations, it is a social process requiring communication, negotiation, and persuasion, etc. In general it involves knowledge about and skills regarding the *social* process of organizational transformation and change. Without knowledge about social coordination in organizations, "installing mechanisms" may prove to be a bridge too far, frustrating the whole process of learning to interpret, diagnose, and design organizations in terms of the Viable System Model.

Another, complementary, way of dealing with the "system-specific-knowledge" required to diagnose and design organizations by means of the Viable System Model, is to use available system-specific knowledge about organizations as a "template" onto which the Viable System Model can be projected. Here, Luhmann's theory of organizations can be helpful, for this theory focuses on the type of systems organizations are (cf. Luhmann 1995).

Two examples may illustrate that social systemic theory such as Luhmann's may be of help. The first example pertains to a problem regarding the application of the Viable System Model in organizations. The second example pertains to "category mistakes" that may be made when applying the Viable System Model to different types of systems. Both examples highlight the importance of knowing both the type of system one is dealing with and its system-specific "mechanisms" available for designing infrastructures.

Example 1: overcoming self-inhibition

One of the hardest problems to overcome when applying the Viable System Model in organizations is that organizations that most "need" an intervention based on the Model are least "open" to it. Luhmann's ideas about decisions and decision premises can be used to explain why this is so.

If we take Luhmann seriously, the Viable System Model can be interpreted as one of many models competing to become a "high-level" decision premise in organizations, affecting lots of other decision premises and decisions. From a social systemic point of view, applying the Viable System Model in an organization implies changing the set of existing decision premises. This process of change is a social process. Particularly in organizations that function according to decision premises that are at odds with the principles advocated by the Viable System Model, this social process of change may be quite difficult. One of many difficulties is the phenomenon of "self-inhibition."

If organizational decisions are structured by decision premises that are at odds with the Viable System Model (e.g., by decision premises favoring autocratic decision making or constraining the semi-autonomy of primary activities), the organization's potential for adaptation and realization of goals will be probably diminished. In Luhmann's terms, this means that decision premises are in place that decrease the organization's potential to change its decision premises. Such organizations seem unable to break out of their deadlock, because the decision premises that are so desperately in need of replacement inhibit the social transformation needed for this replacement. They inhibit the transformation process needed to lift this inhibition. Applying the Viable System Model in this type of organizations may be highly problematic, and designers seeking to apply the Viable System Model in organizations should be well aware of such self-inhibiting processes and of the social mechanisms that can be used to counter them.

In general, from a social systemic point of view, applying the Viable System Model in organizations implies changing well-established "high level" decision premises. This is an essentially social process. Social systems theory provides a conceptual framework that may be used to better understand the dynamics of social transformation and may help designers to deal with the problems involved in it.

Example 2: avoiding category mistakes

Avoiding "category mistakes" when applying the Viable System Model is a second example of the possible usefulness of social systemic theory such as Luhmann's. By making a "category mistake" we mean: "blindly" treating a particular type of system, for instance, a psychic system *as if* it were a system of another type, for instance, an organization. This may be dangerous because different system types will have different types of "mechanisms" that can be used for design purposes. When designing infrastructures to meet the requirements of the Viable System Model it is important to keep in mind the specificity of the mechanisms of the type of system that is the object of design, i.e., it is important to avoid "category mistakes."

Now, as argued above, the Viable System Model is not about organisms, psychic systems, interaction systems, organizations, or societies *as such*. It rather is a perfectly general model of *all* these systems *insofar* as they are viable. This is a good thing. However, without an awareness of the differences between system types, this general character of the Model may easily lead to "category mistakes."

For example, take a designer who wants to improve the infrastructure of an organization. To this purpose, she unfolds its primary activities down to the level of semi-autonomous teams. At this point, she asks how she should continue. For the purpose of her design, she designates semi-autonomous activities of team members as the next level of recursion. These activities can be viewed in organizational terms, i.e., in terms of decisions and decision premises (e.g., in terms of roles, decision programs, communication pathways), and related decisions. However, they can also be viewed as activities of more or less viable individual human beings that become the object of organizational redesign. Once she views them in the latter way, she makes a "category mistake."

By selecting individual human beings as the next level of recursion that is the object of her *organizational* design, she transcends the boundary of the organization as a particular type of social system, treating this other type of system as if it were a subsystem of an organization. Making such a category mistake is dangerous, for diagnosing and redesigning the infrastructure of *this* (new) type of system requires taking into account other elements and structures than the decisions and decision premises that, in the end, constitute organizations. It cannot and should not be treated as if it were an organization.

Another example of a "category mistake" concerns the application of lessons learned in organizations *to* the diagnosis and design of *other* types of (social) systems than organizations. As a functional model, the Viable System Model can be used to diagnose and design all types of systems as long as they belong to the class of viable systems.

However, this may be a dangerous thing to do. For other types of systems consist of other elements and structures (this is what makes them into *another* type of systems). Given these different elements and structures, different types of infrastructures may have to be built and different types of intervention strategies may be required to apply the Viable System Model. The danger of applying the expertise acquired in organizations to other types of systems, then, is that these other systems that are *not* organizations are treated *as if* they are. *As if* they provide the same elements and structures as organizations do to build infrastructures that realize the desired effects specified by the Viable System Model.

For instance, both organizations and the societal subsystem “economy” may be considered as viable systems. From this point of view, they fall within the same system class. However, from the point of social systems theory (see Chap. 4, Sect. 4.3), “organizations” and “societal subsystems” are quite different types of social systems, each with its own elements and structures. Given their specific differences, the two types of social systems may set different constraints to both the design and the transformation of their infrastructure. Given these differences, treating the economic subsystem as if it were an organization, may be dangerous. For instance, designers may overlook coordination mechanisms specific to the economic subsystem, overestimate its changeability, or underestimate the impact of events in other subsystems (e.g., the political or the religious subsystem) in the environment of the economic subsystem on its functioning. In the end, such “category mistakes” may impair the entire design effort.

Therefore, to decrease the probability of “category mistakes,” designers should be aware of the type of system they intervene in, of its elements and structures available for intervention. Luhmann’s social system theory provides exactly this type of knowledge because it focuses on the specificity of different types of social systems and provides a conceptual framework to describe their relevant “mechanisms.”

To sum up, applying the Viable System Model to organizations as a particular type of social systems and combining it with social systems theory can, in the first place, support organizational processes of adaptation and realization. For, applied to organizations as a particular type of social systems consisting of decisions (as their elements) produced against a changing background of decision premises (as their structures), the Viable System Model prescribes in what decision domains decisions should be made and how these decisions should be related if organizations are to remain viable.

In the second place, combining the Viable System Model and social systems theory may enable the design of organizational infrastructures needed for experiments with organizational survival. The Viable System Model *prescribes* the *effects* that should be produced by infrastructures if a system is to be viable. Social systems theory *describes* the elements and structures of organizations defining the logical space of *all possible available “mechanisms”* that may support the realization of the desired effects specified by the Viable System Model. In this sense, the Viable System Model and social systems theory "complement" each other. The Viable System Model is *prescriptive* and *general*. It prescribes what viable systems (irrespective their specific make-up) should be able to do. Social systems theory is *descriptive* and *specific*. It provides a framework for the description of all possible "mechanisms" contributing to the autopoiesis of organizations as a particular type of social systems. By combining the Viable System Model with social systems theory, we know what is desired for viability and what "mechanisms" are available to build the infrastructures to realize what is desired.

In the third place, combining the Viable System Model and social systems theory may avoid the occurrence of "category mistakes," i.e., of blindly treating one type of system as if it is another type of system. As argued, it is a strong point of the Viable System Model that it is an invariant. As such, all (viable) systems can be "interpreted" in terms of the model. However, if one is in the process of designing an infrastructure for a particular type of social system by means of the Viable System Model, the model should be used with care. Different system types have different elements and structures, allowing for different types of infrastructures and transformation processes to implement them. Blindly treating one type of system as if it were another may be dangerous, for the latter system may behave or react in quite unexpected ways. For this reason, knowledge about system specific elements, structures, and "mechanisms" may be useful. Social systems theory provides this type of knowledge.

References

Achterbergh, J., & Vriens, D. (2002). Managing viable knowledge. *Systems Research and Behavioral Science, 19*, 223–241.

Ashby, R. W. (1958). *An introduction to cybernetics*. London: Chapman and Hall.

Beer, S. (1972). *Brain of the firm*. London: Allen Lane the Penguin Press.

Beer, S. (1995). *The Heart of Enterprise*. Chichester: Wiley.

Beer, S. (1996). *Diagnosing the system for organizations*. Chichester: Wiley.

Christis, J. (2002). *Causal explanation and functional analysis in Luhmann's sociology*. Unpublished paper.

Espejo, R., Schumann, W., Schwaninger, M., & Bilello, U. (1996). *Organizational transformation and learning: A cybernetic approach to management*. Chichester: Wiley.

Luhmann, N. (1991). *Soziologische Aufklärung 1*. Opladen: Westdeutscher Verlag.

Luhmann, N. (1995). *Social Systems*. J. Bednarz, Jr. & D. Baecker (Trans.). Palo Alto, CA: Stanford University Press.

Chapter 7
Specific Design Principles: de Sitter's Organizational Structures

7.1 Introduction

In the previous chapter, we discussed Beer's Viable System Model; a functional model specifying desired effects required for viability. These effects can be used as criteria for diagnostic or design purposes. We also pointed at limitations of the Viable System Model. As a *functional* model it does not address the question of the *embodiment* of functions. Although it specifies desired effects, it does not positively address the question of how to design their realization. Simply put, the strength of the Viable System Model is stating *what* effects should be realized, not *how* they should be realized. For instance, functions three and four should engage in a relatively complex and balanced dialogue about plans for innovation, but what is needed to realize this dialogue? How should one distribute tasks and responsibilities among organizational members, so that this dialogue can be carried out properly? How should one select, allocate, and train the people involved in these dialogues, and how does one design the technological infrastructure supporting the complex communication processes required for innovation? In short: how does one design the infrastructure realizing the desired effects for viability?

In this chapter, we focus on the question how to build an infrastructure realizing the functions mentioned by Beer and proceed our inquiry into conditions for experiments with organizational survival, As we discussed in previous chapters, the infrastructure consists of three related parts – division of work, HR-systems and technology. In the present chapter, we focus on the division of work as the most basic and relevant part of the infrastructure. As a consequence, we narrow the question "which infrastructures realize the VSM-functions and increase the organization's potential to conduct experiments" down to "which divisions of work realize the VSM-functions and increase the organization's experimental potential." In this way we focus on what we have called *specific* design principles (Chap. 1), i.e., principles tied to designing a particular part of the infrastructure.

From our discussion of the VSM, we know that there is not a one-to-one relation between functions and tasks. Different functions may be concentrated into one task.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_7, © Springer-Verlag Berlin Heidelberg 2010

For instance, someone running a one-man business should perform all five functions. Moreover, many tasks can be involved in realizing the same function. For instance, realizing the intelligence function may involve someone with the task of gathering data, another with the task of analyzing it, and yet another with the task of processing the analyzed data. Given this relation between functions and tasks, it becomes relevant to ask how to distribute functions into a network of related tasks. In textbooks, we can find this question under the heading of the *design of the distribution of work*.

Everyone who has ever worked in an organization is aware of the importance of the design of the distribution of work. "Doing one's job" relates directly to this design, for the breakdown of work and its synthesis into "tasks" define what "one's job" is. A job is a part of a larger network of jobs, and dependent on the layout of this network, "doing one's job" relates to others "doing their jobs." Dependent on one's job and its place in the network, dependency relations between jobs exist that influence the efficiency and effectiveness of one's job and the quality of one's working life.

This awareness of the importance of the distribution of work is particularly alive in professional or machine bureaucracies struggling to cope with increasingly rapid developments in their environments. Yet, even in "lean," "post modern," "knowledge intensive," and "network" organizations where tasks may be diffuse and changing, workers engage in communication dependent on the (changing) distribution of work and experience its obstacles and benefits. Therefore, if organizations are to be viable, we have to take a closer look at how the distribution of work inhibits or enables social experiments.

To this purpose, we select the work of Ulbo de Sitter. De Sitter is the founder of the Modern Dutch Sociotechnical Approach, also called the Integral Organizational Design Approach. This approach explicitly and systematically uses cybernetics to formulate rules and principles for designing "viable" distributions of work. To discuss these rules and principles as well as their cybernetic origin, we choose as our point of departure de Sitter's book "Synergetisch Produceren" (1994) and his article "From complex organizations with simple jobs to simple organizations with complex jobs" (de Sitter and den Hertog, 1997).

We are well aware that de Sitter's work is almost only available in Dutch (exceptions are: de Sitter 1973; de Sitter, et al. 1990, 1997) and that, for this reason, it is not as known as that of, for instance, Thompson (1967); Galbraith (1973) or Mintzberg (1983). The reason for selecting De Sitter's work is that it is explicitly rooted in cybernetics, is far more detailed than that of other authors on organization design, and fits rather well into the overall intention of the second part of this book: formulating infrastructural conditions for organizations (as social systems) to conduct experiments.

Particularly in "Synergetisch Produceren," de Sitter specifies how a designer should distribute work to keep organizations viable. Inspired by cybernetics in general and Ashby's theory in particular, he spells out how to design distributions of work (1) attenuating disturbances and (2) amplifying regulatory potential to deal with disturbances impinging on "relevant organizational variables."

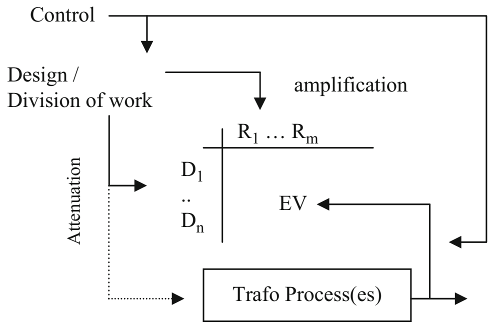


Fig. 7.1 Overview of Ashby's regulatory logic and the role of divisions of work

In fact, de Sitter's theory may be regarded as a re-specification of Ashby's regulatory logic in the realm of distributing work. As the reader may recall from Chap. 2, this logic contained the activities control, design and operational regulation with respect to (a) transformation process(es). Control meant setting the target, i.e., specifying the essential variables for the transformation process and setting their norm-values. The main goal of design was to arrive at a "mechanism" embodying an adequate relation between essential variables, disturbances and regulatory actions. Design entailed "constructing a mechanism" contributing to attenuation (decreasing the chance of disturbances) and amplification (increasing the number of regulatory actions).

Based on Fig. 7.1 we can phrase de Sitter's contribution in terms of Ashby's theory as formulating *principles for designing divisions of work* (as part of the infrastructure). In a nutshell, these principles indicate that a division of work should attenuate as much disturbances as possible and build regulatory potential (amplification) with respect to organizational processes into an infrastructure. If applied, these principles result in a part of the infrastructure (a division of work) based on which it is possible to realize control, design, operational regulation and transformation processes (see also Sect. 1.2).

To understand de Sitter's design principles, i.e., to see that and how a division of work, as part of the organizational infrastructure, can attenuate or amplify, one should translate all aspects of Ashby's regulatory logic to the case of organizations. Ashby's theory does not give this translation yet, since his theory is about regulating all kinds of systems. De Sitter, however, does. In particular, he tries to clarify (1) the relevant *organizational* essential variables, (2) the nature of disturbances in an organizational context, and (3) what regulation in the case of organizations means. Given this translation of relevant concepts to an organizational context, de Sitter systematically tries to show (4) how a division of work can attenuate disturbances and how regulatory potential can be amplified by it.

De Sitter not only gives an answer to these four questions, he also specifies principles for *arriving* at a cybernetically sound division of work (one that attenuates disturbances and amplifies regulatory potential).

In this chapter we will follow in de Sitter's footsteps and eventually hope to show how particular divisions of work (from now on we will refer to them as organizational structures) may support the organizational experiment. In Sect. 7.2, we start with discussing the desired effects of organizational structures: attenuation en amplification. In order to discuss these effects, we will discuss de Sitter's "translations" of Ashby's relevant regulatory concepts into the realm of organizations. In Sect. 7.3 we discuss which organizational structures (according to de Sitter) are "cybernetically sound structures," i.e., structures that actually attenuate and amplify. While discussing them, we present de Sitter's principles for designing organizational structures. In the last section we deal with the question of "adequate organizational structures for conducting experiments."

7.2 Designing Organizational Structures: Aiming at Attenuation and Amplification

In the previous section we stated that organizational structures should have the capacity to attenuate disturbances and amplify regulatory potential to deal with them. In fact, referring to Ashby, one could specify a regulatory table regarding these characteristics of organizational structures – see Fig. 7.2.

To understand how organizational structures can attenuate and amplify, we need to clarify:

1. Organizational structures
2. Relevant organizational variables
3. The nature of the disturbance impinging on these variables
4. Regulatory potential in the context of organizations
5. How organizations can attenuate and amplify

Each of the next sub-sections is devoted to one of the above issues.

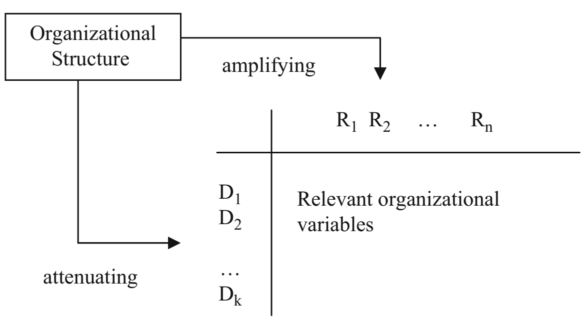


Fig. 7.2 Designing organizational structures cast in Ashby's terms (part I)

7.2.1 Organizational Structures

Roughly speaking, an organizational structure (or distribution or division of work) is a network of related tasks. To understand, in some more detail, what "a network of related tasks" is, we will discuss tasks in organizations (in particular, we will go into their definition and introduce two relevant types of organizational tasks). Second, we will discuss how tasks can be related into an organizational structure.

7.2.1.1 Tasks in Organizations I: Defining Tasks

To define tasks in organizations we use Ashby's concept of transformation. As we discussed in Chap. 2, a transformation describes a change of values of a set of variables from a begin state (the operand) to an end state (the transform) – see Table 7.1.

Table 7.1 A transformation

T↓	Begin state (operand; set of values of the variables at time = t)
	End state (transform; set of values of the variables at time = t + 1)

In the context of discussing organizational structures it is more convenient to represent a transformation slightly differently – see Fig. 7.3:

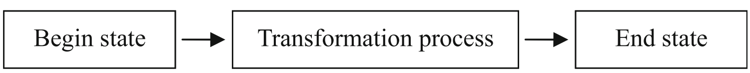


Fig. 7.3 Representation of a transformation

The representation in Fig. 7.3 highlights that there is "something" that causes the begin state to change into the end state: the transformation process. Emphasizing this process supports our explanation of tasks in organizations – as will become apparent. For now, a task is associated with "realizing the end state by performing a particular transformation process."

Transformations can be decomposed into sub-transformations. For instance, the transformation "doing the dishes" may be decomposed into four sub-transformations: "collecting"; "cleaning," "drying," and "storing." And, each of these sub-transformations can become the task of a separate individual.

There are at least two ways to decompose transformations into sub-transformations: (1) into "*parts*" and (2) into "*aspects*." In actual practice, these two types of decomposition may, and often do, occur simultaneously.

To explain both types of decomposition, take the example of Energeco, the company providing ecological energy introduced in the previous chapter. Its transformation consists of a begin state: consumers who are not provided with eco-energy (but want to be) and an end state or desired effect: consumers who are provided with eco-energy (see Fig. 7.4).

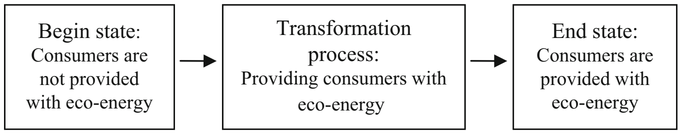


Fig. 7.4 Energeco's main transformation

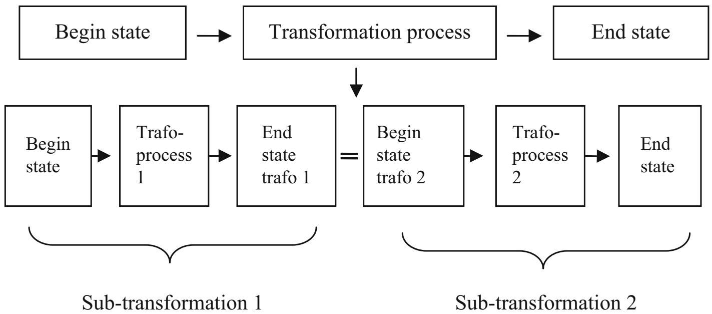


Fig. 7.5 Two sub-transformations resulting from a decomposition into parts

Decomposition into *parts* involves defining desired effects *between* the begin state and end state of the whole transformation. Fig. 7.5 shows two sub-transformations resulting from decomposition of a transformation into parts. Note, that both sub-transformations have a begin state and end state. The two sub-transformations are connected because the end-state of one transformation is the begin-state of the next one.

In the case of "Energeco," it is possible to decompose "providing eco-energy" into parts by defining three sub-transformations: "generating," "storing," and "diffusing" ecological energy. Together, these sub-transformations constitute the process of providing eco-energy to consumers.

Decomposition into *aspects* involves defining one or more characteristics (aspects) of the whole transformation and using them to define sub-transformations. For instance, it is possible to decompose "providing eco-energy to consumers" into "providing solar energy," "providing wind energy," and "providing water energy" to consumers. The characteristic, used to define sub-transformations is "type of energy provided." In this case, the decomposition does not define desired effects *between* the begin state and end state of the whole transformation, but identifies different *aspects* of the *whole* transformation (which, in this case results in three more or less parallel sub-transformations).[1]

[1] It should be remarked that the resulting sub-transformations may differ in degree of "autonomy." Sometimes, they can be treated as autonomous "flows" of production (which, as Beer (1995) suggests, may even be "hived off" from the original organization), and in other cases, one aspectual sub-transformation may be tied intimately to the other. This is the case with defining regulatory and operational sub-transformations (see explanation below).

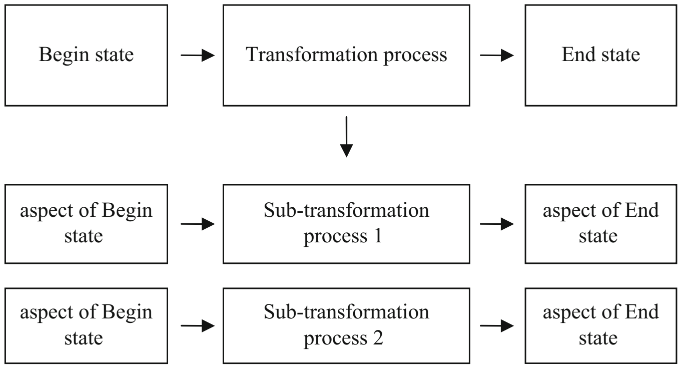


Fig. 7.6 Two sub-transformations resulting from a decomposition into aspects

Figure 7.6 shows two sub-transformations resulting from a decomposition of a transformation into aspects. Note, that the appearing sub-transformations consist of an aspect of both the original begin and end state, and a sub-transformation process connecting these two states.

Given some main organizational transformation, designers can apply all kinds of criteria to decompose it into parts or aspects. In the Energeco example, the source of the energy (sun, water, wind) functions as a criterion for a decomposition into aspects. However, it is also possible to choose other criteria, such as types of users (households, companies) or geographical areas (regions north and south) – see, for instance, Mintzberg (1983), or de Sitter (1994) about the nature of such criteria.

Both ways of decomposing transformations can be applied repeatedly and interchangeably. Every sub-transformation, resulting from decomposition can itself be decomposed into parts or aspects, resulting in new sub-transformations, and so on. Repeating this process of decomposition results, eventually, in a set of sub-transformations that can be regarded as a set of building blocks for "assembling *tasks*."

We can now define a *task* as a specific grouping of (sub-) transformations. From a set of sub-transformations resulting from different decomposition steps, many possible tasks can be assembled by set specific sub-transformations together. A criterion for defining a specific grouping of sub-transformations into a task is that such a grouping can be assigned to some "operational unit" – someone or something able to realize the set of sub-transformations. For example, the transformation "doing the dishes" can be decomposed into parts: e.g., "sorting out," "cleaning," "drying" and "storing." These four sub-transformations can be grouped into different tasks and assigned to different individuals. For instance, task 1 may comprise the sub-transformations "sorting out," "drying" and "storing" and may be assigned to one individual. In this case, task 2 (assigned to someone else) only contains the sub-transformation "cleaning."

Tasks can also be assigned to operational units other than individuals. In the case of Energeco, the decomposition into parts resulted in the sub-transformations, "generation," "storage," and "diffusion." Based on these three transformations, two tasks may be specified: one task in which generating and storing energy are grouped together and another task containing the sub-transformation diffusing energy. These tasks can be assigned to two different organizational units.

The above two examples make apparent that it is possible to specify tasks at different levels of aggregation in organizations – depending on the "operational unit" to which they are assigned. At the lowest level, we find the "job" of an individual member of the organization. This job consists of a set of transformations to be realized by that member. Jobs fit into the task of larger organizational units such as, for instance, teams, departments, or business units. Finally, the tasks of these units fit within the task of the organization as a whole. At this level the task reflects the organization's main transformation.

7.2.1.2 Tasks in Organizations II: Operational and Regulatory Transformations

In the previous section we used Ashby's notion of transformation and introduced possible ways to decompose transformations in order to define an organizational task. One may argue, at this point, that the number possible sub-transformations resulting from decomposing the organization's main transformation can be rather large, and, as a result, so is the number of possible groupings of sub-transformations into tasks. Moreover, since organizational structures are networks of related tasks, a large variety of tasks also entails a large variety of organizational structures.

To come to terms with this variety it is relevant to discuss two types of tasks: operational and regulatory tasks. Based on their description, it becomes possible to see that some groupings of sub-transformations into tasks, and some networks of related tasks are more suitable than others. According to de Sitter, the type of transformation built into a task (and hence, into the network connecting them) condition both the probability of the occurrence of disturbances and the regulatory potential for dealing with them. For instance, the potential to regulate parts of my work myself depends on whether my task comprises the required "regulatory potential." In such a case, the effectiveness of my task (with respect to dealing with disturbances) depends on the inclusion of a specific type of transformation (a "regulatory transformation") into my task.

In this section, then, we discuss two relevant types of sub-transformations, influencing the effectiveness of (related) tasks: operational and regulatory transformations. Moreover, after their introduction, we will elaborate regulatory transformations.

The Operational and Regulatory Aspect of Transformations

To understand what operational and regulatory transformations are, it is relevant to start with distinguishing an operational and a regulatory *aspect* of a transformation. The *operational* aspect of a transformation concerns the *realization* of its desired effect (its desired end state). In the case of "doing the dishes" it means performing the activities (actually sorting out, cleaning, drying and storing dishes) leading to the end state "dishes done & stored." In the case of Energeco it has to do with performing all operational activities involved in actually realizing providing eco-energy (e.g., actually generating; storing and delivering the energy). The *regulatory* aspect of a transformation concerns "dealing with the (virtual and actual) disturbances the operational sub-transformation faces in realizing the desired output of the transformation" (see de Sitter 1994, p. 10). Formulated in terms of Ashby, de Sitter's regulatory aspect consists of Ashby's control, design and operational regulation.

So, as de Sitter sees it, a transformation contains an operational aspect (realizing the end state of the transformation) and a regulatory aspect (i.e., control, design and operational regulation with respect to the operational aspect of the transformation – see Fig. 7.7).

If these aspects are used to define sub-transformations, the resulting sub-transformations are called "operational" and "regulatory" transformations. Moreover, they can be grouped together into tasks, containing more or less regulatory or operational transformations.[2]

De Sitter points at a "natural way" of thinking about operational and regulatory transformations, related to the difference between the operational "primary" processes in organizations and their regulation. De Sitter refers to the "primary" transformation of an organization – its "*raison d'être*" (for instance, providing

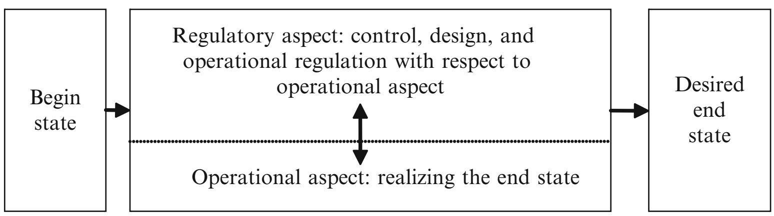


Fig.7.7 The operational and regulatory aspect of a transformation

[2]It may be noted that regulatory and operational sub-transformations are "relative" in two senses. First, they are relative to the selection of a designer modeling a particular transformation. Since operational and regulatory transformations are transformations too, they can also be decomposed regarding their operational and a regulatory aspect. Therefore, what appears to one designer as a regulatory or operational transformation may be selected by another designer as a transformation in focus and decomposed into an operational and a regulatory transformation. This entails that all transformations (including operational and regulatory sub-transformations) always contain an operational and a regulatory aspect. Second, operational and regulatory transformations are also relative to each other. Realizing an operational sub-transformation presupposes a regulatory sub-transformation, otherwise, disturbances cannot be dealt with. Moreover, realizing the regulatory sub-transformation presupposes the operational sub-transformation as its object – it deals with disturbances *of* some operational transformation.

eco-energy, in the Energeco case) as a "root" (consisting of many operational transformations), relative to which it is possible to define regulatory transformations. De Sitter illustrates this relation by means of a "procedure" to define what he calls the production structure of an organization. This "procedure" entails listing all activities, necessary to realize the primary organizational process (for instance: purchase, sales, planning, sawing wood into pieces, drilling holes into pieces of wood, assembling pieces of wood into a chair; maintaining drilling machines, etc.). These activities serve as the background for defining and coupling (operational transformations into) tasks, in which the realization of the primary organizational process is central. The resulting network of tasks can be said to cover the operational aspect of the whole organizational transformation: the *production structure*, as de Sitter calls it.

The tasks in this production structure, as well as groups of tasks have to be regulated. This means that, relative to the network of operational tasks (realizing the primary organizational process), a network of tasks dedicated to dealing with the disturbances in the production structure should be identified. De Sitter calls this network of regulatory tasks the *control structure*.

Four types of Regulatory Transformations

Above, we introduced the distinction between regulatory and operational transformations. To design a network of tasks, de Sitter argues, it is relevant to further differentiate of regulatory transformations in organizations. He holds that two distinctions are important: internal versus external regulation and routine versus non-routine regulation, resulting in four types of regulatory sub-transformations.

To explain these distinctions and the resulting four types of regulation, it is helpful to keep in mind that a task can be seen as a transformation – i.e., a set of coupled sub-transformations with an output, an operational aspect realizing this output, and a regulatory aspect dealing with the disturbances impinging on the realization of the output. Moreover, a task is related to other tasks in a network of tasks. To illustrate, while performing my (operational) task, I may have a relation with:

- A task delivering semi-finished products I have to work on, in order to produce the required output (i.e., a previous task in the operational process)
- A task to which I forward my output (the next task in the operational process)
- A task responsible for the tools I have to work with to produce the required output
- A task from which I receive a planning with regard to what I am supposed to do the coming period
- A task responsible for maintaining the tools I work with
- A task making sure that my knowledge is up-to-date
- Other tasks

Dealing with disturbances affecting the task's operational aspect can be done internally and/or externally. Dealing with a specific disturbance *internally* means that the necessary regulatory activities to deal with it are *part* of the task. Dealing with a specific disturbance *externally*, means that the task needs to involve its environment (e.g., involve other tasks in the network of tasks, or the environment of the organization).

The second distinction introduces the difference between dealing with disturbances in which the task and the network of tasks (the specific coupling of sub-transformations into related tasks) remain *unchanged* (routine regulation) and dealing with disturbances in which the task or network of tasks *changes* (non-routine regulation).

These two distinctions define four classes of regulatory activities:

1. Internal routine regulation (the regulatory activity does not involve other tasks and does not change the task or network of tasks)
2. External routine regulation (the regulatory activity involves other tasks but does not change the task itself or the network of tasks)
3. Internal non-routine regulation (the regulatory activity does not involve other tasks but changes the task's infrastructure and/or adds routine regulatory potential)
4. External non-routine regulation (the regulatory activity either changes the task's essential variables and norms or changes the infrastructure of the task's environment).

To illustrate internal and external routine regulation, imagine a worker in a factory producing wooden chairs, whose task it is to paint pieces of wood before passing them on to someone who assembles them into a chair. A disturbance for this task may be that the surface of the wooden part, as it was passed on to the worker by a previous worker (who has to saw and sand the parts), is not entirely smooth. Internal routine regulation might consist of taking a piece of sandpaper to sand the wooden part. To be an example of *internal* routine regulation it is required that this activity is part of the task. It is an example of internal *routine* regulation because no task in the network and no relation between them is changed.

An example of external routine regulation would be to go to the previous worker and ask him to sand the part again. This regulatory activity may be called *routine* if it is a part of the regulatory repertoire of the task, and if it does not involve changing the task itself nor the network of tasks. It may be called *external* because, to deal with the disturbance, the task's environment (another task in the network) is needed.

To discuss non-routine regulation, it is useful to cast what has been said thus far in terms of Ashby's regulation-table (see Fig. 7.8). The "essential task variables" in the figure refer to the output of the operational sub-transformation of the task. These variables can be subject to (virtual and actual) disturbances ($D_1 \ldots D_k$) and these disturbances may be dealt with by means of internal and external routine regulatory actions ($R_1 \ldots R_n$) that are part of the task.

In order to understand non-routine regulation, a distinction can be made between two classes of disturbances: disturbances arising "within" the task ($D_{I(nternal)}$) and

<table>
<tr><th colspan="2" rowspan="2">Task</th><th>Internal routine</th><th>External routine</th></tr>
<tr><td>$R_1 \ldots R_j$</td><td>$R_{j+1} \ldots R_n$</td></tr>
<tr><td rowspan="3">D_I</td><td>D_1</td><td colspan="2" rowspan="6">Essential Task Variables</td></tr>
<tr><td>…</td></tr>
<tr><td>D_i</td></tr>
<tr><td rowspan="3">D_E</td><td>D_{i+1}</td></tr>
<tr><td>…</td></tr>
<tr><td>D_k</td></tr>
</table>

Fig. 7.8 A regulation table for a task

disturbances arising in the environment of the task ($D_{E(xternal)}$). Disturbances internal to a task may come about because the task’s sub-transformations are poorly defined or coupled (e.g., producing the output of the task is done by means of an inefficient sequence of activities). Another cause of “task-internal” disturbances may be that the human resources tied to the task are not equipped to perform the task’s sub-transformations (because they lack, for instance, motivation or knowledge). Another type of cause for internal disturbances is that the technological infrastructure supporting the task is ineffective (e.g., because the task uses unreliable information and communication technology).

In general, then, three types of sources of internal disturbances may be discerned: (1) structure related disturbances (concerning the way the sub-transformations of the task are defined and coupled); (2) disturbances related to the task’s human resources, and, (3) disturbances due the technological infrastructure of the task. As we already discussed in the chapter on Ashby – a “designer” designs a “mechanism” embodying the regulation table using and combining structural measures, and measures pertaining to human resources and technology, and a specific mix of measures from these three classes of measures was called the infrastructure. If we apply this to this chapter, we find that the task’s infrastructure comprises its structure, human resources and technology. And, in this section we define “internal” disturbances as caused by the infrastructure of the task.

External disturbances originate in the environment of the task – somewhere else in the organizational network (in other tasks or in relations with them) or even in the environment of the organizational network (e.g., because suppliers do not deliver what they are supposed to deliver, or because of a change in demand).

Given these two classes of disturbances, the difference between internal and external non-routine regulation can be explained.

Internal non-routine regulation has to do with “redesigning” the task. It either amplifies the (internal or external) routine regulatory repertoire of the task so that it can henceforward deal routinely with both classes of disturbances, or it attenuates the internal disturbances by redesigning the task’s infrastructure. In this sense, amplification always entails changing the task’s (regulatory) sub-transformations – for it enlarges the routine regulatory potential. Attenuation (i.e., decreasing task-internal disturbances) comes about by changing the infrastructure of the task. The logic behind this way of non-routine regulation is that (parts of) the task’s infrastructure may be seen as a cause of disturbances and that by redesigning (parts of) the infrastructure these disturbances may be removed. Redesigning the infrastructure may include actions related to all three of its parts: e.g., coming up with a different way of performing or coupling the task's operational sub-transformations; allocating different workers to a task, train or motivate them; or change the technology used.

External non-routine regulation deals with disturbances by changing the environment in two ways. It can either change the task's essential variables and /or their norms (thus changing the relation with the environment, and hence, as de Sitter puts it, "the environment's expectations") or it can attenuate disturbances originating in the environment. Attenuation of environmental disturbances can take effect by redesigning the network of tasks (serving as an environment of the task in question). In the example of the two related tasks in the wooden-object factory, given above, external disturbances for the "painting" task arose because of the faulty output of the "sawing and sanding" task. Attenuating disturbances for the painting task by external non-routine regulation now may include changing the infrastructure of the "sawing and sanding" task in such a way that the output of that task is no longer outside acceptable limits. Attenuation by external non-routine regulation may stretch further than the organizational network of tasks and may also come about by changing the relations with the environment of this network (e.g., contracting more reliable suppliers) or even changing the infrastructure of parts of the environment of the organization (e.g., redesign the value chain of which the organization is a part).

To summarize, a task may contain four types of regulation so that it can deal with the (virtual or actual) disturbances influencing its operational sub-transformation – depicted in Fig. 7.9.

This figure may also serve as an anchor for relating the types of regulation of de Sitter to Ashby's regulatory notions control, design and operational regulation (see Chap. 2) – which is given in the following Table 7.2.

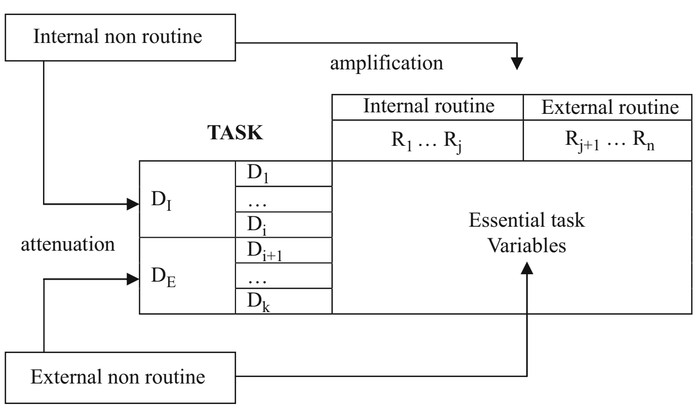


Fig. 7.9 Four types of regulation of a task

Table 7.2 Types of regulation according to Ashby and de Sitter. In the last column an alternative formulation of the types of regulation (also used by de Sitter) is given

Ashby	De Sitter	
Control	External non-routine regulation by changing essential task variables and/or their norms	Strategic regulation
Design	External non-routine regulation: attenuating task-external disturbances by redesigning network of tasks or change environment otherwise;	Regulation by design
	Internal non-routine regulation: attenuating task-internal disturbances by redesigning the tasks internal infrastructure;	
	Internal non-routine regulation by amplification, i.e. increasing routine regulatory potential	
Operational Regulation	Both internal/external routine regulation	Operational regulation

7.2.1.3 Relating Tasks into Organizational Structures

Given a set of operational and regulatory sub-transformations a designer should compose tasks and relate them in such a way that an "adequate configuration of tasks" emerges by means of which the organization's "primary transformation" can be realized. This configuration should be designed in such a way that it attenuates disturbances as much as possible, and, at the same time, amplifies the regulatory potential to deal with the remaining disturbances. In other words, a designer should design "cybernetically sound organizational structures."

A configuration of tasks (or, the *grouping* of sub-transformations into tasks and the *coupling* of tasks resulting from this grouping) is defined by de Sitter as the "organizational structure." To be precise, an *organizational structure* can be defined as: *the grouping and coupling of transformations into tasks and the resulting relations between these tasks relative to orders*.[3] In this definition, an order refers to a request for the realization of some specific desired effect (e.g., a product or service).

De Sitter (1994, p. 91 ff.) also distinguishes two sub-structures of organizational structures: a production structure and a control structure (see also previous section):

1. The *production structure* refers to the grouping and coupling of *operational* transformations into tasks and their relation to orders.

[3]This description is derived from several definitions of de Sitter (pp. 93, 100 and 101). Moreover, de Sitter sometimes uses the term architecture in his definition, to refer to the specific grouping and coupling of transformations (e.g. 1994, p. 93, 100). This seems to be redundant and even de Sitter himself sometimes omits it. Therefore, we do not follow de Sitter in using this term.

2. The *control structure* refers to the grouping and coupling of *regulatory* transformations into tasks and their relation to the production structure

The production structure highlights organizational tasks and their relations insofar as these tasks comprise operational transformations. It consists of the operational aspect of the organizational structure – i.e., the aspect realizing orders. The control structure highlights organizational tasks and their relations insofar as these tasks comprise regulatory transformations. It consists of the regulatory aspect of the organizational structure. The organizational structure is the combination of the production and the control structure. Given these two sub-structures, designing an adequate organizational structure now means designing an adequate (1) production structure relative to orders and (2) control structure relative to the production structure.

7.2.2 Relevant Organizational Variables

The previous sub-section clarified de Sitter's ideas about organizational structures. This was only the first step in understanding how these structures can attenuate disturbances and amplify potentials to deal with them. As indicated, a second step is to make clear what, exactly, the "relevant organizational variables" are, regarding to which disturbances are attenuated and regulatory potential is amplified

In de Sitter's theory, these "relevant organizational variables" are called "functional requirements." De Sitter distinguishes *external* functional requirements – requirements set by modern business environments that should be met, in order to secure the organization's viability. The external requirements can be translated into *internal* functional requirements. These internal requirements should be met by an organization in order to meet external requirements.

De Sitter specifies functional requirements into three categories: the *quality of organization*, the *quality of working relations*, and the *quality of work* (De Sitter 1994, p. 41–42).

By the *quality of organization*, de Sitter refers to an organization's potential to effectively and efficiently realize and adapt its goals. According to de Sitter, the category "quality of organization" consists of three external functional requirements: order flexibility, control over order realization, and potential for innovation.

By the *quality of work*, De Sitter refers to the meaningfulness of jobs and (the possibility to deal with) work related stress. Accordingly, he gives two external functional requirements in this category: a low level of absenteeism and a low level of personnel turnover.[4]

[4]In fact, he adds a third external requirement with respect to the quality of work – one he labels "the need to balance "qualitative demand for work and social and economic developments," by which he means that (1) jobs should fit the demands from the labor market and (2) jobs should help in increasing one's chances on the labor market. We will not elaborate this requirement.

Table 7.3 External and internal functional requirements (adapted from De Sitter 1994, p. 42)

External functional requirements		Internal functional requirements
Quality of organization	order flexibility	Short production-cycle time Sufficient product variations Variable mix of products
	control over order realization	Reliable production and production time Effective control of quality
	potential for innovation	Strategic product development Short innovation time
Quality of work	low levels of absenteeism	Controllable stress-conditions;
	low levels of personnel turnover	Opportunities to (1) be involved, (2) learn, and (3) develop
Quality of working relations	Effective communication	Shared responsibility participation in communication

By the *quality of working relations*, he refers to the effectiveness of communication in organizations.

Table 7.3 lists these three classes of functional requirements. De Sitter argues that modern organizations should strive to reach appropriate levels of all these external requirements. In fact, he states that if they do not, their viability will be threatened.

All external requirements can be translated into one or more "internal requirements." If organizations can realize these internal requirements, the external requirements will be satisfied as well. For example, the external requirement "order flexibility" refers to an organization's potential to (1) be able to deal with fluctuations in demand and to (2) deliver the required demand in time. According to de Sitter, order flexibility can be translated into three internal requirements: short production time, sufficient product variations and a variable mix of products. This means that flexibility (as an external requirement) can, in principle, be satisfied by (1) decreasing production times, (2) producing products in a sufficient number of variations, and (3) producing a sufficient number of different types of products. The last column of Table 7.3 lists the internal requirements, needed to satisfy the external ones.

The list of external and internal functional requirements comprises de Sitter's set of essential variables. If the internal requirements are met, the external requirements are met and the organization's viability is secured. De Sitter further argues that it is the task of a designer to design an organizational structure that aids in meeting all these requirements *at the same time*. Therefore, the overall adequacy of an organizational structure should be evaluated in terms of its capacity to contribute to satisfying all internal, and hence external, requirements.

7.2.3 Disturbances

In order to realize the functional requirements, de Sitter holds that organizational structures should decrease the number of disturbances affecting the three classes of essential organizational variables (attenuation), and increase the potential to deal with the remaining disturbances (amplification). To understand what this means, it is necessary to be more precise about the nature of these disturbances. What, in de Sitter's theory counts as a disturbance regarding the three classes of variables?

In Chap. 2 a disturbance was described as some event (potentially) causing the essential variables to change value. This description of a disturbance is also valid in de Sitter's theory about organizational structures. However, in his theory a distinction is made between the occurrence of some disturbing event and the probability of the occurrence of such an event. De Sitter proposes to evaluate organizational structures with regard to their capacity to decrease the probability of the occurrence of disturbances. He also introduces the concept of "sensitivity to the dispersion of disturbances" – that is, the degree to which a disturbance occurring somewhere in the network of tasks affects other tasks. For instance, the sensitivity to the dispersions of disturbances in a distribution of work in which all tasks are coupled serially, is high, because a disturbance occurring in the first task affects all others.

Given these distinctions, it is, in principle, possible to draw up a table with information about disturbances for some distribution of work – see Table 7.4.

The first column lists all possible disturbing events. The second column lists the probability of their (independent) occurrence, and the last column shows the number of tasks affected by the disturbance divided by the total number of tasks (as a measure of the sensitivity (of the organizational structure) to the dispersion of the disturbance).

This view on disturbances enables a re-specification of attenuation and amplification in the context of organizational structures. A distinction can now be made between:

1. Attenuation – i.e., the degree to which an organizational structure decreases
   a. The probability of the occurrence of disturbances (this incorporates removing disturbances, in which case the probability becomes 0).
   b. The sensitivity to the dispersion of disturbances (that is, the proportion of affected tasks is decreased)

Table 7.4 Disturbances: events, probability and dispersion

Disturbing event	P(disturbing event)	Proportion of affected tasks
D_1	$P(D_1) = \ldots$	# tasks affected by D_1/total # tasks
D_2	$P(D_2) = \ldots$	...
...	...	
D_n	$P(D_n) = \ldots$	

2. Amplification – i.e., the regulatory potential built into an organizational structure to actively deal with virtual or actual disturbances. As we discussed earlier, this can be realized by four types of regulation: internal routine, external routine, internal non-routine and external non-routine regulation – at the level of a task, a part of the network of tasks, or at the level of the network of related tasks itself.

De Sitter (1994, pp. 23–26) also gives a specification of structure-related causes of disturbances. He argues that the probability of disturbances for a particular task depends on four different things: (1) the number of relations a task has with its environment, (2) the variability of these relations, (3) the nature of environmental changes, and (4) the specificity of the norms regarding the output and the way the task should be carried out.

First, the number of relations a task has with its environment (with other tasks in the network or with the environment outside the network) can be a reason for an increased probability of disturbances. Every relation introduces a possible source of disturbances. So, the higher the number of relations, the higher the probability a disturbance will occur.

A second reason for disturbances to occur is the variability of these relations. By this, de Sitter refers to the degree to which the content of the relation varies. As de Sitter argues (1994, p.26): The chance that something goes wrong regarding a specific relation, increases if the number of different messages, instructions, requests, material, etc., associated with that relation (i.e. its *variability*), increases. This is true for two reasons. First, because every different form of "relational content" (message, instruction, request, material, etc.) can itself be a source of disturbances. Second, a high level of variability also leads to an increased probability of disturbances because one has to deal with all these different contents in one task.

As a simple example, consider the probability of something going wrong in the task of a cook, who gets orders for food from one waiter. There is only one relation between two tasks. The variability of this relation (i.e. the number of messages passed on from waiter to cook) depends on the number of items on the menu. Clearly, the probability of a disturbance is high if the number of menu-items is high because every item/message can itself be a source of confusion, and because the cook has to deal in his task with all the different messages (prepare the requested items). In general, the higher the variability, the higher the probability of the occurrence of disturbances. And, consequently, the higher the number of relations with high variability, the higher the probability of disturbances.

De Sitter sees the nature of the change of the environment of a task as a third reason for disturbances. This change can be more or less stable, or predictable. It goes without saying that the probability of a disturbance is higher if the environmental change, causing the disturbance, is harder to predict.

The last reason for task-disturbances, as treated by de Sitter, is the specificity of the norms regarding the output or regarding the way the task should be carried out. The more specific the norms regarding the output the less freedom one has to deal with output-variations by means of external (routine) regulation. In such cases this output-variation cannot be absorbed by the network of tasks, but, instead, becomes an error in it. Also, if the way a task should be performed is rigidly specified, one removes, of course, the potential for dealing with (disturbing) situations that are not covered by the task-specifications.

7.2.4 Attenuation and Amplification by Organizational Structures

Given the above treatment of organizational structures, essential organizational variables, and disturbances in organizations, we can now articulate de Sitter's argumentation with regard to designing structures in some more detail. According to de Sitter, an organizational structure should attenuate (decrease the probability of occurrence of disturbances and decrease the sensitivity to the dispersion of disturbances) and amplify (deal with occurring disturbances – by means of the four types of regulation) regarding three essential organizational variables – quality of organization, quality of work and quality of working relations.[5] The regulatory table from Fig. 7.1 can now be redrawn as in Fig. 7.10.

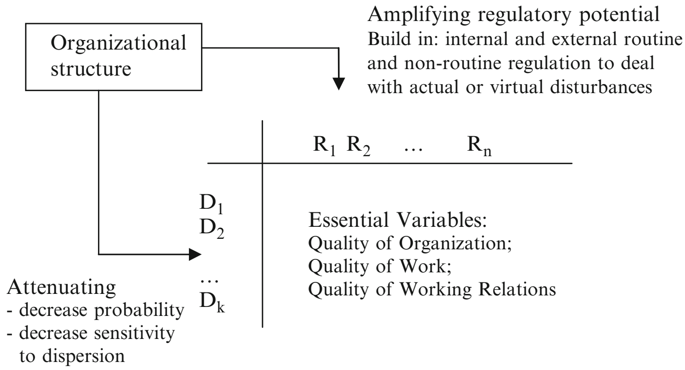


Fig.7.10 Designing organizational structures cast in Ashby's terms (part II)

De Sitter's own term for indicating that organizational structures should both attenuate and amplify regarding disturbances is "*controllability*." According to de Sitter, controllability is the overall design principle for organizational structures (1994, p. 206 ff.). In its general form, this principle is a reformulation of Ashby's cybernetic insights regarding the relation between (the variety of) disturbances and

[5]From this reasoning, it follows that designing an organizational structure means building into the organizational structure the (regulatory) capacity to change the organizational structure itself (in fact, to change all parts of the infrastructure) – compare to Chapter 1. This is so because we defined regulation by design (1) as a way to deal with disturbances – i.e. as a part of the organization's regulatory potential, and (2) in such a way that it included changing the transformation's (task's) infrastructure. In this sense it could be said that one designs for self-design.

regulatory potential. De Sitter (1994, p. 207) formulates controllability as the ratio of the potential for regulation and the required regulation[6] (see formula 7.1).

$$\text{Controllability} = \frac{\text{Potential for regulation}}{\text{Required regulation}} \tag{7.1}$$

If treated as a ratio, optimizing controllability can be achieved by both decreasing the required regulation (attenuation) and increasing the potential for regulation (amplification). As de Sitter explains (1994, p. 204/205) controllability expresses both effectiveness (if the potential for regulation matches the required regulation – i.e., for every situation where regulation is required, the potential for regulation is available) and efficiency (if the potential for regulation does not exceed the potential needed to deal with the required regulation which should be as low as possible – see also footnote 6).

7.3 Principles for Designing Organizational Structures

The goal of the previous section was to describe the desired effect of organizational structures: attenuation and amplification. This desired effect is expressed by what de Sitter calls "controllability." However, although we now have an idea of the aim of organizational structures, it has not yet become apparent how structures that *actually* attenuate and amplify should be designed. The present section will address this question. In order to do so, we start (in 7.3.1) with describing de Sitter's "*design parameters*." These parameters capture relevant characteristics of organizational structures that, according to de Sitter need to have specific values, so that organizational structures are able to attenuate and amplify. Once we know what these design parameters are we can discuss which values they should be given by designers to optimize organizational structures. In fact, de Sitter's specific design principles consist of stating which design parameters should have which value. To discuss these design principles, we introduce (in 7.3.2) two contrasting organizational structures, one with design parameters having "high values" ("high parameter

[6] Actually, he formulates it as a *function* of this ratio. The reason for this has to do with the interpretation of the ratio as a real quotient. To be able to do so, we might define the "potential for regulation" as "the number of situations, calling for regulation that can actually be regulated" and the "required regulation" as "the number of situations calling for regulation." In this case, given a certain "number of situations calling for regulation," a "controllability-value" of 1 expresses maximum effectiveness. However, it might be that this number can still be reduced. To achieve an optimal value of controllability, then, it is needed to decrease the number of "situations calling for regulation" and, next, add (or remove) enough regulatory potential to attain the value to 1. This, one may recall, is exactly how (according to Ashby) any design-process should proceed. Only if "the number of situations calling for action" can not be lowered, one may say that 1 is the optimal value for controllability. The *function* should bring this effect about.

structures") and one with "low values ("low-parameter structures"). Moreover, we will argue that low parameter structures are, according to de Sitter, better equipped to attenuate disturbances and amplify regulatory potential.

7.3.1 Design Parameters

As indicated, design parameters capture characteristics of organizational structures, relevant for attenuation and amplification. If one knows what their optimal values are (with respect to attenuation and amplification) they can be used for diagnosing and designing organizational structures. De Sitter discusses seven design parameters describing organizational structures. Every configuration of parameter-values has a specific effect on the structure's controllability – i.e., on attenuation and amplification. Designers can use this knowledge to assess and evaluate existing and proposed structures and thus try to arrive at adequate ones. Below, we discuss these parameters.[7]

De Sitter's parameters can be divided into three groups. One group of parameters describe the *production structure* (the grouping and coupling of operational transformations and their relation to orders) A second group describe the control structure (the grouping of regulatory transformations into tasks) and one parameter describes the "separation" between the operational and the regulatory transformations (resulting in a separation of tasks with respect to the production and to the control structure).

---

[7] At this point the "recursive use" of the parameter-concept as introduced in the chapter on Ashby should be noted. A parameter was described as something having an influence on the behavior of a system – or more specifically, as variable which values can influence the essential variables. With parameters, it became possible to describe the influence of "disturbances" and "regulatory actions" on essential variables. In the present context of organizational design, two "sets" of parameters and essential variables are identified. The first set has as essential variables the three "quality-variables" and to make sure that adequate values of these variables are attained, parameters such as "the organizational structure," "human resources" and "technology" may be identified. In this chapter on designing organizational structures, we focus on one of these parameters – the "organizational structure." The second set has this parameter, the "organizational structure" as its point of departure. The question becomes: how can we design an adequate organizational structure, and, hence, new essential variables (related to the adequacy of organizational structures – in terms of its capacity to attenuate disturbances and amplify regulatory potential) are defined and new parameters are identified. These parameters (we discuss them below) are called design-parameters and giving them specific values has the purpose of reaching an adequate organizational structure – or put differently – of meeting the essential variables expressing the adequacy of the organizational structure.

7.3.1.1 Parameters describing the production structure

Parameter 1: the level of functional concentration

De Sitter describes functional concentration as a parameter, referring to "the grouping of operations [operational tasks] with respect to orders" (p. 98). He goes on to describe the two endpoints of what can be called a "functional-concentration scale." A maximum value of functional concentration means that all operational tasks *of the same type* are concentrated into specialized departments, where they are performed with respect to (potentially) all orders. A minimum value of functional concentration means that all operational tasks (*of a different type*) required for realizing some order are grouped together into a "production-flow" (p.98). To understand this parameter it should be clear what operational tasks "of the same type" and "of a different type" amount to. To explain this, again consider the case of a factory producing tables and chairs. In this case, the operational sub-transformation comprises all the operational activities needed to realize orders (e.g., sales, planning, maintenance, sawing, drilling, assembling parts, and distributing tables and chairs). These activities may serve as a starting point to further decompose the operational sub-transformation. One type of decomposition specializes them into sub-transformations that are different – with respect to their output as well as to their transformation process. These sub-transformations are said to be different in type, and if one assigns these sub-transformations to tasks, these tasks can also be said to be different in type. So, for instance, the task covering the sub-transformation "sales," the task covering the sub-transformation "planning," etc. differ in type. It is also possible to design several tasks covering the same sub-transformation – e.g., assign the sub-transformation "sawing" to 15 different workers. In this case, the tasks all have a comparable transformation process (and output) and are said to be of the same type. Based on these considerations, we can now understand de Sitter's "level of functional concentration" (cf. de Sitter 1994, p.43). If it has a maximum value, all operational tasks of the same type are (potentially) coupled to *all* order-types. Moreover, in practice, they are concentrated in specialized (functional) departments (e.g., in a "sales-department," a "planning-department," etc.). In a structure with maximum functional concentration, workers perform tasks related to all possible order-types. Fig. 7.11 presents a structure in which the overall operational sub-transformation has functionally specialized departments, coupled to both types of orders.[8] If the level of functional concentration is minimal, all tasks, necessary for the production of some order-type,

[8]The figure should actually be drawn as a decomposition into parts, i.e. the departments should be "separate" sub-transformations producing output that is the input for the next department – see also section 7.2.1.1.

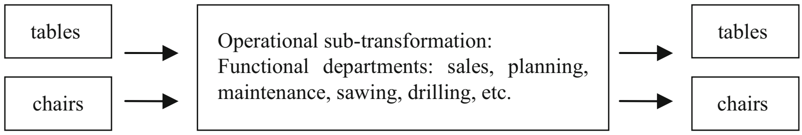


Fig.7.11 Functional concentration: all operational sub-transformations are (potentially) coupled to all (two) order-types

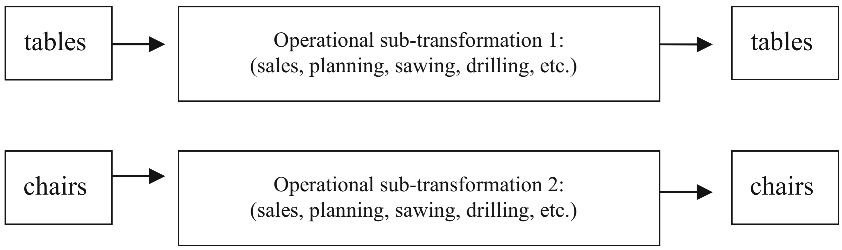


Fig. 7.12 Functional de-concentration: all order-types have their own specific set of operational sub-transformations

are only coupled to this specific order-type – and grouped together into a production-flow. In a structure with minimal functional concentration, so-called (more or less autonomous) "parallel flows" coupled to types of orders may appear.[9] In such a structure, a worker only performs a task related to one order-type – not for others. In Fig. 7.12, a structure with two flows is presented.

Parameter 2: the level of differentiation of operational transformations

The second parameter to describe the production structure is called the level of differentiation of operational transformations. In order to understand this parameter, it is necessary to see that, according to de Sitter (1994) three types of operational sub-transformations can be differentiated: "making," "preparing," and "supporting." Making refers to the actual, direct realization of the output of the transformation (or in the case of organizations: actually producing an order) – for instance, performing the sequence of activities, needed to make a table. Preparation refers to providing the necessary conditions for performing the sequence of "make"-operations (e.g.: scheduling "workers"; providing the necessary raw

[9] It should be noted that many criteria can be used to distinguish "types" of orders, e.g. their physical similarity (tables, chairs), type of customer (e.g. industry, or retail), etc., see also section 7.2.1.1.

materials, tools and equipment for producing a table). Both making and preparing are directly tied to the transformation's specific output (orders in the case of organizations): each time a specific table is made for a *specific* customer, and for this *specific* order, preparations are made. Typical preparation activities in organizations are purchase, sales or planning. Supporting refers to all operational activities that are indirectly tied to realizing the output (producing orders) such as maintenance, human resources planning, or technical service. The level of differentiation of operational transformations is "maximal" if operational sub-transformations are grouped into "make," "prepare," and "support" tasks, and minimal if operational tasks contain make, prepare and support sub-transformations.

Parameter 3: level of specialization of operational transformations

The level of specialization of operational transformations refers to "how much tasks are split up into short (-cycled) sub-tasks." Specialization of operational transformations increases as operational transformations become more specialized and these specialized transformations become separate tasks. For instance, the operational transformation "doing the dishes" can be specialized, i.e., decomposed into the smaller sub-transformations "sorting out," "cleaning," "drying," and "storing" dishes and each of these transformations can become a separate task, realized by an individual worker. Specialization decreases as specialized sub-transformations of a transformation are integrated, and become one task. For instance, "doing the dishes" may be treated as one separate task. It should be noted that specialization may concern make, support as well as preparatory transformations.

Parameter 4: the level of separation between operational and regulatory transformations

Separation between operational and regulatory transformations in tasks is maximal if (1) operational transformations are grouped into tasks that are maximally stripped from their regulatory potential (i.e., they contain very few regulatory sub-transformations) and (2) regulatory transformations are grouped into tasks separated from its operational aspect (as much as possible). In this case, "operational tasks" depend for regulation on separate "regulatory tasks." Separation is minimal if a task consists of both operational sub-transformations *and* the regulatory sub-transformations needed to regulate them. Much separation leads to a separation of the control and production structure – that is, two networks of tasks exist: one devoted to producing orders and a second one, devoted to regulating the first. Little separation leads to one network of tasks, comprising both operational and regulatory sub-transformations.

7.3.1.2 Parameters describing the control structure

Parameter 5: the level of differentiation of regulatory transformations into aspects

The level of differentiation is maximal if the three Ashby-related types of regulation (control or strategic regulation; regulation by design, and operational regulation) are grouped into different tasks. An example of this is an organization where the board deals with strategic regulation, a staff of engineers takes care of regulation by design, and workshop managers take care of operational regulation. The level of differentiation of regulatory tasks is minimal if these three forms of regulation are combined into one task. For instance, J. Jones (a butcher) who makes his own strategic decisions, redesigns his own shop, and regulates the servicing of his clientele to his own standards.

Parameter 6: the level of differentiation of regulatory transformations into parts

With this parameter, De Sitter focuses on the particular decomposition of every regulatory transformation into a "monitoring," "assessing," and "acting" part. According to de Sitter (1994, p.94), every regulatory activity necessarily involves three activities: monitoring, assessing and acting. That is, to deal with the disturbances regarding the operational aspect of a transformation, it is necessary to measure the actual values of variables defining the operational sub-transformation (monitoring). In addition, these values have to be compared with a norm (assessment). If there is a difference between the actual value of one or more variables and the norms, actions reducing the difference between the actual and the norm values of the variables should be selected and performed (acting). The effect of the regulatory actions should be that the relevant variables regarding the operational sub-transformation return within their limits. Because the operational sub-transformation may be interrupted by disturbances at any time, the three regulatory activities should also be conducted continuously. The value of this parameter is maximal if one differentiates regulation into "monitoring," "assessing," and "acting" and assigns these sub-transformations to separate regulatory tasks. The value of this parameter is minimal if these regulatory aspects are integrated into one task.

Parameter 7: the level of specialization of regulatory transformations

As with specialization of the operational transformation, this parameter refers to the level of "splitting up" regulatory transformations into small sub-transformations. The value of this parameter increases as the decomposition of a particular regulatory transformation increases, and as these regulatory sub-transformations become separate tasks. For instance, operational regulation may be decomposed into several regulatory sub-transformations: product quality, efficiency, personnel, etc., and each of these specialized transformations may become a separate task.

Table 7.5 De Sitter's design parameters

Parameter	Parameter type
1. Level of functional concentration (relative to orders)	Parameters to describe production structure
2. Level of differentiation of operational transformations in tasks	
3. Level of specialization of operational transformations in tasks	
4. Level of separation of operational and regulatory transformations in tasks	Parameter to describe the separation between the production and control structure
5. Level of differentiation into aspects of regulatory transformations in tasks	Parameters to describe the control structure
6. Level of differentiation into parts of regulatory transformations in tasks	
7. Level of specialization of regulatory transformation in tasks	

Specialization decreases as sub-transformations of a regulatory transformation are integrated into one task.

Table 7.5 summarizes de Sitter's seven design-parameters for describing and assessing organizational structures. In order to describe and assess an organizational structure, each parameter might be given a "value".[10] This value lies somewhere on the parameter-dimension for which we discussed the minimum and maximum end points. The configuration of seven parameter-values may be taken as a description of the whole organizational structure. In Fig. 7.13 a graph is drawn, tentatively picturing the configuration of parameter-values describing an organizational structure.

De Sitter's contention is that different organizational structures have a different effect on the organization's controllability. More specifically, he holds that these effects can be assessed by means of the configurations of parameter-values. Moreover, an assessment of organizational structures based on configurations of parameter-values can, in turn, be used to (re-) design them. The question, then, becomes: what the effect of specific configurations of parameter-values on controllability is (i.e., on attenuating disturbances and amplifying regulatory potential). Based on this knowledge one can treat these parameters as *design*-parameters – that is, one can select specific values of the parameters and use them as principles, as guidelines to compose an adequate network of related tasks. What designers need to

[10] To hold that these parameters can be given "exact values" on some linear scale would be stretching things. In the text we defined the dimensions on which they can be "scored" and gave their end points. Given these dimensions and end points, a relative (qualitative) valuation may be possible: relative to the end points (e.g. "the value for this parameter lies closer to a maximum value than to a minimum one") and to other parameter-values (e.g. "the value for this parameter seems to be higher than others, given a particular problem in the organizational structure"). A more adequate description of the value of a parameter might be "too high, too low or ok regarding the attenuation or amplification of organizational disturbances". Moreover, the dimensions are not entirely independent. Nevertheless, designers should somehow refer to an (implicit) version of such a parameter-value graph in their design efforts.

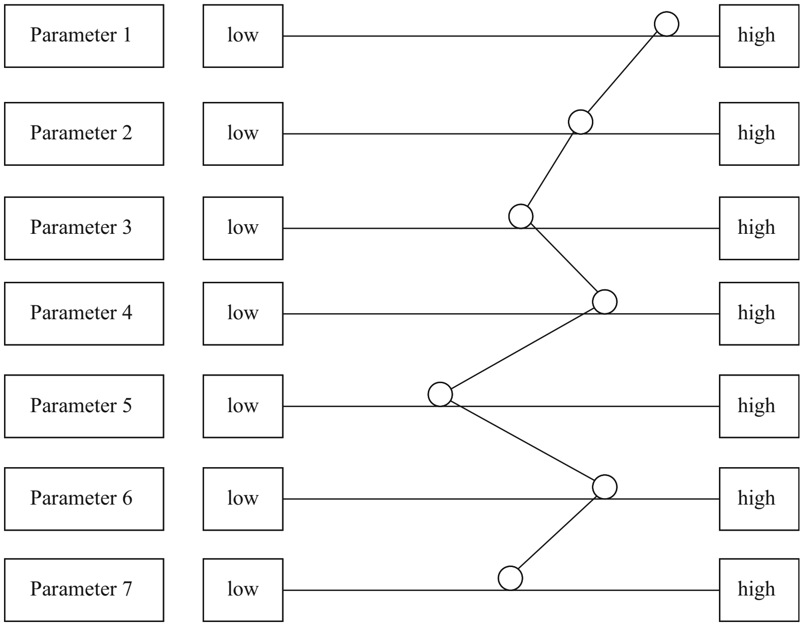


Fig.7.13 A set of parameter-values describing an organizational structure

know, then, is the nature of the impact of different parameter-value configurations – the topic of the next section.

7.3.2 Using Design parameters to Formulate Design Principles

Based on our discussion of the design parameters in the previous section, we now unfold de Sitter's *design principles*. These principles provide the rationale to select a particular organizational structure out of a set of possible ones. They provide de Sitter's answer to what an organizational structure should look like – if it is to attenuate disturbances and amplify regulatory potential. This answer rests on the effect of different parameter-configurations on the structure's controllability – as discussed in the previous section. The design principles allow a designer to make well-reasoned decisions as to the grouping of transformations into tasks and the relations between tasks at every step in the design process. Here, relevant questions are for instance, "what should the production structure at the macro-level of the organization look like given specific environmental requirements?" or "what regulatory tasks should be allocated to a team?"

De Sitter formulates one overall design-principle that provides a heuristic for designing organizational structures. This principle, in essence, reads: "An adequate organizational structure (with optimal controllability) is designed by setting the seven parameter values as low as possible". In formulating this principle, he explicitly chooses for one (and against other) specific configurations of parameter values. To understand that this principle makes sense we have to explore the reasons why setting the parameters as low as possible does, indeed, help in optimizing controllability. We will do so be contrasting two configurations: one in which the parameter-values are set as high as possible and one in which these values are as low as possible. We discuss the effects of these two configurations on the production and the control-structure and thus gain insight into how two (extreme) configurations affect controllability.

7.3.2.1 The Effect of an Organizational Structure with Maximum Parameter Values on Controllability

To understand the impact of organizational structures with high parameter-values on their controllability it is useful to explain what such structures (i.e., their combined production and control structures) look like. To start with, three design parameters are at stake when describing the production structure: the level of functional concentration; the level of differentiation and the level of specialization. If these three parameters have maximum values, they result in a production structure that may be represented by Fig. 7.14

The production structure in Fig. 7.14 contains the three differentiated operational sub-transformations that are specialized into several parts (making and support into four, and preparation into three parts). Moreover, all three order-types (X, Y and Z) pass through all preparation and making sub-transformations. In such structures, there are many relations between support and making; between preparation and making and between orders, preparation and making (see arrows). In practice, organizations with such a structure have departments in which

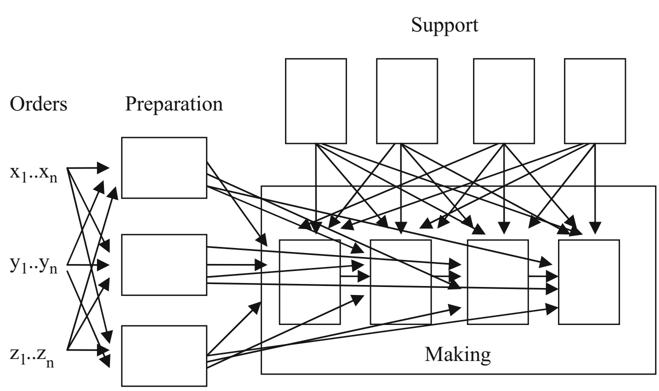


Fig. 7.14 Production structure with a high level of operational differentiation, separation and functional concentration

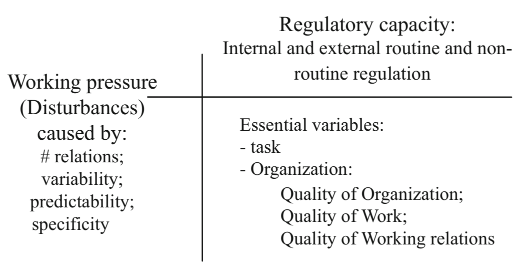
capacities are devoted to one specific sub-transformation. For example, “preparation”-departments like “planning,” “purchase,” “sales”; “support”-departments like “personnel,” “maintenance,” and “making” departments all containing capacities regarding one operational sub-transformation – e.g., a “sawing,” “drilling,” “painting,” and “assembling” department. Although the structure in this figure already gives some idea of the complexity of the resulting network of tasks, it does not – by far – represent the complexity faced in “real-life maximum-parameter-value” organizations, which have far more functional departments (for preparation, support and making), and in which the specialization into small tasks may be taken much further.[11]

In the organizations we are describing in this section, the level of separation between operational and regulatory sub-transformations is high – meaning that the control structure is detached from the production structure: i.e., besides a network of operational tasks, a separate network of regulatory tasks exists. For individual operational tasks, this entails that all four forms of regulation are diminished as far as possible.

To describe the control structure, de Sitter proposes three parameters: the level of differentiation of regulatory transformations into parts and into aspects and the level of their specialization. A control structure with high values on these parameters typically consists of a hierarchy of (small) regulatory tasks. Moreover, in such control structures the three regulatory sub-transformations monitoring, assessment and action-selection are assigned to different tasks, and specific tasks (or sub-networks of tasks) are devoted to strategic regulation, regulation by design, and operational regulation.

Given this description of an organizational structure with high values on all parameters, the question for this section is what effect it has on relevant organizational variables. To evaluate this effect, we use de Sitter’s re-specification of Ashby’s model of regulation (see Fig. 7.15).

Fig. 7.15 Organizational disturbances and regulatory potential



[11] In the text, production organizations are taken as examples. However, as de Sitter stresses, the same reasoning holds for organizations providing services (e.g. governmental organizations; consultancy, etc.)

As discussed earlier, de Sitter argues that disturbances for a task may be caused by the number of relations a tasks has with its environment, the variability of these relations, the predictability of environmental change, and the specificity of the task. These parameters reflect what de Sitter calls "working pressure." To deal with this pressure a task or a network of tasks requires sufficient regulatory potential (in terms of the four ways of regulation we discussed earlier). A mismatch between regulatory potential and working pressure (i.e., a problematic value of controllability) results, at the level of a task, in the situation that norms of the relevant variables of its operational sub-transformation (its output) are not met. At the level of the network of tasks (the organization) this mismatch results in problematic values of the three "quality-variables." And, in accordance with Ashby, de Sitter holds that by means of organizational design, working pressure may be attenuated and regulatory potential may be amplified.

Based on this model we are now in a position to evaluate the effect of the organizational structure with maximum parameter values on its controllability. We do this by showing how high parameter values of the production and control structure lead to an increased working pressure and, at the same time, to a decreased regulatory potential. Next, we briefly discuss the impact of high working pressure and diminished regulatory potential on realizing tasks and on the three organizational quality variables.

The Effect of High Parameter Values of the Production Structure

The effect of high functional concentration and differentiation (into preparation, make and support transformations) is that (1) all preparation and make transformations are potentially coupled to all order-types, (2) support transformations are potentially coupled to all make transformations, and (3) that all transformations of the same type are clustered into separate departments. High specialization adds to this that these operational transformations are decomposed into a large number of small sub-transformations.

A production structure with these characteristics leads to an increase in working pressure. Differentiation and specialization introduce new sub-transformations and tasks that have to be coupled and thus lead to an increase in relations. As a simple example, suppose that I operate two machines to produce something and that I am also able to maintain both machines as I see fit. For "efficiency purposes," a separate "maintenance" task may be introduced (differentiation). Moreover, this task may be split up into two sub-tasks tailored to maintaining a specific machine (specialization). This introduces new relations, for each time something is wrong with a machine I have to consult one of the maintenance experts. And, as the Sitter argues, every relation adds a possible source of disturbances. These disturbances may be generated by chance (illness, malfunctioning of equipment, etc.) or by the production structure itself. For instance, a shared support-function introduces possible conflict between relations: simultaneous consults have to be scheduled – resulting in delay in the network (see also Galbraith 1973). In general, then, differentiation and specialization lead to an increase in the number of relations in a network.

A high level of functional concentration results in a high level of variability. De Sitter explains why this is the case. Suppose an organization produces two types of products: tables and chairs, and that both tables and chairs come in three different variations. In this case, at any moment in time, the environment can ask one of $2^6–1$ combinations of different order-combinations.[12] This means that the production-structure should be capable of "dealing" with all these combinations, which, in turn, entails that the relevant (order-related) network of operational tasks should (because they are, potentially, coupled to all orders) be capable of dealing with these combinations. Every order combination means specific planning (of *this* combination); purchase of raw materials (for *this* combination); preparation of tools and equipment (for the production of *this* combination); sequences of make-operations (tailored to *this* combination), etc. And, because the tasks are related to realize the order-combinations, the relations between the tasks should, potentially, be as variable as the number of order-combinations. De Sitter concludes: the higher the number of products; the higher the number of order-combinations (which increases exponentially with the number of orders see de Sitter 1994, p. 53 ff.) and, hence, the more variable the relations in a functionally concentrated production structure. As we have indicated earlier, a high level of variability leads to an increase in the probability of disturbances. In this case: each new order-combination introduces a new possible source of disturbances, and a high number of order-combinations means that tasks themselves may become more complex (e.g., planning order-routings for six products (63 order-combinations) seems to be more complicated than for five products (31 order-combinations), which means that more can go wrong.

A production structure with high parameter values has a large number of relations and a high level of variability. De Sitter goes on to point out that both effects lead to a lack of predictability. As he puts it: the higher the number of relations and the more variable these relations, the lower the possibility to predict the input to some task (1994, p.144; p.158). Moreover, because sub-transformations of the same type are clustered in different departments the overview over orders is decreased to a considerable degree. De Sitter (1994, p. 43) describes: as soon as orders are placed, they are "made invisible" by "translating them into a sequence of required operational activities and sorting them with regard to the first operational activity" (1994, p. 43). Throughout the production process piles of parts of semi-finished products arrive at an operational department. Next, they are distributed among sub-transformations to undergo some operational activity, and finally they

[12] At any moment, a specific product (table, type 1; table, type 2; table, type 3; chair type 1; chair type 2; chair type 3) can either be ordered or not. This results in 2^6 possible combinations of orders because 6 specific products can be ordered or not. However, the combination in which no products are ordered does not count as an *order*-combination, so the total number of possible order-combination is $2^6–1$. This number is even an underestimation, because it does not yet include e.g. possibly differing order quantities.

are grouped together for shipment to the next department. Individual workers realizing a small operational transformation, cannot, as a consequence, recognize the order as order. They only know that a certain activity regarding some piece of material has to be performed. In other words: at the level of an operational task one can only see the effect of the operations belonging to it, one can neither see the order one contributes to nor the rest of the operational process leading up to it.

For someone who has to distribute the work among operational workers, the order is also hard to recognize – most of the time incoming material from some department has to be distributed in order to undergo some operational treatment according to some fixed planning. Throughout the process, then, only piles of material (representing parts of semi-finished products) are visible – not the orders or products themselves. Not being able to recognize orders or the process of producing them decreases the predictability and hence increases the working pressure. De Sitter: the situation in which "workers neither know what is coming, where it is coming from, nor where it is going to" (1994, p. 31) describes low predictability. A lack of overview due to the design of the production structure has many (other) consequences which will be discussed below.

The Effect of High Parameter Values of the Control Structure

Above, we argued that high values of the production structure parameters cause an increase in working pressure. On top of that, a separated control structure with high parameter values further increase working pressure and leads to a decrease in regulatory potential. Separation, differentiation and specialization of regulatory sub-transformations obviously increase the number of regulatory tasks, and therefore, the number of relations in the network. Just as we discussed in the context of specialization and differentiation of the production structure, this increases working pressure. Generally speaking, it leads to a complex network of regulatory tasks each with a small regulatory scope, with regulators lacking insight in and necessarily detached from the operational transformation they are supposed to regulate (de Sitter 1994, Chap. 9). Moreover, it has specific consequences for the regulatory capacity of individual tasks and of the network as a whole. To see this, it is useful to remember the four ways of dealing with disturbances discussed earlier:

1. Internal routine regulation (regulatory activity does not involve other tasks and does not change the task or network of tasks)
2. External routine regulation (regulatory activity involves other tasks but does not change the task itself of the network of tasks)
3. Internal non-routine regulation (the regulatory activity does not involve other tasks but changes the task's infrastructure and/or adds routine regulatory potential)
4. External non-routine regulation (the regulatory activity either changes the task's essential variables and norms or changes the infrastructure of the task's environment).

Separation, differentiation and specialization of the regulatory aspect of a transformation have an effect on all four forms of regulation.

To start with, at the level of individual tasks, it decreases the *internal routine regulatory capacity*. Specialization of tasks in the production structure already resulted in a reduction of the *object* of regulation. For instance, due to specialization I now no longer make a chair or a table but I am supposed to saw pieces of wood. This kind of specialization obviously reduces the object of regulation – for now I deal with disturbances regarding sawing pieces of wood – not regarding making a chair or a table. Therefore, reducing the object of regulation implies a reduction of the (now redundant) internal routine regulatory repertoire. Separation, differentiation and specialization of the regulatory aspect of the task further reduce internal routine regulatory potential: they introduce new regulatory tasks that all realize parts of the internal routine regulatory capacity of the original task. For instance, due to separation I may now have a boss, responsible for monitoring the output, comparing it to some standard and proposing adjustments. And, the more separation, differentiation and specialization, the more the original internal regulation gets dispersed throughout the regulatory network, which, in turn, means a loss of the possibility to react immediately to occurring disturbances.

De Sitter also notes that high parameter values lead to a loss of internal routine regulatory capacity because the number of combinations of regulatory actions decreases. For instance, suppose that, originally, I had a task involving six different operational sub-transformations. If something went wrong while completing the fourth operational sub-transformation, I might have dealt with it by solving the problem with this sub-transformation and performing sub-transformations five and six more quickly. Such "overall regulatory solutions" are combinations of individual regulatory actions. Now, if, by specialization of the production structure my task contains fewer sub-transformations, or if by separation, differentiation or specialization of the control structure my task contains fewer regulatory actions, the number of regulatory combinations, and therefore the internal regulatory potential, also decreases. De Sitter explains that this is an exponential decrease (p. 161). In one task with six regulatory actions, $2^6–1$ regulatory combinations are possible. Removing one regulatory action results in $2^5–1$ regulatory combinations, etc.. Eventually, in tasks with only one regulatory action, internal regulatory combinations can, of course, no longer be formed. To be sure, such combinations may still be possible – but only by means of more complex process-regulation (involving regulators external to the task and with an overview over the process).

In all, high values of the parameters describing both the production and control structure result in tasks with a low internal routine regulatory capacity. They also result in tasks with a low *external routine regulatory capacity*. This capacity means trying to deal with disturbances involving other tasks in the network, without changing the network of tasks. For instance, asking a previous work-station to fix some error. High parameter values of both the production and control structure seriously impair the external routine regulatory capacity. Because of the involvement of other tasks in dealing with disturbances, external regulation requires an overview of the process – i.e., one has to know which other tasks can be involved.

This requires, at least, knowing what the previous tasks in the process are (to enable determining where the input for a task comes from) and what the next tasks in the process are (for determining the tasks dealing with the output of a task). Given this knowledge, external routine regulation also requires communication between tasks to actually deal with the disturbance. High parameter-values of the production structure, however, increase a lack of process overview and make communication virtually impossible. As we argued above, in these structures the process of making orders is made invisible for individual workers, performing operational sub-transformations – they only see the small operational part they are assigned to. As de Sitter puts it: these workers do not know where it [task input] comes from and where it [task output] goes to (see above). Moreover, due to specialization and separation, "communication with other tasks" as regulatory activity is no longer part of a task. To put the matter bluntly: the required communication is impossible because it is no part of the job, and, if even it were a part of the job, it would be useless, because one would not know with whom to communicate.

Even if one knew which other tasks one had to involve to deal with a disturbance externally, the reduced *internal* regulatory capacity of this other task probably would not allow its being involved in dealing with the disturbances of *your* task. This raises an important point: high values of parameters regarding production and control structure diminish the *internal* routine regulatory potential of a task and, at the same time, the *external* routine regulatory potential in terms of its being able to assist in dealing with the disturbances of other tasks is also decreased.

Separation, differentiation, and specialization of the control structure make matters even worse: external routine regulation of an operational task (if it exists at all in such structures) now involves many regulatory tasks (some regulator, who passes on the request to someone else, who passes on the request, . . .) before it eventually reaches relevant other operational tasks. The same holds for dealing with disturbances in the regulatory network itself. In such structures, de Sitter holds, external routine regulation is, because of their structure, nearly impossible, (1994).

Above, we discussed the effect of high value parameter structures on routine regulation. We will now turn our attention to the impact of these structures on the effectiveness of non-routine regulation. *Internal non-routine regulation* at the level of a task means dealing with disturbances by changing the task itself – i.e., by changing its infrastructure. One problem in structures with high parameter values is that, because the task itself is small, the object of infrastructural change is also very small. That is, improvements brought about by internal non-routine regulation only pertain to small parts of the process. Moreover, because of a lack of overview over the whole process one cannot observe the consequences of a change in the task's infrastructure for (the infrastructure of) the whole process – leading to sub-optimization at best. However, in most organizations with high-value parameters, internal non-routine regulation is removed from operational transformations and assigned to other tasks. At the level of these "other tasks" (regulators responsible for "regulation by design") an overview of the whole process is still required – something that is hard to obtain in these structures. At the same time, de Sitter notes that separating and differentiating internal non-routine regulation may lead to regulators who are too detached from the tasks they are supposed to change. It results, as he puts it, in " detached regulation," in which regulators can, because of their lack of involvement in the process of performing operational sub-transformations, neither be really sensitive to nature of its disturbances nor to the effects of the proposed "improvements." They structurally miss relevant "process-information" (de Sitter; p. 329 ff).[13]

External non-routine regulation at the level of a task means dealing with disturbances by changing either the goals of the own task or of other tasks; changing the infrastructure of (parts of) the network of tasks; or changing the environment of the organization. This form of non-routine regulation is greatly impaired by (if not, made nearly impossible in) a structure with high parameter values. If a task has only small specialized operational sub-transformations and does not allow an overview over the process producing orders, the effect of changes to goals or to the network's overall infrastructure cannot be appreciated. To give an example, process innovations (meaning a change in the infrastructure of the network) are very hard to initiate because mostly, disturbances stem from the complexity of the network of tasks and not so much from the complexity of the tasks themselves (de Sitter 1994, p. 50). And it is this network-complexity that cannot be overseen by tasks. Moreover, because of the reduced regulatory aspect of the task, it is impossible to experiment with and learn about effects of "improvements." Suggesting process improvements, then, necessarily becomes a matter of those not directly involved in the processes they want to improve. But this may be problematic – precisely because of their lack of involvement.

To cut a long story short, in structures with high parameter values, internal and external non-routine regulation are separated from operational sub-transformations, and assigned to differentiated, specialized tasks in the network – leading to an increase in the number of tasks and their relations in the network (and therefore to an increase in the probability of disturbances), and to a decrease of the effectiveness of regulation, because disturbances cannot be countered immediately, but should await "interventions" of structurally detached regulatory tasks.

The Effect of High Working Pressure and Reduced Regulatory Potential on Realizing Tasks and on Organizational Quality

Above, we set out to argue that tasks in organizations with high values on all parameters face high working pressure and have low regulatory potential. Needless to say, this affects realizing the task's output. While realizing a task one has to take into account a large number of relations, which are highly variable and rather unpredictable. This increases the probability of disturbances. At the same time the task's decreased regulatory potential impairs dealing with them. To deal with

[13] This is all too often observed in situations where "ICT-departments" offer (impose) "ICT-solutions" that do not support the process they are supposed to improve.

disturbances, one has to rely on a complex network of regulators, which can be a source of disturbances itself, because of its complexity. Being poorly equipped to deal with disturbances, tasks tend to infect other tasks and thus propagate disturbances throughout the network.

Structural consequences at the micro-level (the task-level) thus translate themselves into consequences at a meso-level (groups of tasks – such as departments) and at the macro-level (the level of the organization as a whole). To describe the consequences at a macro level, de Sitter proposes to use the three "quality-variables" we already discussed: quality of organization, quality of work and quality of working relations. Below, we will briefy address consequences for these three variables.

Quality of Organization

To start with, de Sitter states that organizations with high values on all parameters cannot meet the functional requirements with regard to *quality of organization* (see Table 7.4). The complexity of the network, the increased probability of disturbances and the increased lack of adequate regulatory potential necessarily affect production cycle times. Disturbances are not dealt with immediately, but by means of a separated and differentiated regulatory network, which, of course, takes time. Moreover, if disturbances tend to affect other tasks as well, the production process is even more disrupted, and requires more regulation – and hence, more time. De Sitter discusses a particular detrimental effect of maximum functional concentration on production time: the structural accumulation of stock. In functional concentrated structures operational sub-transformations of the same type are clustered together in "functional departments" (e.g., a "sawing department," a "drilling department," an "assembly department" etc.). In such structures departments are separated, and it would be rather tiresome to bring some piece of material to the next department every time an operational activity is finished. A common practice is to schedule work in batches. That is, after a fixed period of time all semi-finished products (output of the department) are collected and delivered to the next department. However, a consequence of this practice is that stock accumulates within departments, leading to longer production cycle times. De Sitter even calculates that the actual production time of a product in an organization with a functionally concentrated structure is often only 5% of the total production cycle time – meaning that a product is "stuck in stock" for 95% of the time it takes an organization to actually finish it (de Sitter 1994, p. 47).

Another functional requirement of the quality of organization is the level of control over order realization. This, too, is problematic in high value structures. Planning of orders is complex and, due to a lack of overview over the process it is very hard to trace individual orders. It is also hard to ensure the effective control of quality of products. One reason is that quality control (in essence: making sure that products meet their required demands) is dispersed throughout the regulatory network. Moreover, because of functional concentration, one structurally lacks insight in the interrelation between the different operational sub-transformations and its effect on the quality of the end-product (ibid. p. 48). This means that one cannot anticipate (during the performance of operational sub-transformations) possible failures of the products. Quality-control, then, is, at best, a reactive activity.

The last functional requirement of the variable "quality of organization" is the potential for innovation. As we discussed earlier, process-innovation is almost impossible, because of a lack of overview over the process. It is necessarily confined to the small sub-processes one is assigned to (if, that is, one has the required possibilities in a task to think of new ways of doing things and to experiment with them). Product-innovation is also difficult – because, as de Sitter argues, it requires a thorough understanding of the whole process of developing, making and selling products and therefore an intricate co-operation between marketing, product-development, production design, production, logistics, sales and service. But, such co-operation is difficult because of differentiation and, again, because of a lack of overview over the whole process.

Quality of Work

De Sitter holds that quality of work mainly depends on (1) the stress associated with task-performance, (2) the (im)possibility to be/feel involved, and (3) the opportunities a task offers to learn and develop. Not surprisingly, de Sitter's contention is that a low level of stress, and opportunities for involvement and learning have a positive effect on the quality of work. However, organizations with high values on the structural parameters fall short to meet these requirements.

To start with, the probability of task-related stress in these organizations is high. As de Sitter points out, feelings of stress with regard to work have to do with "the situation in which you face problems but are unable to solve them" (p. 21).[14] Given this description, de Sitter further operationalizes stress experienced while performing a task as "the possibility one has to deal with disturbances." Now, if the internal regulatory potential in a task is low, one has to depend on the task's environment to deal with disturbances. However, in organizations with high values on the structural parameters, this is problematic because other tasks also have low internal regulatory potential. Therefore, the external regulatory potential is also low. In such organizations, then, the probability of task-related stress is high because tasks face high working pressure (potential problems) and, at the same time, tasks have limited regulatory potential to solve them.[15]

A second problematic consequence of high parameter value organizations for the quality of work is that it is hard to "feel involved." In fact, the probability of alienation of workers is rather high in such organizations. Feelings of involvement

[14] See also Christis 1998, for a more elaborate explanation of the concept of stress based on de Sitter's theory.

[15] De Sitter shows (p.28) how results of a study by Karasek (1979) empirically support his conception of stress.

with regard to work may come about in at least two ways. One may feel "socially" involved because one is – while performing a job – actively engaged in a network of social relations associated with the job. One may also feel "intrinsically" involved because one is able to see and appreciate the point of the task – i.e., its contribution to the process of producing something as well as to the product itself. It is possible to feel intrinsically involved in a passive or more active way. If one can see the point of what one is doing and can attach meaning to one's task, one can be said to be feeling intrinsically involved in a passive way. As de Sitter (and, of course, many sociologists with him) argues: not being able to see the point of what one is doing can be seen as alienation from the task. And, alienation leads to decreased responsibility. Such passive intrinsic involvement requires an overview over the process – which can only come about in organizations that have not split up tasks too much. If one is also actively involved in co-defining or changing the task's goals, the involvement can be said to be active. Both social and active intrinsic involvement require an overview and external regulation. External regulation implies being actively involved in a network of social relations; it is by means of this network that one tries to deal with problems in a routine or non-routine way. Moreover, by definition, external non-routine regulation is (in part) about being able to (re) define goals. Thus, given this view on involvement, it is easy to see that organizations in which external regulatory capacity has been taken away from tasks, workers who are performing such tasks may not feel involved – both socially as well as intrinsically.

The last aspect attached to the quality work that is problematic in high-value parameter organizations concerns the opportunity to learn and develop. Separation, differentiation and specialization tend to make work uninteresting and repetitive. Jobs, then, neither challenge workers to learn new things nor mobilize them to contribute to the best of their ability to the organization's viability. What is more, these jobs make such contributions virtually impossible, for, as we noted earlier, workers are deprived of the possibility to think of new ways of doing their work, experiment with them and learn from such experimentation. Learning is necessarily confined to the small task one is assigned to, which often means that there is not much to learn. Leaning about the process of producing orders and about how to improve one's contribution to this process is, again, virtually impossible, because of a lack of overview.[16]

In all, the quality of work tends to be low in organizations with high values on all parameters. An increased risk of experiencing stress, of feeling non-involved, and few learning opportunities often lead to high levels of absenteeism and personnel turnover. One may conclude that such organizations do not offer workers to live their working lives to their fullest extent – a point taken up in Chap. 11.

[16]These points can be made, based on a more-or-less common sense view of the concept of learning. De Sitter, however, even unfolds his own (functional) model of learning in organizations (1994, chapter 10), in which learning is tied to regulation. He then uses this model to explore the (structural) requirements for learning. In the context of our book, however, it suffices to discuss structural impairments for learning in general.

Quality of working Relations

The last functional requirement to evaluate organizational structures de Sitter uses is the quality of "working relations." This has to do with effective communication and cooperation within the organization. To be effective, it requires knowledge about and involvement in the process one has to communicate about. As we discussed earlier, members in organizations with high values on parameters structurally lack an overview over the process and are often not involved. Therefore, the possibility to deal with disturbances by means of cooperation is structurally diminished. And, at the same time, non-involvement affects shared responsibility, necessary for effective communication and cooperation. As de Sitter puts it, in such organizations communication is often limited to what one "really" has in common – which, unfortunately, often boils down to "the workspace, toilets, coffee, [. . .]" (1994, p. 52).

7.3.2.2 The Effect of an Organizational Structure with Low Parameter Values on Controllability

What we now have is an overview of how the structure of an organization with high values on parameters leads to high working pressure and low regulatory capacity and how it negatively affects both performing individual tasks and maintaining the viability of the organization as a whole (as described in terms of the three quality-variables). In this section, we discuss de Sitter's alternative to these structures: those with low values on all parameters – and argue how they manage to reduce working-pressure and, at the same time, increase regulatory potential – at the level of individual tasks and at the level of networks of tasks. In many respects this reasoning mirrors the one given in the previous section – the main idea is to show that low parameter values attenuate disturbances (by reducing the number of relations in the network of tasks and the variability of these relations) and, at the same time, amplify all forms of regulation (internal, external, routine and non-routine).

Our strategy showing the effect of low parameter structures is similar to that of the previous section: we first describe their effect on the production and control structure and then move on to the consequences of these structures for realizing tasks and organizational quality (in terms of the three "quality-variables").

Before we can go into the effect of organizational structures with low parameter values, it is a good idea to give a short description of their related production and control structure. A production structure is described by three design parameters: the level of operational differentiation; the level of separation; and of functional concentration. If these parameters have low values, a production structure as given in Fig. 7.16 may result. As compared to a production structure with high parameter-values (see Sect. 7.3.2.1 / Fig. 7.14) order-types are produced in parallel "order-flows" due to a low level of functional concentration. Each order-flow is dedicated to producing one type of orders. Because of a low level of differentiation

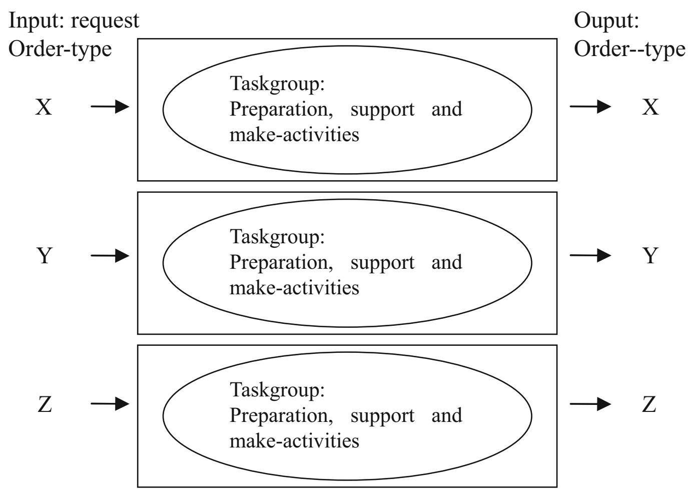


Fig. 7.16 A production structure with a low level of operational differentiation, separation and functional concentration

in each flow, making, support and preparation activities are integrated as much as possible. In addition, because of a low level of specialization, activities are not split up in many sub-transformations. As compared to the production structure with high parameter values, this structure does not have "functional departments" in which specific activities are central (e.g., specific preparation, make or support departments – see previous section). Instead, activities are organized around more or less autonomous parallel flows producing a specific type of output (e.g., a specific type of order – e.g., qua product; type of customer or geographical area). Within these flows, activities, required to produce the output are integrated into so-called "task-groups" or semi-autonomous groups (de Sitter, Chap. 8) see also Fig. 7.16. These groups resemble what Galbraith (1974) called "self-contained tasks," a specific configuration of tasks in which "[. . .] the subtask grouping [is changed] from resource (input) based to output based categories and [. . .] each group [. . .is given. . .] the resources it needs to supply the output. For example, the functional organization could be changed to product groups. Each group would have its own product engineers, process engineers, fabricating and assembly operations, and marketing activities. In other situations, groups can be created around product lines, geographical areas, projects, client groups, markets, etc." (Galbraith 1974, p. 31).

Ideally, within these order-flows, task-groups contain all possible (prepare, support and make) activities to produce the output they are assigned to (e.g., some product or order-type). Workers to whom such tasks are assigned, can, in principle, carry out all required activities for producing this output of the flow. As compared to the specialized tasks in the high-value structures, tasks in low-value structures contain a "redundancy of functions" so as to be able to quickly react to disturbances (see below).

In practice, however, producing some product often requires a large number of complex activities, and, hence, task-groups can not cover the entire production flow. In this case, de Sitter proposes to introduce (serial or parallel) segments in the flow, and assign task-groups to these segments. Defining segments should be done while keeping in mind that the level of functional concentration should be as low as possible. The idea is to define sub-output within a flow and assign a task group dedicated to producing that output; *not* to define segments containing *activities of the same type* which would re-introduce functional concentration.[17] It should, in other words lead to more-ore-less autonomous units within autonomous units – see next section for an elaboration of this idea.

In a low-value organizational structure, the level of separation between operational and regulatory sub-transformations (parameter 4) is low. This means that the production and control structure are integrated as much as possible. Instead of a separate network of tasks aimed at the regulation of operational tasks of the production structure; relevant regulatory tasks are *part* of operational tasks.

The control-structure of a low-parameter value organization has low values on (1) the level of differentiation of regulatory transformations into parts, (2) the level of differentiation of regulatory transformations into aspects, and (3) the level of specialization of regulatory transformations. Instead of a hierarchy of many small regulatory tasks, all aimed at different aspects of the control cycle and at different levels of regulation (operational, design and strategic) – which is characteristic of the high-value parameter control structure, this control structure is "flat" and consists of "integrated" regulatory tasks, including as many regulatory aspects, parts and levels as possible, covering as much of the production structure as possible.

Ideally, this leads to a structure in which the task-groups not only have the operational means, but also the regulatory means to produce the desired output. However, it should be noted that it is impossible and undesirable for complex production processes to integrate all regulatory tasks into a task-group. In our discussion below, we will assume organizations in which the values of the parameters concerning the control structure are as low as possible – e.g., in which the tasks-groups also have the regulatory capacity needed to produce the output (or even to change the infrastructure in which they produce their output). De Sitter (1994) gives numerous examples of organization in which these parameters have low values. At the end of this chapter we will discuss the issue of "how low" the parameter values of both structures can get; in the mean time it is helpful to keep in mind that we discuss the advantages of lowering the parameter-values as much as possible.

In all, if high value structures can be characterized as complex structures (large complex networks) with simple jobs, low value structures are simple structures (more or less autonomous parallel flows with a built-in control structure) containing

[17] In fact, segments should be defined in such a way that they may contain highly interdependent activities, but the dependence between segments should be as low as possible.

complex jobs (integrated tasks – involving both many operational as well as regulatory sub-transformations).

Given this idea of what an organizational structure with low parameter-values may look like, we will now turn to discussing its effect (1) on the production and control structure and (2) on the essential organizational and tasks variables. The recurrent theme in this discussion is that low parameter-values lead to a structure attenuating working pressure and, at the same time, amplifying regulatory potential. And, in turn, decreased working pressure and increased regulatory potential will lead to higher levels of the organizational quality variables.

The Effect of Low Parameter Values of the Production Structure

In a production structure with a low level of functional concentration parallel "orders-flows" emerge – "units" tied to producing a particular order type. In these flows, due to a low level of differentiation preparation, make and support activities are integrated as much as possible into tasks. A low level of specialization ensures that these tasks are not split up in too many short-cycled sub-tasks. Instead, "integrated" tasks are designed covering a large part of the production process.

As compared to a production structure with high values on these parameters, de Sitter argues that working pressure in low-value structures is far less. To be more precise: in these structures variability decreases dramatically (mainly by means of creating parallel order-flows); and the number of relations decreases (because of integrating operational tasks). To illustrate these effects, let us consider a high-value production structure of a plant producing wooden furniture, as given in Fig. 7.17. This structure has a high level of differentiation, specialization and functional concentration. It produces 8 different products: three types of chairs (x); two types of tables (y), and three types of cupboards (z). Due to a high level of functional concentration, all make and prepare departments are, in principle, coupled to all of these products (see earlier). This leads to a high level of variability:

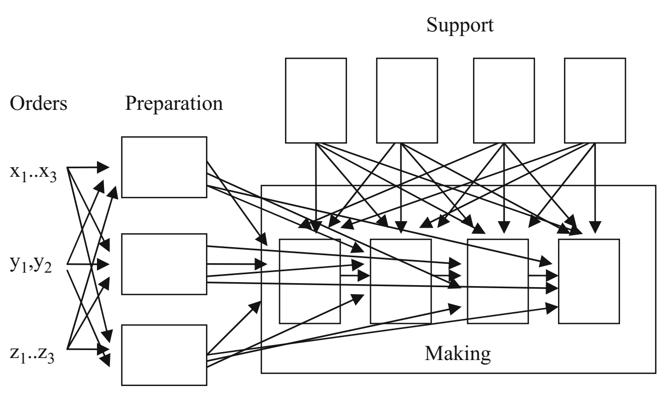


Fig. 7.17 A production structure with high values of design parameters

each prepare and make activity should be able to deal with every possible order-combination (a particular combination of different products, demanded by customers, at some moment in time). A particular order-combination in a particular week may, for instance, be: “2 chairs, type 1, 4 chairs type 2, 18 tables, type 2 and, 24 cupboards, type 3”. For the sake of simplicity, we forget about the specific product-type quantities, and just consider the different types of products in an order-combination. In that case, there are 255 different order-combinations (= 2^8–1). As we have argued in the previous section, this production-structure must be able to process each of these combinations. Every combination requires a different way of doing things (e.g., in terms of planning, purchasing raw material, making-activities like sawing, drilling, assembling, etc.). Moreover, the relations in this network should also be able to handle this variety: every order-combination requires a different interaction between operational tasks (in terms of information or material). So, the variability (the variety of the interactions in the network of related tasks) is tied to 255 different order-combinations. It is convenient to use this number of order-combinations to express the variability of the network (and it is still an underestimation, because of the simplifications we’ve just made).

De Sitter argues that a high level of variability has a negative effect on working pressure. As we discussed before, the higher the variability, the higher the demand on the task processing it. Moreover, this effect increases exponentially with more relations Because the content of the relations between the tasks can differ considerably from time to time (qua information or material), the chance of something going wrong is also substantial (it is considerably higher than in the situation where the content of the relations is always the same).

Now, suppose that the level of functional concentration is decreased in such a way that the production structure contains three parallel substructures – each tied to a particular order-type: chairs, tables and cupboards. Figure 7.18 shows this situation. In this production structure, the variability is reduced enormously. In the substructure dealing with chairs and cupboards, the variability is now 7 (2^3–1), while in the substructure producing tables, the variability is even as low as 3. This is a reduction of complexity in the production structure from 255 to 7. In this new production structure the probability of something going wrong (as a function of the variety passing through the interaction-channels in the network) is, of course, far less. As a result, activities required in the sub-structures become simpler. For example, planning is much easier in the new structure. Also, because each flow only has to deal with one product (in two or three types) production can go faster; for it means that less time is needed to change material or to set up machines (a result that cannot be underestimated in many complex production processes).

The example shows the reduction of variability due to a decrease in functional concentration. De Sitter admits that it may not always be possible (or even efficient) to define such “autonomous sub-structures.” However, it should be kept in mind that the example is primarily meant to show the effect of lowering the value of functional concentration – i.e., reduction of variability. At the same time, de Sitter proposes designers to start with looking for autonomous flow-oriented sub-structures (like the ones in the example). If these substructures cannot be found at

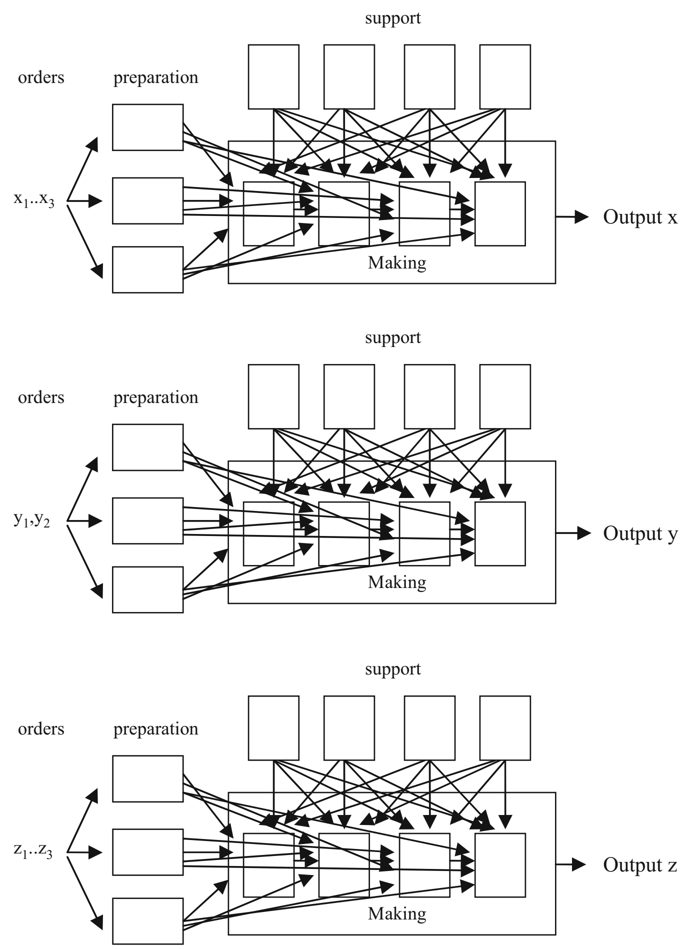


Fig. 7.18 A production structure containing three parallel production flows coupled to order-types

the level of complete orders (or order-combinations), functional concentration should be diminished within the existing structure, by identifying parallel segments in it and assign ("smaller") semi-autonomous sub-structures to these segments (cf. de Sitter 1994, Chap. 7 for the many ways in which this could be accomplished). For example, in some cases he proposes to define autonomous flow-like sub-structures tied to producing product-components that share a particular route. This may lead to different more or less parallelized semi-autonomous teams each producing product-parts that have a different route[18]. The main idea is that one

[18] In this case the "orders," relative to which functional concentration is defined is an "internal" concept; they comprise the requested product-components by the production-process.

should define a complete "in-out"-configuration and cluster tasks in such a way that they are dedicated to that particular output. This can be done at different levels – at the level of complete orders or at the level of parts of the production of an order. It should, however, NOT lead to a configuration of tasks that can, in principle, be tied to every order.

If the levels of specialization and differentiation are also lowered, (1) preparation, make and support activities are part of one task as much as possible (low level of differentiation), and (2) activities are not split up in too many short-cycled tasks, but integrated as much as possible (low level of specialization). For instance, each flow may consist of one or more "task-groups" – groups of workers capable of performing many (if not all) activities required producing the output (see Fig. 7.16 for an example of such a production structure with three different types of products).

The main effect of these parameter settings is that the number of interfaces between tasks is reduced. This has two related results. Firstly, since every additional interface introduces an additional source of disturbances in the production process, task-integration means reduction of disturbances. Secondly, if tasks are not split up, and if different types of tasks are not grouped together in separate departments, the overview over the production process increases, which greatly enhances regulatory potential (see below).

For instance, if a worker in a differentiated structure faces a problem with the machine he operates, he needs to file a request for support. Only after processing this request (which entails passing on the request to the relevant support-unit; planning support activities; and additional communication about the problem) the problem may be resolved. In a non-differentiated structure such problems may be resolved by the worker himself – which saves the time needed to file and process the request for support.

Integrating tasks in terms of de-specialization, also reduces the amount of interfaces between tasks (and hence disturbances). For instance, integrating "make"-tasks (1) means a reduction of disturbances (due to the many interfaces), (2) increases the overview over the production-process, and (3) also often means a reduction of stock. Typically, in a highly specialized production structure, semi-finished products undergo some operational activity and are grouped together for shipment to another department (i.e., they are produced in batches). This has two consequences: (1) every department has its own "stock of semi-finished products," and (2) there is virtually no overview over the production-process as a whole – the overview over the production process is very low. As we said earlier, one cannot see the product one is contributing to nor the rest of the operational process leading up to it. Now, integration of make-tasks into a flow offers the possibility to stop producing in batches (which reduces stock-size dramatically – cf de Sitter 1994, or Womack et al. 1990) and it also means an increased overview, since the new tasks now contains at least two previously separated tasks.

As with functional concentration, de Sitter is well aware of the fact that for many complex production processes, it may not be feasible to integrate all relevant activities into one task (performed by a task-group). In such a case, segments should be identified within a flow. They should be defined in such a way, that a task-group can be defined, consisting of tasks dedicated to produce the output of the segment. Moreover, these segments should – according to Simon's ideas – consist of activities that depend minimally on other segments (but may have a high internal dependency). We do not go into the details of designing segments within flows; we just want to point at the fact that de Sitter's theory addresses the relevance of identifying segments and gives detailed guidelines for doing so. (1994, Chap. 7)

In all, production structures with low parameter-values have (compared to those with high values) less variability (due to a low level of functional concentration) and a fewer number of relations between tasks (due to low levels of differentiation and specialization). This, in turn, leads, to fewer disturbances (as a function of both variability and number of relations in the network of tasks). In this sense such a production-structure *attenuates* disturbances. Moreover, it may lead to less stock and creates an increased overview of the production-process. Finally, the production structure (because it consists of parallel flows with task-groups) creates chances for lowering values of the control structure; the topic of the next section.

The Effect of Low Parameter Values of the Control Structure

As became apparent in the previous section, low values of parameters concerning the production structure *attenuate* disturbances. In this section we will argue that low values of parameters related to the control-structure (1) further *attenuate* disturbances (by decreasing working pressure) and (2) *amplify* regulatory potential.

In general, a low level of separation (i.e., not separating operational from regulatory transformations) results in tasks containing operational *and* relevant regulatory sub-tasks. A control structure with a low level of differentiation integrates different types of regulatory activities, and a low level of "regulatory" specialization results in the integration of small sub-tasks into tasks with a larger regulatory scope. This results, ideally, in regulation that is not detached from but integrated into operational tasks – enabling prompt reactions to disturbances; and a smaller network of (regulatory) tasks, reducing the number of disturbances. Moreover, tasks with a broader regulatory (and operational) scope lead to an increased regulatory potential and to an increased overview over larger parts of the production process, which, according to de Sitter, has several positive consequences.

To appreciate the result of a control structure with low parameter values in some more detail, we will discuss its effect on the different forms of regulation. To start with, a low level of both separation and regulatory specialization counters regulatory problems with internal and external *routine* regulation encountered in structures with high values on these parameters.

De-*separation* entails that operational tasks include relevant routine regulatory capacity. This inclusion entails that disturbances can be dealt with immediately (dealing with them does not require communication with and action from a network of regulators) and effectively (those dealing with the disturbances are no longer detached from the process they intervene in, but, instead, are operationally immersed in it; a condition for appreciating the effects of alternative regulatory solutions to disturbances). De-*specialization* results in the integration of regulatory tasks into operational tasks, covering a larger part of the (operational) process. This only makes sense, of course, if the production structure is also de-specialized. Together, a low level of both separation and (operational and regulatory) specialization create conditions for immediate, relevant regulation covering a larger part of the production process.

De-separation allows for including both internal and external routine regulatory potential. By decreasing regulatory specialization the routine regulatory potential increases, which means that the number of different combinations of responses to problems also increases. De Sitter points out that this increase is exponential (see the previous section in which we discussed the exponential *loss* of this kind of regulatory variety). Including external regulatory routine potential further increases the variety and flexibility in dealing with disturbances. However, as we have just remarked, this requires that workers have some overview over the production process and that the required communication with other tasks is allowed – otherwise they are unable to involve relevant other workers to help them solving or preventing difficulties.

In organizations with low parameter-values the integration of operational tasks in parallel production flows forms a basis for integrating (routine and non-routine) regulatory potential into operational tasks. Typically (or ideally), in “task-groups” team-members can perform many of the required operational activities and also have the required regulatory routine potential to counter most disturbances. Involving other team-members to prevent and deal with disturbances (external routine regulation) is a natural aspect of working in such teams. Moreover, this external routine regulation is possible in these teams because team members possess the required knowledge about the whole process and because “communication with other tasks” is built into the team-task. In such teams the built-in external regulatory potential of members functions as a back-up in case the internal regulatory potential of some member fails.

This kind of external regulatory potential is tied to individual jobs *within* a task-group. One could also define it as internal regulatory potential of the task-group itself, which should be “built into these groups.” External regulatory potential at the level groups should also be available, if production flows are segmented (and, hence, consist of more than one task-group). However, in such a case, de Sitter points out that it should be (1) direct (between representatives of task-groups; not via a hierarchy or regulators), (2) reciprocal and (3) symmetric; because the information about specific parts of the operational processes and suggested measures for mutual regulation of it can only be appreciated by the respective representatives of the task-groups dealing with these parts. (de Sitter Chap 9).

On top of low levels of separation and regulatory specialization – which result in the inclusion of regulatory potential into (operational) tasks, a low level of regulatory differentiation entails that (many) tasks also contain non-routine regulatory activities. If non-routine regulatory potential is included in a task, it becomes possible to initiate and carry out changes to the infrastructure of task itself, in order to deal with disturbances. This can range from quite simple changes, such as installing new software to assist in certain activities, or small changes to the sequence of steps realizing the task's output into a more efficient one, to more complicated changes such as installing new technological equipment or introducing a new procedure for accomplishing the task. The advantage of introducing such non-routine internal regulation into a task (as compared to define separate non-routine regulatory tasks) is that it can lead to effective and efficient changes to tasks. It can be more efficient, for instance, because one does not depend on a hierarchy of regulators for changes to the task's internal infrastructure, which saves time and extra communication. It can be more effective, because the changes are initiated and implemented by those who carry out the very task that is to be changed. At the same time, including non-routine regulatory potential into a task is also demanding. It requires an overview over the chain of tasks in order to appreciate the effect of certain task-internal changes to the rest of the network of tasks (otherwise it would lead to sub-optimization). In a similar vein, it requires additional "infrastructural" knowledge on the part of those initiating and carrying out the change; on top of the knowledge required for "just" realizing one's task. One needs to know, for instance, how particular new software can aid in performing one's job, and how it needs to be installed, etc.

To deal with such problems, external non-routine potential should be included into tasks, enabling the possibility to discuss changes to the network of tasks / task-groups. Again, de Sitter stresses that such external regulation should be direct, reciprocal and symmetric.

Another way to show the effect of structures with low parameter-values on the control structure is to emphasize the increased quality of information (cf. de Sitter, Chap. 9). De Sitter argues that timeliness, completeness and relevance of information increases. It results in timely information, because there is less "distance" between disturbance and those interpreting it (it does not need to be interpreted by (a hierarchy of) detached regulators). This also leads to less communication and hence possible distortion of the information. Moreover, since operational and regulatory tasks are integrated, information about disturbances, regulatory actions and their effect is directly relevant in terms of its use in learning about them.

The Effect of Low Working Pressure and Increased Regulatory Potential on Realizing Tasks and on Organizational Quality

A structure with low parameter-values has two main effects on realizing tasks in organizations: it attenuates disturbances and amplifies regulatory potential. In Table 7.6, we summarize these effects for each parameter.

Given these effects, we can also consider the impact on de Sitter's quality variables (quality of organization, quality of work and quality of working relations). Because this impact has already been discussed (implicitly) in the previous section, and because it mirrors the impact of high-value organizations, we will only briefly list the effects.

Table 7.6 Effect of low parameter-values

Parameter	Effect	Attenuation/ Amplification
Functional concentration	Decreased variability	Attenuation
Differentiation of operational transformation	Decreased number of relations	Attenuation
Specialization of operation transformation	Decreased number of relations	Attenuation
Separation	Decreased number of relations, increased regulatory potential	Attenuation Amplification
Differentiation of regulatory transformation (strategic / design / operational)	Decreased number of relations, increased regulatory potential	Attenuation Amplification
Differentiation of regulatory transformation (monitoring / assessment / intervention)	Decreased number of relations, increased regulatory potential	Attenuation Amplification
Specialization of regulatory transformation	Decreased number of relations, increased regulatory potential	Attenuation Amplification

Quality of Organization

1. Flexibility will increase because in low parameter structures stock can be reduced significantly and product-cycle times can be increased.
2. Control over the process is increased because of the emergence of a simpler operational structure and the integration of operational tasks and regulatory tasks – leading to a more reliable production process.
3. Low parameter structures create better chances for product and process innovation. Both type of innovation require an overview over larger parts of the process (including an idea of what the client wants). It also requires the possibility to experiment and learn from experimentation. Low-parameter structures provide this overview and can build the experimental requirements into tasks.

Quality of Work

1. Low parameter structures create chances for reducing stress. Stress, as a function of the ability to deal with disturbances adequately, is reduced when relevant regulatory potential increases.
2. Since, in low parameter structures, operational and regulatory tasks are integrated and cover a considerable part of the production process opportunities for work-related learning are created. One is able to experiment with regulatory actions, see their effect and adjust actions accordingly. Working on complex tasks (as compared to high parameter structures) is more demanding for workers, but at the same time more challenging, offering opportunities to develop.
3. Because of the inclusion of relevant external regulation as part of working in task groups, involvement, both socially and intrinsically seems to be secured.

Quality of Working Relations

Working on complex tasks in task groups, equipped with an overview over a large part of the process and with the regulatory potential to deal with disturbances enables *relevant* work-related communication. Moreover, by defining cohesive segments tied to task groups, the quality of internal communication is bound to improve.

Limits to Lowering Parameter-values

In the structures we have discussed in this section all parameters should be decreased as much as possible. However, as de Sitters admits, there are limits to lowering their values and there are circumstances in which decreasing parameters below some point is not feasible. In this last subsection on low-parameter structures we list some of the criteria indicating such limits.

A first class of criteria has to do with the feasibility of defining parallel production flows. For instance, the defined flows should have a considerable degree of capacity utilization – de Sitter thinks of 80%. Below this percentage, it might not be worthwhile to add the flow. Another criterion for defining flows has to do with the number of shared recourses. If this number increases, the autonomy of the flow may be threatened too much. A third criterion has to do with the costs of building extra flows. Sometimes it may just be too expensive.

Another class of criteria pertains to the complexity of tasks, emerging in low-parameter-value organizations. Tasks may become so complex or demanding, that it may turn out to be hard to find qualified workers for them. And, even if one is able to find them, they are more expensive than those working at less complex tasks. Moreover, training workers to fit the new task-requirements may be more expensive and demanding.

Moreover, some activities cannot be integrated, because of their intrinsic complexity or nature. For some complex tasks specialization is required because otherwise, their performance could be threatened. It does not, for instance, seem to be a good idea to include all kinds of monitoring and nursery activities into the task of a brain-surgeon who needs to be focused on the operation. In a similar vein, the principle of segregation of duties imposes limits to integrating tasks: some activities cannot be integrated into one task, for the sake of financial control or accountability.

More criteria may be given, but it is beyond the scope of this book to go into them in much detail. A final remark we would like to offer is that using cost or capacity related criteria is difficult, for it requires a trade-off between the advantages of lowering parameters (in terms of improvement of the quality variables) and the disadvantages in terms of costs or loss of capacity – a trade-off for which no formulas have yet been designed.

7.4 De Sitter's Organizational Structures and Conducting Experiments

The goal of this book is to describe organizations as social systems conducting experiments and to formulate principles for their design. In part I of the book we used Ashby and Luhmann to describe organizations, thus uncovering their experimental and social *arche*. In this part we formulate principles for designing organizational infrastructures to increase the likelihood of "experimental success." To this end, we started off with Beer's Viable System Model, a model providing us with functional requirements for organizational infrastructures. That is, it delivers a set of desired effects of infrastructures – in terms of five interrelated functions. The current chapter moves beyond Beer, and states de Sitter's principles for designing a specific part of the infrastructure: the division of work.

In fact, it is our aim in this chapter to show that de Sitter's theory can be used to design a division of work that creates chances for conducting experiments with meaningful survival. To this end, we introduced several relevant concepts from his theory and showed how they were related to the design of the division of work (or, organizational structures). In particular, we discussed so-called "high" and "low parameter structures" and discussed their impact on relevant organizational variables (quality of organization, work, and working relations). According to de Sitter, low parameter structures are "cybernetically speaking" favorable structures since they attenuate disturbances for the three classes of essential variables (by design) and the required regulatory potential is built into them.

While discussing de Sitter's theory in the previous sections, it may have already become apparent that his low parameter structures not only benefit the realization of the essential organizational variables de Sitter has in mind; they also create chances for conducting experiments with meaningful survival. In this last section, we want to emphasize this effect of low parameter structures. In order to do this we start with a brief discussion of the relation of de Sitter's theory to the theories we used to describe organizations as social systems conducting experiments (i.e., Ashby's and Luhmann's) and to Beer's VSM. Next, we go into the question how low parameter structures may benefit conducting experiments with meaningful survival.

7.4.1 Relating de Sitter to Ashby, Luhmann and Beer

In his theory on designing divisions of work, de Sitter applies Ashby's theory quite directly. To start with, de Sitter defines his own set of essential variables – the requirements "modern organizations" should meet in order to survive (i.e., the variables related to quality of organization, work, and working relations). By doing so, de Sitter performs Ashby's control-step for organizations in general.

Given these variables, a distinction can be made between disturbances, caused by the infrastructure and "general" disturbances; disturbances that are not caused by the infrastructure. According to de Sitter, designing a division of work should be directed at removing as much as possible the disturbances related to the infrastructure and equip the organization with the potential to deal with the remaining ones. In this way, de Sitter applies Ashby's two basic design heuristics: attenuating disturbances and amplifying regulatory potential. De Sitter points out that only low parameter structures really attenuate disturbances and have the required regulatory potential. These structures, according to de Sitter, enable control, design and operational regulatory activities. In all, it can be seen quite easily that de Sitter's theory is firmly rooted in Ashby's cybernetics. In particular the related notions of attenuation and amplification acted as heuristics for de Sitter to develop his design-principles.

It is less straightforward, however, how his theory relates to Luhmann's. To see that and how his design-principles are related to social systems theory it may be helpful to turn to Chap. 5, in which cybernetics and social systems theory were related. Since de Sitter's theory is a direct application of Ashby's cybernetics to designing divisions of work, it can be linked to social systems theory by means of the argumentation given in that chapter. Luhmann defines organizations as social systems consisting of "decisions producing decisions." Moreover, the production of decisions by decisions is guided by the organization's structure: decision premises. Now, any division of work is the result of a process of decision-making in organizations, and can serve as an infrastructural subject in further decision-making. One could also say that a division of work serves as a decision-premise, since it conditions decision-making in organizations. For example, it structures decision-pathways, it defines what can be expected of members in organizations and it determines who are to communicate with each other. However, these relations between divisions of work and social system theory hold for all kinds of organizational structures. They are not specific for the ones propagated by de Sitter. De Sitter's main contribution consists in principles leading to *specific* divisions of work: his low parameter structures. Although one might argue that these structures allow for swift and relevant communication and lead to a different coherence of decisions (and hence to a different instantiation of an organization as a social system) than high parameter structures, this moves beyond the way Luhmann's reasons about organizations as social systems, for (when defining organizations) he is not interested in specific instantiations of organizations.

As we described earlier, part II of the book deals with principles for designing organizational infrastructures. Beer's VSM was used to derive general principles; i.e., desired effects, infrastructures should meet in order to maintain their viability. Thus, Beer's main contribution is to deliver a functional model of infrastructures. It does not go into how these functions should be *realized by a specific infrastructure* – it does not answer the question what kind of infrastructure creates better chances for viability.

By contrast, de Sitter's does treat this question. It is his contribution to organizational cybernetics that, *given* the requirements for viability, he persistently and systematically directs the attention to the question of how to design organizational structures to meet these requirements. This question remains underexposed in the Viable System Model. Although it specifies the desired effects organizational structures should realize, and in this sense is useful for diagnosing and designing organizational structures, the Viable System Model does not solve the question of how we should go about when grouping transformations into tasks to (1) implement the functions required for viability, (2) attenuate disturbances, (3) and amplify regulatory potential.

In this sense de Sitter goes beyond the Viable System Model. Just as Beer, de Sitter is fascinated by organizational viability. Unlike Beer, de Sitter explicitly directs his attention to unfolding and solving the problem of building organizational structures contributing to realizing viability.

7.4.2 How Do Low Parameter Structures Benefit Organizational Experimentation?

It is our contention that of all possible structures de Sitter's low parameter structures create the best chances for experimenting with rich meaningful survival. To start with, they enable meaningful survival. Based on de Sitter's essential organizational variables as desired effects for many modern organizations, we hope to have shown that low parameter structures attenuate disturbances and amplify regulatory potential – more than other structures.

Moreover, low parameter structures enable experiments in organizations. In fact, they build the capacity to experiment and learn into structures as much as possible. They do so at different levels, i.e., by including non-routine and routine regulatory potential into individual or team-tasks, they allow for strategic, design and operational experimentation. For instance, by adding extra routine regulatory potential into (operational) tasks, the number of possibilities to deal with disturbances increases. In such tasks, workers are required to deal with disturbances as they see fit. This includes experimenting and learning about what works in particular circumstances. By adding non-routine regulatory potential, low parameter structures also include the potential to experiment with the infrastructure and even goals. In low parameter structures, innovation and strategy-formulation is not a matter of detached designers and strategy-makers, instead these structures allow for the participation of those who are involved in operational processes in experimenting with strategic and design issues.

By integrating routine and non-routine regulatory potential into operational tasks, and by creating an overview over the transformation process, low parameter structures have a built-in capacity for experimenting and learning. According to de Sitter, this is essential. Because disturbances cannot be organized away, it is best to equip tasks –at every level, be it individual or team – with the potential to deal with them; and this should include experimenting if one is confronted with new or unfamiliar (combinations of) disturbances. Building in the capacity to experiment at all levels is also essential for innovation: for instance, it allows for swift and relevant changes of the infrastructure governing operational transformations and their regulation by including those who are directly involved in it.

In fact, many of the advantages of low parameter structures, discussed in this chapter can be phrased in terms of their ability to allow and support experimentation at all levels. Conversely, many of the disadvantages of high parameter structures can be phrased in terms of not possessing this experimental ability.

A last point we address in this section is that low parameter structures create chances for *rich* meaningful survival. That is, they create chances for "developing our characteristically human capacities" in organizations. As we pointed out earlier, they allow for learning (skills) and involvement – both intrinsically and socially. Moreover, they enable members at different levels of the organization to see the point of their work, and, hence, to reflect on the (societal) goals the organization contributes to. Although much more can be said about the relation between organizational structures and rich meaningful survival, we are at this point, not yet equipped with the appropriate concepts to identify "rich" meaningful survival. In Chap. 10, we elaborate on the notion of rich survival, and in Chap. 11, we return to discussing its relation with organizational structures.

References

Beer, S. (1995). *The Heart of Enterprise*. Chichester: Wiley.

Christis, J. (1998). *Arbeid, Organisatie en Stress*. Amsterdam: Het Spinhuis.

Galbraith, J. (1973). *Designing complex organizations*. Reading: Addison Wesley.

Karasek, R. (1979). Job demands, job decision latitude and mental strain: Implications for job design. *Administrative Science Quarterly, 24*, 285–307.

Mintzberg, H. (1983). *Structures in fives: Designing effective organizations*. New Jersey: Prentice Hall.

Sitter, L.U. de (1973). A system-theorethical paradigm of social interaction: Towards a qualitative system dynamics. *Annals of System research, 3*, 109–140.

Sitter, L.U. de, (1994). Synergetisch Produceren. Assen: van Gorcum.

Sitter, L.U. de, & den Hertog, J. F. (1997). From complex organizations with simple jobs to simple organizations with complex jobs. *Human Relations, 50*(5), 497–534.

Sitter, L.U. de, den Hertog, J.F., & van Eijnatten, F.M. (1990). Simple organisations, complex jobs: The Dutch sociotechnical approach. In: *Proceedings of the American Academy of Management*.

Thompson, J. D. (1967). *Organizations in Action*. New York: McGraw-Hill.

Womack, P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world*. New York: McMillan.

Chapter 8
Epilogue to Part II: functional and specific design principles

8.1 Introduction

In the previous two chapters, we unfolded Beer's functional and de Sitter's specific design principles. As argued, these principles can be used to diagnose and design organizational (infra)structures supporting experiments with meaningful survival. This means that we have realized the objective set for Part II of the book. In this epilogue, we summarize these principles (8.2) and reflect on their status (8.3). In this reflection, we argue that the design principles are not contingent and risky, like the selections figuring in the experiment, but necessary and certain. Section 8.4 marks the transition to Part III of the book.

8.2 Summary of the Design Principles

To summarize the principles discussed in the previous chapters, we stick to the distinction between functional and specific design principles.

Functional design principles comprise sets of (related) effects that should be produced if a system is to be able to continue its risky experiments with its meaningful survival. Within the class of functional design principles, we distinguish principles specifying: (1) what is desired for effective control in general and (2) effects a system should attain to adapt and realize the goals of its primary activities.

Specific design principles specify which system-specific infrastructural parameters are relevant to realize desired effects specified by the functional design principles and what values these parameters should have if these effects are to be realized. For organizations, there are three relevant classes of parameters: (1) the organizational structure, (2) HR-systems, and (3) technical systems. Because the organizational structure provides the foundation for the other two classes of parameters, we only explored the relevant parameters in this class and their required values.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_8, © Springer-Verlag Berlin Heidelberg 2010

8.2.1 Functional Design Principles

Regarding the functional design principles, we in the first place, summarize those related to effective control in general. These principles comprise Ashby's functional description of regulation and its clues for the design of effective and efficient regulators (Table 8.1).

Table 8.1 Functional design principles related to effective control

Effective control in general
"an essential feature of the good regulator is that it blocks the flow of variety from disturbances to essential variables" (Ashby, 1958, p. 201)
Attenuation: decreasing disturbances as much as possible
Amplification: increasing regulatory potential as much as needed
Control: determining goals
Design: specifying and implementing the infrastructure needed for regulation and performing transformations
Operational regulation: : regulating transformations
Performing transformations: realizing goals

In the second place, the class of functional design principles comprises sets of effects that should be produced if a system is to be able to conduct its experiments with meaningful survival (see Table 8.2).

Table 8.2 Functions supporting ongoing experiments with meaningful survival

Functions needed to conduct experiments with meaningful survival:	
Adaptation of goals	Realization of goals
Policy: setting goals and regulating the discussion between intelligence and control	Control: regulating the cohesion and synergy of the primary functions by: direct commands and reports, resource bargaining, audits, and monitoring coordination
Intelligence: scanning the environment and contributing to the formulation of goals	Coordination: supporting the primary activities by reducing oscillations
Control: contributing to the formulation of goals	Primary activities: realizing the raison d'être of the system

8.2.2 Specific Design Principles

Specific design principles specify relevant system-specific infrastructural parameters and their desired values needed to realize the desired effects specified by the functional design principles. These design principles are "specific," because in different types of systems, different classes of parameters may be relevant. For instance, in multi-cellular organisms or in a societal subsystem like the economy other parameters are relevant than in organizations.

As explained, in the case of organizations, relevant classes of parameters are the organizational structure, HR-systems, and technical systems. De Sitter is one of few

Table 8.3 De Sitter's parameters of the organizational structure and their desired values

Parameters relevant to the organizational structure		Desired value
Parameters relevant to the production structure	1. Level of functional concentration relative to orders 2. Level of differentiation of operational transformations in tasks 3. Level of specialization of operational transformations in tasks	As low as (is economically and technically) possible
Parameter relevant to the relation between the production and control structure	4. Level of separation of operational and regulatory transformations in tasks	
Parameters relevant to the control structure	5. Level of differentiation into aspects of regulatory transformations in tasks 6. Level of differentiation into parts of regulatory transformations in tasks 7. Level of specialization of regulatory transformation in tasks	

authors who both explicitly defines the parameters of the organizational structure and specifies what their value should be in order to realize the functional design principles. De Sitter's specific design principles for organizational structures can be summarized as follows (see Table 8.3).

De Sitter argues that if an organizational structure – described in terms of the abovementioned parameters – satisfies the desired values, it realizes a high:

1. *Quality of organization*, optimally enabling the organization to adapt and realize its primary activities by means of effective and efficient control, design, operational regulation
2. *Quality of work*, creating potentials for worker involvement, work-related learning and reducing the risk of work-related stress
3. *Quality of working relations*, providing conditions for effective communication.

8.3 The Status of the Design Principles

In this book, we conceive of organizations as social systems conducting risky experiments. Selected goals, infrastructural designs, regulatory actions, and primary operations are the "objects" in this experiment. They are contingent and risky hypotheses about what should be done in order to survive.

However, according to Beer, the functional principles are necessary and sufficient. Against this, it could be argued that because they figure in the experiment, they are as contingent and risky as the other "objects." Yet, if this were the case, why devote much attention to them? They would be just as contingent and risky as the "objects" involved in the experiment, and might be outdated tomorrow.

To establish whether the design principles are just a contingent and risky temporary fad or something more, we need to reflect on their status.

To gain insight in the status of the design principles it is useful to specify *how* they are involved in organizational experiments with meaningful survival. On closer examination, two relevant modes of involvement can be distinguished: (1) they are principles ***of*** these experiments and (2) they may be design principles figuring ***in*** these experiments.

8.3.1 Principles of Experiments with Meaningful Survival

The main argument until now can be summarized as follows. Organizations have an experimental and a social "arche." Zooming in on the experimental "arche," we argued that "meaningful survival" is what is at stake and the selections about "objects" involved in the experiments are contingent and risky hypotheses about what should be done in order to survive. Making these selections requires the realization of particular desired effects. Beer's functional design principles specify these effects.

To establish the status of the functional design principles, we depart from the supposition that, in our culture and society, historically contingent phenomena called "organization" exist. Based on reflection, we can say that organizations have an experimental "arche" that cannot be negated of them without negating organizations altogether. That is, if we grant that there are organizations, it can be argued with certainty that this "experimental" feature is a necessary feature of them.

The functional design principles, in turn, specify in terms of desired effects what is implied in the experimental "arche." They result from a reflection on what is necessarily involved in conducting experiments with meaningful survival. As such, the functional design principles specify necessary principles *of* all possible experiments with meaningful survival. They specify with certainty the necessary conditions of their possibility.

This line of reasoning implies that if a particular organization stops to realize one or more of these effects, it loses its potential to conduct experiments with meaningful survival. If this happens, its experimental "arche" is destroyed, destroying the phenomenon called "organization" with it. For this reason, *all* existing organizations realize the desired effects specified by the functional design principles.

However, *not all* organizations do this at the same level. *Some* organizations realize these effects better than others. Using Beer's terminology, one can say that all existing organizations are viable, i.e., they all realize the desired effects specified by the functional design principles, but some of them are more viable than others.

Therefore, the question becomes, what is it that makes one organization more and other organizations less viable? This is where the specific design principles come in.

We know that whether, and if so, how well or how badly, a system realizes the effects specified by the functional design principles – i.e., whether and how viable it is – depends on the design of its infrastructure. For an organization, this means that dependent on the design of its organizational structure, HR-systems, and technical systems, it is more or less able to conduct its experiments. The infrastructure "embodies" the organization's ability to conduct these experiments.

The specific design principles specify requirements to the organization's infrastructure if it is to realize the desired effects specified by functional design principles. In this sense, the specific design principles are (system-specific) principles of experiments with meaningful survival.

To the extent that these principles are deduced from either functional design principles or other "archai" of organizations (e.g., the social "arche"), they are necessary and certain as well. For instance, de Sitter deduces the parameters of the organizational structure and their required values from (1) decision premises involved in the social "arche" of organizations (goal program, task, communication pathway, etc.) and (2) Ashby's theory of regulation (his functional model, attenuation, and amplification). To this extent, his specific design principles can be called necessary and certain.

To conclude, functional design principles specify in terms of desired effects what is necessarily implied in all possible experiments with meaningful survival. They are necessary – and not contingent – principles that can be established with certainty by means of a reflection on the functional conditions of the possibility of the experiments constituting the experimental "arche." To the extent that specific design principles can be deduced from both the social "arche" and functional design principles, they share in the status of the functional design principles (see Fig. 8.1).

8.3.2 Design Principles Figuring in Experiments with Meaningful Survival

Above, we concluded that we have a set of principles specifying with certainty necessary conditions of all possible organizational experiments with meaningful survival. Based on Chaps. 6 and 7, we know that these principles can be used to guide selections with regard to the design of organizational infrastructures. The principles *of* all possible experiments with meaningful survival can thus figure as design principles *in* these experiments.

If this happens, necessary and certain principles are applied in concrete organizations to design concrete infrastructures. For instance, based on de Sitter's approach, we know that to improve the quality of organization, the level of functional concentration should be low. This principle holds for all organizations. However, in

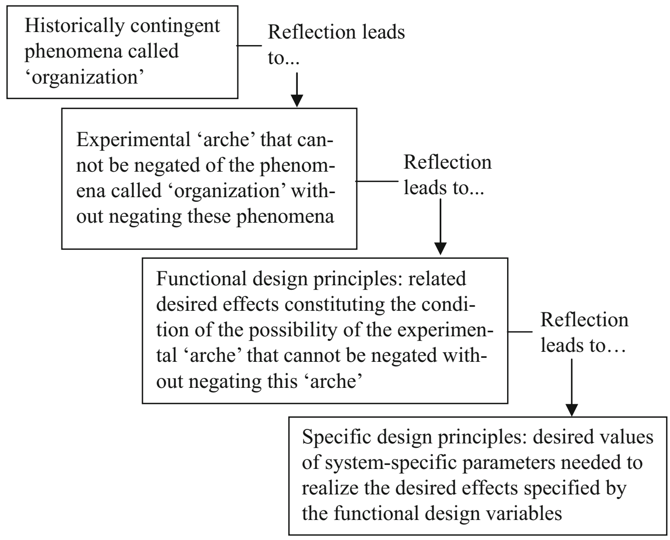


Fig. 8.1 From phenomena to necessary conditions

order to design the actual organizational structure of a particular hospital or energy company, this principle has to be applied to the particulars of these organizations.

Once design principles figure in the experiment, there seems to be a “gap” between them and the concrete situation to which they are applied. The design principles are necessary and hold for all possible organizational experiments. In this sense, they are *general*. The situations in which they are applied are concrete, i.e., they are applied to design this or that organization.

To “bridge” this gap between general principle and concrete situation, a “designer” seems to be needed. This designer (1) knows the design principles, (2) is sensitive to the particulars of concrete design problems, and (3) has the deliberative capacity and experience to apply these principles to the particulars of the problem. This designer is able to “judge” whether particular infrastructures are more or less able to realize the desired effects specified by the functional design principles.

Please note that these judgments are *not* necessary and certain; they do not *guarantee* success with necessity and certainty. Because they involve concrete, complex, and changing particulars, they are contingent and risky. However, because they also involve knowledge of design principles, experience, and deliberative capacity, and apply these to the particulars of the design problem, they can be said to be “better” and more “informed” than the judgments of those who do not have the aforementioned knowledge and competencies.

Designer’s judgments bridging the gap between the necessary design principles and the concrete particulars of a design problem do not guarantee success with necessity and certainty. They increase the probability of success; they increase the probability that the organization to which they are applied is able to experiment. For this reason, selections regarding the organization's infrastructure, even if they are based on necessary and certain design principles, remain contingent and risky as was argued in Chap. 1.

8.4 Transition to Part III: Poor and Rich Survival

In Chaps. 6 and 7, we discussed design principles that, if they are applied by able designers, support organizational experiments with meaningful survival. However, in Chap. 1, we distinguished two modalities of meaningful survival: poor and rich survival.

We defined "poor" survival as maintaining a separate meaningful existence by selecting *whatever* goals, for *whatever* reason, and realizing these goals in *whatever* way. In other words, poor survival means that as long as organizational goals and means to realize them meet whatever criteria considered meaningful for whatever reason, meaningfulness is ensured, whatever the possible detrimental consequences for organizational members or the society the organization is a part of.

Rich survival was defined as maintaining a separate and meaningful existence by selecting and realizing goals that contribute to the creation of societal conditions enabling human beings to develop and realize their humanity. Rich survival makes organizational contributions to society into a serious issue. It is about organizations contributing to the improvement of society by improving conditions for its citizens to flourish.

Developing a rich concept of meaningful organizational survival is especially relevant because of the type of society we live in. In spite of all kinds of other characterizations (e.g., network society, risk society, world society, information society), our society can still be called an "organization society." By this we mean that it is a society in which organizing and organizations are vital for realizing societal values such as "health," "justice," "education," "safety," in fact, for a lot of the things we both value and take for granted in our daily lives. At the same time, we know that organizations, their functioning and its consequences, constitute a threat to all of the aforementioned values: they also contribute to huge societal problems, undermining conditions enabling us to live a fulfilled life. And, finally, to counter these problems, we still need the complexity-resolving potential of organizing and organizations.

Therefore, to attenuate the probability of organizations causing societal problems in the first place and to amplify the organizational potential to cope with societal problems, it seems to be required that organizations learn to act in a socially responsible way, contributing to societal conditions enabling human beings to live a meaningful human life. This is why we think it is worthwhile to develop the idea of organizations as social systems conducting experiments with *rich* meaningful survival, for contributing to a society enabling its members to live a fulfilled life is exactly what "rich meaningful survival" means.

This importance of socially responsible behavior seems to be increasingly acknowledged in society, in organizations (albeit sometimes reluctantly), and in management literature. Forced by public opinion, important stakeholders, or by law, organizations introduce "codes of conduct" (as instances of compliance-based approaches to business ethics) and/or encourage the development of "professional virtues" (as instances of integrity-based approaches). Here, we do not go into the question whether "codes" and "virtues" are a sign of real concern or only are there to placate important stakeholders. We just want to point at the growing acknowledgement of the importance of improving the societal record of organizations.

However, discussions and literature about "codes" and "virtues" seem to be limited to what is wanted of organizational members and how this contributes to socially responsible behavior. They pay less attention to infrastructural conditions needed to enable rich survival.

In Part III of the book, we want to address this question. We not only want to know what rich survival entails, we also want to know which infrastructural conditions can enable it. Given our focus in this book on organizational structure, we limit this question to the organizational structures that can support rich survival.

This means that it is our purpose in Part III of the book to unfold what we mean by rich survival and explain which organizational structures can support it. To this purpose, we organized this part into three chapters.

In Chap. 9, we start by presenting a vivid example of *poor* survival. We do this, to create an awareness of both the destructive potentials of organizations and the cybernetic and social systemic principles underpinning them. Given an understanding of poor survival, we develop a counter-model to it (Chap. 10). Central to this counter-model, i.e., central to the model of rich survival, is the idea of human beings "living a fulfilled human life." To explore what this idea entails, we turn to Aristotle's ethics. Finally, in Chap. 11 we discuss specific principles for the design of organizational structures supporting rich survival.

Chapter 7
Specific Design Principles: de Sitter's Organizational Structures

7.1 Introduction

In the previous chapter, we discussed Beer's Viable System Model; a functional model specifying desired effects required for viability. These effects can be used as criteria for diagnostic or design purposes. We also pointed at limitations of the Viable System Model. As a *functional* model it does not address the question of the *embodiment* of functions. Although it specifies desired effects, it does not positively address the question of how to design their realization. Simply put, the strength of the Viable System Model is stating *what* effects should be realized, not *how* they should be realized. For instance, functions three and four should engage in a relatively complex and balanced dialogue about plans for innovation, but what is needed to realize this dialogue? How should one distribute tasks and responsibilities among organizational members, so that this dialogue can be carried out properly? How should one select, allocate, and train the people involved in these dialogues, and how does one design the technological infrastructure supporting the complex communication processes required for innovation? In short: how does one design the infrastructure realizing the desired effects for viability?

In this chapter, we focus on the question how to build an infrastructure realizing the functions mentioned by Beer and proceed our inquiry into conditions for experiments with organizational survival, As we discussed in previous chapters, the infrastructure consists of three related parts – division of work, HR-systems and technology. In the present chapter, we focus on the division of work as the most basic and relevant part of the infrastructure. As a consequence, we narrow the question "which infrastructures realize the VSM-functions and increase the organization's potential to conduct experiments" down to "which divisions of work realize the VSM-functions and increase the organization's experimental potential." In this way we focus on what we have called *specific* design principles (Chap. 1), i.e., principles tied to designing a particular part of the infrastructure.

From our discussion of the VSM, we know that there is not a one-to-one relation between functions and tasks. Different functions may be concentrated into one task.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_7, © Springer-Verlag Berlin Heidelberg 2010

For instance, someone running a one-man business should perform all five functions. Moreover, many tasks can be involved in realizing the same function. For instance, realizing the intelligence function may involve someone with the task of gathering data, another with the task of analyzing it, and yet another with the task of processing the analyzed data. Given this relation between functions and tasks, it becomes relevant to ask how to distribute functions into a network of related tasks. In textbooks, we can find this question under the heading of the *design of the distribution of work*.

Everyone who has ever worked in an organization is aware of the importance of the design of the distribution of work. "Doing one's job" relates directly to this design, for the breakdown of work and its synthesis into "tasks" define what "one's job" is. A job is a part of a larger network of jobs, and dependent on the layout of this network, "doing one's job" relates to others "doing their jobs." Dependent on one's job and its place in the network, dependency relations between jobs exist that influence the efficiency and effectiveness of one's job and the quality of one's working life.

This awareness of the importance of the distribution of work is particularly alive in professional or machine bureaucracies struggling to cope with increasingly rapid developments in their environments. Yet, even in "lean," "post modern," "knowledge intensive," and "network" organizations where tasks may be diffuse and changing, workers engage in communication dependent on the (changing) distribution of work and experience its obstacles and benefits. Therefore, if organizations are to be viable, we have to take a closer look at how the distribution of work inhibits or enables social experiments.

To this purpose, we select the work of Ulbo de Sitter. De Sitter is the founder of the Modern Dutch Sociotechnical Approach, also called the Integral Organizational Design Approach. This approach explicitly and systematically uses cybernetics to formulate rules and principles for designing "viable" distributions of work. To discuss these rules and principles as well as their cybernetic origin, we choose as our point of departure de Sitter's book "Synergetisch Produceren" (1994) and his article "From complex organizations with simple jobs to simple organizations with complex jobs" (de Sitter and den Hertog, 1997).

We are well aware that de Sitter's work is almost only available in Dutch (exceptions are: de Sitter 1973; de Sitter, et al. 1990, 1997) and that, for this reason, it is not as known as that of, for instance, Thompson (1967); Galbraith (1973) or Mintzberg (1983). The reason for selecting De Sitter's work is that it is explicitly rooted in cybernetics, is far more detailed than that of other authors on organization design, and fits rather well into the overall intention of the second part of this book: formulating infrastructural conditions for organizations (as social systems) to conduct experiments.

Particularly in "Synergetisch Produceren," de Sitter specifies how a designer should distribute work to keep organizations viable. Inspired by cybernetics in general and Ashby's theory in particular, he spells out how to design distributions of work (1) attenuating disturbances and (2) amplifying regulatory potential to deal with disturbances impinging on "relevant organizational variables."

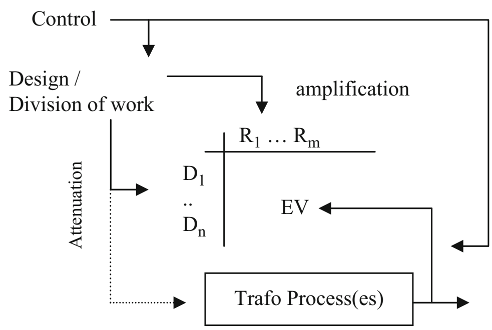


Fig. 7.1 Overview of Ashby's regulatory logic and the role of divisions of work

In fact, de Sitter's theory may be regarded as a re-specification of Ashby's regulatory logic in the realm of distributing work. As the reader may recall from Chap. 2, this logic contained the activities control, design and operational regulation with respect to (a) transformation process(es). Control meant setting the target, i.e., specifying the essential variables for the transformation process and setting their norm-values. The main goal of design was to arrive at a "mechanism" embodying an adequate relation between essential variables, disturbances and regulatory actions. Design entailed "constructing a mechanism" contributing to attenuation (decreasing the chance of disturbances) and amplification (increasing the number of regulatory actions).

Based on Fig. 7.1 we can phrase de Sitter's contribution in terms of Ashby's theory as formulating *principles for designing divisions of work* (as part of the infrastructure). In a nutshell, these principles indicate that a division of work should attenuate as much disturbances as possible and build regulatory potential (amplification) with respect to organizational processes into an infrastructure. If applied, these principles result in a part of the infrastructure (a division of work) based on which it is possible to realize control, design, operational regulation and transformation processes (see also Sect. 1.2).

To understand de Sitter's design principles, i.e., to see that and how a division of work, as part of the organizational infrastructure, can attenuate or amplify, one should translate all aspects of Ashby's regulatory logic to the case of organizations. Ashby's theory does not give this translation yet, since his theory is about regulating all kinds of systems. De Sitter, however, does. In particular, he tries to clarify (1) the relevant *organizational* essential variables, (2) the nature of disturbances in an organizational context, and (3) what regulation in the case of organizations means. Given this translation of relevant concepts to an organizational context, de Sitter systematically tries to show (4) how a division of work can attenuate disturbances and how regulatory potential can be amplified by it.

De Sitter not only gives an answer to these four questions, he also specifies principles for *arriving* at a cybernetically sound division of work (one that attenuates disturbances and amplifies regulatory potential).

In this chapter we will follow in de Sitter's footsteps and eventually hope to show how particular divisions of work (from now on we will refer to them as organizational structures) may support the organizational experiment. In Sect. 7.2, we start with discussing the desired effects of organizational structures: attenuation en amplification. In order to discuss these effects, we will discuss de Sitter's "translations" of Ashby's relevant regulatory concepts into the realm of organizations. In Sect. 7.3 we discuss which organizational structures (according to de Sitter) are "cybernetically sound structures," i.e., structures that actually attenuate and amplify. While discussing them, we present de Sitter's principles for designing organizational structures. In the last section we deal with the question of "adequate organizational structures for conducting experiments."

7.2 Designing Organizational Structures: Aiming at Attenuation and Amplification

In the previous section we stated that organizational structures should have the capacity to attenuate disturbances and amplify regulatory potential to deal with them. In fact, referring to Ashby, one could specify a regulatory table regarding these characteristics of organizational structures – see Fig. 7.2.

To understand how organizational structures can attenuate and amplify, we need to clarify:

1. Organizational structures
2. Relevant organizational variables
3. The nature of the disturbance impinging on these variables
4. Regulatory potential in the context of organizations
5. How organizations can attenuate and amplify

Each of the next sub-sections is devoted to one of the above issues.

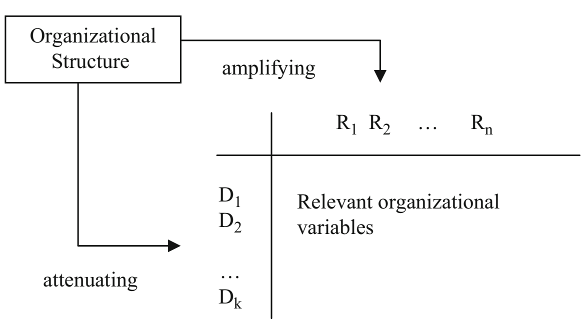


Fig. 7.2 Designing organizational structures cast in Ashby's terms (part I)

7.2.1 *Organizational Structures*

Roughly speaking, an organizational structure (or distribution or division of work) is a network of related tasks. To understand, in some more detail, what "a network of related tasks" is, we will discuss tasks in organizations (in particular, we will go into their definition and introduce two relevant types of organizational tasks). Second, we will discuss how tasks can be related into an organizational structure.

7.2.1.1 Tasks in Organizations I: Defining Tasks

To define tasks in organizations we use Ashby's concept of transformation. As we discussed in Chap. 2, a transformation describes a change of values of a set of variables from a begin state (the operand) to an end state (the transform) – see Table 7.1.

Table 7.1 A transformation

T↓	Begin state (operand; set of values of the variables at time = t)
	End state (transform; set of values of the variables at time = t + 1)

In the context of discussing organizational structures it is more convenient to represent a transformation slightly differently – see Fig. 7.3:

Begin state → Transformation process → End state

Fig. 7.3 Representation of a transformation

The representation in Fig. 7.3 highlights that there is "something" that causes the begin state to change into the end state: the transformation process. Emphasizing this process supports our explanation of tasks in organizations – as will become apparent. For now, a task is associated with "realizing the end state by performing a particular transformation process."

Transformations can be decomposed into sub-transformations. For instance, the transformation "doing the dishes" may be decomposed into four sub-transformations: "collecting"; "cleaning," "drying," and "storing." And, each of these sub-transformations can become the task of a separate individual.

There are at least two ways to decompose transformations into sub-transformations: (1) into "*parts*" and (2) into "*aspects*." In actual practice, these two types of decomposition may, and often do, occur simultaneously.

To explain both types of decomposition, take the example of Energeco, the company providing ecological energy introduced in the previous chapter. Its transformation consists of a begin state: consumers who are not provided with eco-energy (but want to be) and an end state or desired effect: consumers who are provided with eco-energy (see Fig. 7.4).

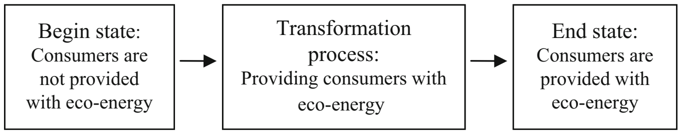


Fig. 7.4 Energeco's main transformation

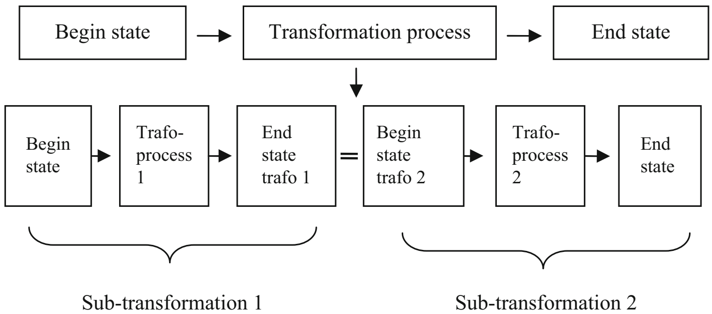


Fig. 7.5 Two sub-transformations resulting from a decomposition into parts

Decomposition into *parts* involves defining desired effects *between* the begin state and end state of the whole transformation. Fig. 7.5 shows two sub-transformations resulting from decomposition of a transformation into parts. Note, that both sub-transformations have a begin state and end state. The two sub-transformations are connected because the end-state of one transformation is the begin-state of the next one.

In the case of "Energeco," it is possible to decompose "providing eco-energy" into parts by defining three sub-transformations: "generating," "storing," and "diffusing" ecological energy. Together, these sub-transformations constitute the process of providing eco-energy to consumers.

Decomposition into *aspects* involves defining one or more characteristics (aspects) of the whole transformation and using them to define sub-transformations. For instance, it is possible to decompose "providing eco-energy to consumers" into "providing solar energy," "providing wind energy," and "providing water energy" to consumers. The characteristic, used to define sub-transformations is "type of energy provided." In this case, the decomposition does not define desired effects *between* the begin state and end state of the whole transformation, but identifies different *aspects* of the *whole* transformation (which, in this case results in three more or less parallel sub-transformations).[1]

[1]It should be remarked that the resulting sub-transformations may differ in degree of "autonomy." Sometimes, they can be treated as autonomous "flows" of production (which, as Beer (1995) suggests, may even be "hived off" from the original organization), and in other cases, one aspectual sub-transformation may be tied intimately to the other. This is the case with defining regulatory and operational sub-transformations (see explanation below).

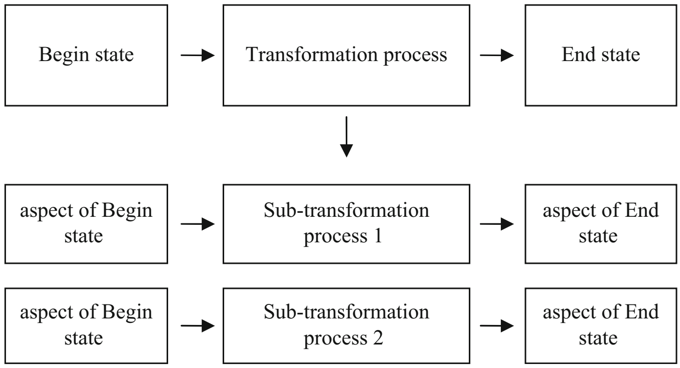


Fig. 7.6 Two sub-transformations resulting from a decomposition into aspects

Figure 7.6 shows two sub-transformations resulting from a decomposition of a transformation into aspects. Note, that the appearing sub-transformations consist of an aspect of both the original begin and end state, and a sub-transformation process connecting these two states.

Given some main organizational transformation, designers can apply all kinds of criteria to decompose it into parts or aspects. In the Energeco example, the source of the energy (sun, water, wind) functions as a criterion for a decomposition into aspects. However, it is also possible to choose other criteria, such as types of users (households, companies) or geographical areas (regions north and south) – see, for instance, Mintzberg (1983), or de Sitter (1994) about the nature of such criteria.

Both ways of decomposing transformations can be applied repeatedly and interchangeably. Every sub-transformation, resulting from decomposition can itself be decomposed into parts or aspects, resulting in new sub-transformations, and so on. Repeating this process of decomposition results, eventually, in a set of sub-transformations that can be regarded as a set of building blocks for "assembling *tasks*."

We can now define a *task* as a specific grouping of (sub-) transformations. From a set of sub-transformations resulting from different decomposition steps, many possible tasks can be assembled by set specific sub-transformations together. A criterion for defining a specific grouping of sub-transformations into a task is that such a grouping can be assigned to some "operational unit" – someone or something able to realize the set of sub-transformations. For example, the transformation "doing the dishes" can be decomposed into parts: e.g., "sorting out," "cleaning," "drying" and "storing." These four sub-transformations can be grouped into different tasks and assigned to different individuals. For instance, task 1 may comprise the sub-transformations "sorting out," "drying" and "storing" and may be assigned to one individual. In this case, task 2 (assigned to someone else) only contains the sub-transformation "cleaning."

Tasks can also be assigned to operational units other than individuals. In the case of Energeco, the decomposition into parts resulted in the sub-transformations, "generation," "storage," and "diffusion." Based on these three transformations, two tasks may be specified: one task in which generating and storing energy are grouped together and another task containing the sub-transformation diffusing energy. These tasks can be assigned to two different organizational units.

The above two examples make apparent that it is possible to specify tasks at different levels of aggregation in organizations – depending on the "operational unit" to which they are assigned. At the lowest level, we find the "job" of an individual member of the organization. This job consists of a set of transformations to be realized by that member. Jobs fit into the task of larger organizational units such as, for instance, teams, departments, or business units. Finally, the tasks of these units fit within the task of the organization as a whole. At this level the task reflects the organization's main transformation.

7.2.1.2 Tasks in Organizations II: Operational and Regulatory Transformations

In the previous section we used Ashby's notion of transformation and introduced possible ways to decompose transformations in order to define an organizational task. One may argue, at this point, that the number possible sub-transformations resulting from decomposing the organization's main transformation can be rather large, and, as a result, so is the number of possible groupings of sub-transformations into tasks. Moreover, since organizational structures are networks of related tasks, a large variety of tasks also entails a large variety of organizational structures.

To come to terms with this variety it is relevant to discuss two types of tasks: operational and regulatory tasks. Based on their description, it becomes possible to see that some groupings of sub-transformations into tasks, and some networks of related tasks are more suitable than others. According to de Sitter, the type of transformation built into a task (and hence, into the network connecting them) condition both the probability of the occurrence of disturbances and the regulatory potential for dealing with them. For instance, the potential to regulate parts of my work myself depends on whether my task comprises the required "regulatory potential." In such a case, the effectiveness of my task (with respect to dealing with disturbances) depends on the inclusion of a specific type of transformation (a "regulatory transformation") into my task.

In this section, then, we discuss two relevant types of sub-transformations, influencing the effectiveness of (related) tasks: operational and regulatory transformations. Moreover, after their introduction, we will elaborate regulatory transformations.

The Operational and Regulatory Aspect of Transformations

To understand what operational and regulatory transformations are, it is relevant to start with distinguishing an operational and a regulatory *aspect* of a transformation. The *operational* aspect of a transformation concerns the *realization* of its desired effect (its desired end state). In the case of "doing the dishes" it means performing the activities (actually sorting out, cleaning, drying and storing dishes) leading to the end state "dishes done & stored." In the case of Energeco it has to do with performing all operational activities involved in actually realizing providing eco-energy (e.g., actually generating; storing and delivering the energy). The *regulatory* aspect of a transformation concerns "dealing with the (virtual and actual) disturbances the operational sub-transformation faces in realizing the desired output of the transformation" (see de Sitter 1994, p. 10). Formulated in terms of Ashby, de Sitter's regulatory aspect consists of Ashby's control, design and operational regulation.

So, as de Sitter sees it, a transformation contains an operational aspect (realizing the end state of the transformation) and a regulatory aspect (i.e., control, design and operational regulation with respect to the operational aspect of the transformation – see Fig. 7.7).

If these aspects are used to define sub-transformations, the resulting sub-transformations are called "operational" and "regulatory" transformations. Moreover, they can be grouped together into tasks, containing more or less regulatory or operational transformations.[2]

De Sitter points at a "natural way" of thinking about operational and regulatory transformations, related to the difference between the operational "primary" processes in organizations and their regulation. De Sitter refers to the "primary" transformation of an organization – its "*raison d'être*" (for instance, providing

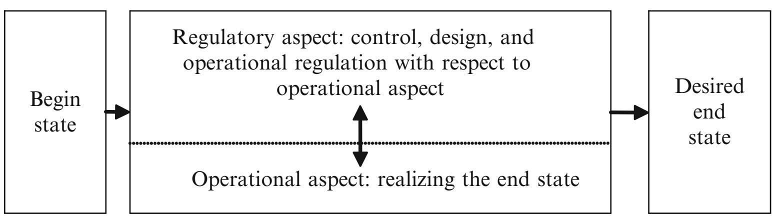


Fig.7.7 The operational and regulatory aspect of a transformation

[2]It may be noted that regulatory and operational sub-transformations are "relative" in two senses. First, they are relative to the selection of a designer modeling a particular transformation. Since operational and regulatory transformations are transformations too, they can also be decomposed regarding their operational and a regulatory aspect. Therefore, what appears to one designer as a regulatory or operational transformation may be selected by another designer as a transformation in focus and decomposed into an operational and a regulatory transformation. This entails that all transformations (including operational and regulatory sub-transformations) always contain an operational and a regulatory aspect. Second, operational and regulatory transformations are also relative to each other. Realizing an operational sub-transformation presupposes a regulatory sub-transformation, otherwise, disturbances cannot be dealt with. Moreover, realizing the regulatory sub-transformation presupposes the operational sub-transformation as its object – it deals with disturbances *of* some operational transformation.

eco-energy, in the Energeco case) as a "root" (consisting of many operational transformations), relative to which it is possible to define regulatory transformations. De Sitter illustrates this relation by means of a "procedure" to define what he calls the production structure of an organization. This "procedure" entails listing all activities, necessary to realize the primary organizational process (for instance: purchase, sales, planning, sawing wood into pieces, drilling holes into pieces of wood, assembling pieces of wood into a chair; maintaining drilling machines, etc.). These activities serve as the background for defining and coupling (operational transformations into) tasks, in which the realization of the primary organizational process is central. The resulting network of tasks can be said to cover the operational aspect of the whole organizational transformation: the *production structure*, as de Sitter calls it.

The tasks in this production structure, as well as groups of tasks have to be regulated. This means that, relative to the network of operational tasks (realizing the primary organizational process), a network of tasks dedicated to dealing with the disturbances in the production structure should be identified. De Sitter calls this network of regulatory tasks the *control structure*.

Four types of Regulatory Transformations

Above, we introduced the distinction between regulatory and operational transformations. To design a network of tasks, de Sitter argues, it is relevant to further differentiate of regulatory transformations in organizations. He holds that two distinctions are important: internal versus external regulation and routine versus non-routine regulation, resulting in four types of regulatory sub-transformations.

To explain these distinctions and the resulting four types of regulation, it is helpful to keep in mind that a task can be seen as a transformation – i.e., a set of coupled sub-transformations with an output, an operational aspect realizing this output, and a regulatory aspect dealing with the disturbances impinging on the realization of the output. Moreover, a task is related to other tasks in a network of tasks. To illustrate, while performing my (operational) task, I may have a relation with:

- A task delivering semi-finished products I have to work on, in order to produce the required output (i.e., a previous task in the operational process)
- A task to which I forward my output (the next task in the operational process)
- A task responsible for the tools I have to work with to produce the required output
- A task from which I receive a planning with regard to what I am supposed to do the coming period
- A task responsible for maintaining the tools I work with
- A task making sure that my knowledge is up-to-date
- Other tasks

Dealing with disturbances affecting the task's operational aspect can be done internally and/or externally. Dealing with a specific disturbance *internally* means that the necessary regulatory activities to deal with it are *part* of the task. Dealing with a specific disturbance *externally*, means that the task needs to involve its environment (e.g., involve other tasks in the network of tasks, or the environment of the organization).

The second distinction introduces the difference between dealing with disturbances in which the task and the network of tasks (the specific coupling of sub-transformations into related tasks) remain *unchanged* (routine regulation) and dealing with disturbances in which the task or network of tasks *changes* (non-routine regulation).

These two distinctions define four classes of regulatory activities:

1. Internal routine regulation (the regulatory activity does not involve other tasks and does not change the task or network of tasks)
2. External routine regulation (the regulatory activity involves other tasks but does not change the task itself or the network of tasks)
3. Internal non-routine regulation (the regulatory activity does not involve other tasks but changes the task's infrastructure and/or adds routine regulatory potential)
4. External non-routine regulation (the regulatory activity either changes the task's essential variables and norms or changes the infrastructure of the task's environment).

To illustrate internal and external routine regulation, imagine a worker in a factory producing wooden chairs, whose task it is to paint pieces of wood before passing them on to someone who assembles them into a chair. A disturbance for this task may be that the surface of the wooden part, as it was passed on to the worker by a previous worker (who has to saw and sand the parts), is not entirely smooth. Internal routine regulation might consist of taking a piece of sandpaper to sand the wooden part. To be an example of *internal* routine regulation it is required that this activity is part of the task. It is an example of internal *routine* regulation because no task in the network and no relation between them is changed.

An example of external routine regulation would be to go to the previous worker and ask him to sand the part again. This regulatory activity may be called *routine* if it is a part of the regulatory repertoire of the task, and if it does not involve changing the task itself nor the network of tasks. It may be called *external* because, to deal with the disturbance, the task's environment (another task in the network) is needed.

To discuss non-routine regulation, it is useful to cast what has been said thus far in terms of Ashby's regulation-table (see Fig. 7.8). The "essential task variables" in the figure refer to the output of the operational sub-transformation of the task. These variables can be subject to (virtual and actual) disturbances ($D_1 \ldots D_k$) and these disturbances may be dealt with by means of internal and external routine regulatory actions ($R_1 \ldots R_n$) that are part of the task.

In order to understand non-routine regulation, a distinction can be made between two classes of disturbances: disturbances arising "within" the task ($D_{I(nternal)}$) and

<table>
<tr><th colspan="2" rowspan="2">Task</th><th>Internal routine</th><th>External routine</th></tr>
<tr><th>$R_1 \ldots R_j$</th><th>$R_{j+1} \ldots R_n$</th></tr>
<tr><td rowspan="3">D_I</td><td>D_1</td><td colspan="2" rowspan="6">Essential Task Variables</td></tr>
<tr><td>…</td></tr>
<tr><td>D_i</td></tr>
<tr><td rowspan="3">D_E</td><td>D_{i+1}</td></tr>
<tr><td>…</td></tr>
<tr><td>D_k</td></tr>
</table>

Fig. 7.8 A regulation table for a task

disturbances arising in the environment of the task ($D_{E(xternal)}$). Disturbances internal to a task may come about because the task's sub-transformations are poorly defined or coupled (e.g., producing the output of the task is done by means of an inefficient sequence of activities). Another cause of "task-internal" disturbances may be that the human resources tied to the task are not equipped to perform the task's sub-transformations (because they lack, for instance, motivation or knowledge). Another type of cause for internal disturbances is that the technological infrastructure supporting the task is ineffective (e.g., because the task uses unreliable information and communication technology).

In general, then, three types of sources of internal disturbances may be discerned: (1) structure related disturbances (concerning the way the sub-transformations of the task are defined and coupled); (2) disturbances related to the task's human resources, and, (3) disturbances due the technological infrastructure of the task. As we already discussed in the chapter on Ashby – a "designer" designs a "mechanism" embodying the regulation table using and combining structural measures, and measures pertaining to human resources and technology, and a specific mix of measures from these three classes of measures was called the infrastructure. If we apply this to this chapter, we find that the task's infrastructure comprises its structure, human resources and technology. And, in this section we define "internal" disturbances as caused by the infrastructure of the task.

External disturbances originate in the environment of the task – somewhere else in the organizational network (in other tasks or in relations with them) or even in the environment of the organizational network (e.g., because suppliers do not deliver what they are supposed to deliver, or because of a change in demand).

Given these two classes of disturbances, the difference between internal and external non-routine regulation can be explained.

Internal non-routine regulation has to do with "redesigning" the task. It either amplifies the (internal or external) routine regulatory repertoire of the task so that it can henceforward deal routinely with both classes of disturbances, or it attenuates the internal disturbances by redesigning the task's infrastructure. In this sense, amplification always entails changing the task's (regulatory) sub-transformations – for it enlarges the routine regulatory potential. Attenuation (i.e., decreasing task-internal disturbances) comes about by changing the infrastructure of the task. The logic behind this way of non-routine regulation is that (parts of) the task's infrastructure may be seen as a cause of disturbances and that by redesigning (parts of) the infrastructure these disturbances may be removed. Redesigning the infrastructure may include actions related to all three of its parts: e.g., coming up with a different way of performing or coupling the task's operational sub-transformations; allocating different workers to a task, train or motivate them; or change the technology used.

External non-routine regulation deals with disturbances by changing the environment in two ways. It can either change the task's essential variables and /or their norms (thus changing the relation with the environment, and hence, as de Sitter puts it, "the environment's expectations") or it can attenuate disturbances originating in the environment. Attenuation of environmental disturbances can take effect by redesigning the network of tasks (serving as an environment of the task in question). In the example of the two related tasks in the wooden-object factory, given above, external disturbances for the "painting" task arose because of the faulty output of the "sawing and sanding" task. Attenuating disturbances for the painting task by external non-routine regulation now may include changing the infrastructure of the "sawing and sanding" task in such a way that the output of that task is no longer outside acceptable limits. Attenuation by external non-routine regulation may stretch further than the organizational network of tasks and may also come about by changing the relations with the environment of this network (e.g., contracting more reliable suppliers) or even changing the infrastructure of parts of the environment of the organization (e.g., redesign the value chain of which the organization is a part).

To summarize, a task may contain four types of regulation so that it can deal with the (virtual or actual) disturbances influencing its operational sub-transformation – depicted in Fig. 7.9.

This figure may also serve as an anchor for relating the types of regulation of de Sitter to Ashby's regulatory notions control, design and operational regulation (see Chap. 2) – which is given in the following Table 7.2.

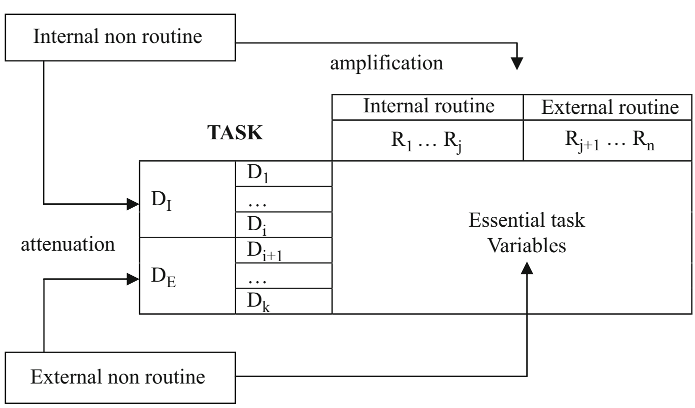


Fig. 7.9 Four types of regulation of a task

Table 7.2 Types of regulation according to Ashby and de Sitter. In the last column an alternative formulation of the types of regulation (also used by de Sitter) is given

Ashby	De Sitter	
Control	External non-routine regulation by changing essential task variables and/or their norms	Strategic regulation
Design	External non-routine regulation: attenuating task-external disturbances by redesigning network of tasks or change environment otherwise;	Regulation by design
	Internal non-routine regulation: attenuating task-internal disturbances by redesigning the tasks internal infrastructure;	
	Internal non-routine regulation by amplification, i.e. increasing routine regulatory potential	
Operational Regulation	Both internal/external routine regulation	Operational regulation

7.2.1.3 Relating Tasks into Organizational Structures

Given a set of operational and regulatory sub-transformations a designer should compose tasks and relate them in such a way that an "adequate configuration of tasks" emerges by means of which the organization's "primary transformation" can be realized. This configuration should be designed in such a way that it attenuates disturbances as much as possible, and, at the same time, amplifies the regulatory potential to deal with the remaining disturbances. In other words, a designer should design "cybernetically sound organizational structures."

A configuration of tasks (or, the *grouping* of sub-transformations into tasks and the *coupling* of tasks resulting from this grouping) is defined by de Sitter as the "organizational structure." To be precise, an *organizational structure* can be defined as: *the grouping and coupling of transformations into tasks and the resulting relations between these tasks relative to orders*.[3] In this definition, an order refers to a request for the realization of some specific desired effect (e.g., a product or service).

De Sitter (1994, p. 91 ff.) also distinguishes two sub-structures of organizational structures: a production structure and a control structure (see also previous section):

1. The *production structure* refers to the grouping and coupling of *operational* transformations into tasks and their relation to orders.

[3]This description is derived from several definitions of de Sitter (pp. 93, 100 and 101). Moreover, de Sitter sometimes uses the term architecture in his definition, to refer to the specific grouping and coupling of transformations (e.g. 1994, p. 93, 100). This seems to be redundant and even de Sitter himself sometimes omits it. Therefore, we do not follow de Sitter in using this term.

2. The *control structure* refers to the grouping and coupling of *regulatory* transformations into tasks and their relation to the production structure

The production structure highlights organizational tasks and their relations insofar as these tasks comprise operational transformations. It consists of the operational aspect of the organizational structure – i.e., the aspect realizing orders. The control structure highlights organizational tasks and their relations insofar as these tasks comprise regulatory transformations. It consists of the regulatory aspect of the organizational structure. The organizational structure is the combination of the production and the control structure. Given these two sub-structures, designing an adequate organizational structure now means designing an adequate (1) production structure relative to orders and (2) control structure relative to the production structure.

7.2.2 Relevant Organizational Variables

The previous sub-section clarified de Sitter's ideas about organizational structures. This was only the first step in understanding how these structures can attenuate disturbances and amplify potentials to deal with them. As indicated, a second step is to make clear what, exactly, the "relevant organizational variables" are, regarding to which disturbances are attenuated and regulatory potential is amplified

In de Sitter's theory, these "relevant organizational variables" are called "functional requirements." De Sitter distinguishes *external* functional requirements – requirements set by modern business environments that should be met, in order to secure the organization's viability. The external requirements can be translated into *internal* functional requirements. These internal requirements should be met by an organization in order to meet external requirements.

De Sitter specifies functional requirements into three categories: the *quality of organization*, the *quality of working relations*, and the *quality of work* (De Sitter 1994, p. 41–42).

By the *quality of organization*, de Sitter refers to an organization's potential to effectively and efficiently realize and adapt its goals. According to de Sitter, the category "quality of organization" consists of three external functional requirements: order flexibility, control over order realization, and potential for innovation.

By the *quality of work*, De Sitter refers to the meaningfulness of jobs and (the possibility to deal with) work related stress. Accordingly, he gives two external functional requirements in this category: a low level of absenteeism and a low level of personnel turnover.[4]

[4]In fact, he adds a third external requirement with respect to the quality of work – one he labels "the need to balance "qualitative demand for work and social and economic developments," by which he means that (1) jobs should fit the demands from the labor market and (2) jobs should help in increasing one's chances on the labor market. We will not elaborate this requirement.

Table 7.3 External and internal functional requirements (adapted from De Sitter 1994, p. 42)

External functional requirements		Internal functional requirements
Quality of organization	order flexibility	Short production-cycle time Sufficient product variations Variable mix of products
	control over order realization	Reliable production and production time Effective control of quality
	potential for innovation	Strategic product development Short innovation time
Quality of work	low levels of absenteeism	Controllable stress-conditions;
	low levels of personnel turnover	Opportunities to (1) be involved, (2) learn, and (3) develop
Quality of working relations	Effective communication	Shared responsibility participation in communication

By the *quality of working relations*, he refers to the effectiveness of communication in organizations.

Table 7.3 lists these three classes of functional requirements. De Sitter argues that modern organizations should strive to reach appropriate levels of all these external requirements. In fact, he states that if they do not, their viability will be threatened.

All external requirements can be translated into one or more "internal requirements." If organizations can realize these internal requirements, the external requirements will be satisfied as well. For example, the external requirement "order flexibility" refers to an organization's potential to (1) be able to deal with fluctuations in demand and to (2) deliver the required demand in time. According to de Sitter, order flexibility can be translated into three internal requirements: short production time, sufficient product variations and a variable mix of products. This means that flexibility (as an external requirement) can, in principle, be satisfied by (1) decreasing production times, (2) producing products in a sufficient number of variations, and (3) producing a sufficient number of different types of products. The last column of Table 7.3 lists the internal requirements, needed to satisfy the external ones.

The list of external and internal functional requirements comprises de Sitter's set of essential variables. If the internal requirements are met, the external requirements are met and the organization's viability is secured. De Sitter further argues that it is the task of a designer to design an organizational structure that aids in meeting all these requirements *at the same time*. Therefore, the overall adequacy of an organizational structure should be evaluated in terms of its capacity to contribute to satisfying all internal, and hence external, requirements.

7.2.3 Disturbances

In order to realize the functional requirements, de Sitter holds that organizational structures should decrease the number of disturbances affecting the three classes of essential organizational variables (attenuation), and increase the potential to deal with the remaining disturbances (amplification). To understand what this means, it is necessary to be more precise about the nature of these disturbances. What, in de Sitter's theory counts as a disturbance regarding the three classes of variables?

In Chap. 2 a disturbance was described as some event (potentially) causing the essential variables to change value. This description of a disturbance is also valid in de Sitter's theory about organizational structures. However, in his theory a distinction is made between the occurrence of some disturbing event and the probability of the occurrence of such an event. De Sitter proposes to evaluate organizational structures with regard to their capacity to decrease the probability of the occurrence of disturbances. He also introduces the concept of "sensitivity to the dispersion of disturbances" – that is, the degree to which a disturbance occurring somewhere in the network of tasks affects other tasks. For instance, the sensitivity to the dispersions of disturbances in a distribution of work in which all tasks are coupled serially, is high, because a disturbance occurring in the first task affects all others.

Given these distinctions, it is, in principle, possible to draw up a table with information about disturbances for some distribution of work – see Table 7.4.

The first column lists all possible disturbing events. The second column lists the probability of their (independent) occurrence, and the last column shows the number of tasks affected by the disturbance divided by the total number of tasks (as a measure of the sensitivity (of the organizational structure) to the dispersion of the disturbance).

This view on disturbances enables a re-specification of attenuation and amplification in the context of organizational structures. A distinction can now be made between:

1. Attenuation – i.e., the degree to which an organizational structure decreases
   a. The probability of the occurrence of disturbances (this incorporates removing disturbances, in which case the probability becomes 0).
   b. The sensitivity to the dispersion of disturbances (that is, the proportion of affected tasks is decreased)

Table 7.4 Disturbances: events, probability and dispersion

Disturbing event	P(disturbing event)	Proportion of affected tasks
D_1	$P(D_1) = \ldots$	# tasks affected by D_1/total # tasks
D_2	$P(D_2) = \ldots$	...
...	...	
D_n	$P(D_n) = \ldots$	

2. Amplification – i.e., the regulatory potential built into an organizational structure to actively deal with virtual or actual disturbances. As we discussed earlier, this can be realized by four types of regulation: internal routine, external routine, internal non-routine and external non-routine regulation – at the level of a task, a part of the network of tasks, or at the level of the network of related tasks itself.

De Sitter (1994, pp. 23–26) also gives a specification of structure-related causes of disturbances. He argues that the probability of disturbances for a particular task depends on four different things: (1) the number of relations a task has with its environment, (2) the variability of these relations, (3) the nature of environmental changes, and (4) the specificity of the norms regarding the output and the way the task should be carried out.

First, the number of relations a task has with its environment (with other tasks in the network or with the environment outside the network) can be a reason for an increased probability of disturbances. Every relation introduces a possible source of disturbances. So, the higher the number of relations, the higher the probability a disturbance will occur.

A second reason for disturbances to occur is the variability of these relations. By this, de Sitter refers to the degree to which the content of the relation varies. As de Sitter argues (1994, p.26): The chance that something goes wrong regarding a specific relation, increases if the number of different messages, instructions, requests, material, etc., associated with that relation (i.e. its *variability*), increases. This is true for two reasons. First, because every different form of "relational content" (message, instruction, request, material, etc.) can itself be a source of disturbances. Second, a high level of variability also leads to an increased probability of disturbances because one has to deal with all these different contents in one task.

As a simple example, consider the probability of something going wrong in the task of a cook, who gets orders for food from one waiter. There is only one relation between two tasks. The variability of this relation (i.e. the number of messages passed on from waiter to cook) depends on the number of items on the menu. Clearly, the probability of a disturbance is high if the number of menu-items is high because every item/message can itself be a source of confusion, and because the cook has to deal in his task with all the different messages (prepare the requested items). In general, the higher the variability, the higher the probability of the occurrence of disturbances. And, consequently, the higher the number of relations with high variability, the higher the probability of disturbances.

De Sitter sees the nature of the change of the environment of a task as a third reason for disturbances. This change can be more or less stable, or predictable. It goes without saying that the probability of a disturbance is higher if the environmental change, causing the disturbance, is harder to predict.

The last reason for task-disturbances, as treated by de Sitter, is the specificity of the norms regarding the output or regarding the way the task should be carried out. The more specific the norms regarding the output the less freedom one has to deal with output-variations by means of external (routine) regulation. In such cases this output-variation cannot be absorbed by the network of tasks, but, instead, becomes an error in it. Also, if the way a task should be performed is rigidly specified, one removes, of course, the potential for dealing with (disturbing) situations that are not covered by the task-specifications.

7.2.4 Attenuation and Amplification by Organizational Structures

Given the above treatment of organizational structures, essential organizational variables, and disturbances in organizations, we can now articulate de Sitter's argumentation with regard to designing structures in some more detail. According to de Sitter, an organizational structure should attenuate (decrease the probability of occurrence of disturbances and decrease the sensitivity to the dispersion of disturbances) and amplify (deal with occurring disturbances – by means of the four types of regulation) regarding three essential organizational variables – quality of organization, quality of work and quality of working relations.[5] The regulatory table from Fig. 7.1 can now be redrawn as in Fig. 7.10.

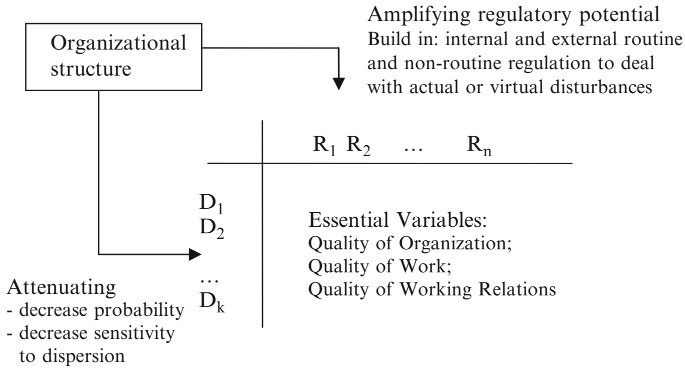


Fig.7.10 Designing organizational structures cast in Ashby's terms (part II)

De Sitter's own term for indicating that organizational structures should both attenuate and amplify regarding disturbances is "*controllability*." According to de Sitter, controllability is the overall design principle for organizational structures (1994, p. 206 ff.). In its general form, this principle is a reformulation of Ashby's cybernetic insights regarding the relation between (the variety of) disturbances and

[5] From this reasoning, it follows that designing an organizational structure means building into the organizational structure the (regulatory) capacity to change the organizational structure itself (in fact, to change all parts of the infrastructure) – compare to Chapter 1. This is so because we defined regulation by design (1) as a way to deal with disturbances – i.e. as a part of the organization's regulatory potential, and (2) in such a way that it included changing the transformation's (task's) infrastructure. In this sense it could be said that one designs for self-design.

regulatory potential. De Sitter (1994, p. 207) formulates controllability as the ratio of the potential for regulation and the required regulation[6] (see formula 7.1).

$$\text{Controllability} = \frac{\text{Potential for regulation}}{\text{Required regulation}} \tag{7.1}$$

If treated as a ratio, optimizing controllability can be achieved by both decreasing the required regulation (attenuation) and increasing the potential for regulation (amplification). As de Sitter explains (1994, p. 204/205) controllability expresses both effectiveness (if the potential for regulation matches the required regulation – i.e., for every situation where regulation is required, the potential for regulation is available) and efficiency (if the potential for regulation does not exceed the potential needed to deal with the required regulation which should be as low as possible – see also footnote 6).

7.3 Principles for Designing Organizational Structures

The goal of the previous section was to describe the desired effect of organizational structures: attenuation and amplification. This desired effect is expressed by what de Sitter calls "controllability." However, although we now have an idea of the aim of organizational structures, it has not yet become apparent how structures that *actually* attenuate and amplify should be designed. The present section will address this question. In order to do so, we start (in 7.3.1) with describing de Sitter's "*design parameters*." These parameters capture relevant characteristics of organizational structures that, according to de Sitter need to have specific values, so that organizational structures are able to attenuate and amplify. Once we know what these design parameters are we can discuss which values they should be given by designers to optimize organizational structures. In fact, de Sitter's specific design principles consist of stating which design parameters should have which value. To discuss these design principles, we introduce (in 7.3.2) two contrasting organizational structures, one with design parameters having "high values" ("high parameter

[6] Actually, he formulates it as a *function* of this ratio. The reason for this has to do with the interpretation of the ratio as a real quotient. To be able to do so, we might define the "potential for regulation" as "the number of situations, calling for regulation that can actually be regulated" and the "required regulation" as "the number of situations calling for regulation." In this case, given a certain "number of situations calling for regulation," a "controllability-value" of 1 expresses maximum effectiveness. However, it might be that this number can still be reduced. To achieve an optimal value of controllability, then, it is needed to decrease the number of "situations calling for regulation" and, next, add (or remove) enough regulatory potential to attain the value to 1. This, one may recall, is exactly how (according to Ashby) any design-process should proceed. Only if "the number of situations calling for action" can not be lowered, one may say that 1 is the optimal value for controllability. The *function* should bring this effect about.

structures") and one with "low values ("low-parameter structures"). Moreover, we will argue that low parameter structures are, according to de Sitter, better equipped to attenuate disturbances and amplify regulatory potential.

7.3.1 Design Parameters

As indicated, design parameters capture characteristics of organizational structures, relevant for attenuation and amplification. If one knows what their optimal values are (with respect to attenuation and amplification) they can be used for diagnosing and designing organizational structures. De Sitter discusses seven design parameters describing organizational structures. Every configuration of parameter-values has a specific effect on the structure's controllability – i.e., on attenuation and amplification. Designers can use this knowledge to assess and evaluate existing and proposed structures and thus try to arrive at adequate ones. Below, we discuss these parameters.[7]

De Sitter's parameters can be divided into three groups. One group of parameters describe the *production structure* (the grouping and coupling of operational transformations and their relation to orders) A second group describe the control structure (the grouping of regulatory transformations into tasks) and one parameter describes the "separation" between the operational and the regulatory transformations (resulting in a separation of tasks with respect to the production and to the control structure).

---

[7] At this point the "recursive use" of the parameter-concept as introduced in the chapter on Ashby should be noted. A parameter was described as something having an influence on the behavior of a system – or more specifically, as variable which values can influence the essential variables. With parameters, it became possible to describe the influence of "disturbances" and "regulatory actions" on essential variables. In the present context of organizational design, two "sets" of parameters and essential variables are identified. The first set has as essential variables the three "quality-variables" and to make sure that adequate values of these variables are attained, parameters such as "the organizational structure," "human resources" and "technology" may be identified. In this chapter on designing organizational structures, we focus on one of these parameters – the "organizational structure." The second set has this parameter, the "organizational structure" as its point of departure. The question becomes: how can we design an adequate organizational structure, and, hence, new essential variables (related to the adequacy of organizational structures – in terms of its capacity to attenuate disturbances and amplify regulatory potential) are defined and new parameters are identified. These parameters (we discuss them below) are called design-parameters and giving them specific values has the purpose of reaching an adequate organizational structure – or put differently – of meeting the essential variables expressing the adequacy of the organizational structure.

7.3.1.1 Parameters describing the production structure

Parameter 1: the level of functional concentration

De Sitter describes functional concentration as a parameter, referring to "the grouping of operations [operational tasks] with respect to orders" (p. 98). He goes on to describe the two endpoints of what can be called a "functional-concentration scale." A maximum value of functional concentration means that all operational tasks *of the same type* are concentrated into specialized departments, where they are performed with respect to (potentially) all orders. A minimum value of functional concentration means that all operational tasks (*of a different type*) required for realizing some order are grouped together into a "production-flow" (p.98). To understand this parameter it should be clear what operational tasks "of the same type" and "of a different type" amount to. To explain this, again consider the case of a factory producing tables and chairs. In this case, the operational sub-transformation comprises all the operational activities needed to realize orders (e.g., sales, planning, maintenance, sawing, drilling, assembling parts, and distributing tables and chairs). These activities may serve as a starting point to further decompose the operational sub-transformation. One type of decomposition specializes them into sub-transformations that are different – with respect to their output as well as to their transformation process. These sub-transformations are said to be different in type, and if one assigns these sub-transformations to tasks, these tasks can also be said to be different in type. So, for instance, the task covering the sub-transformation "sales," the task covering the sub-transformation "planning," etc. differ in type. It is also possible to design several tasks covering the same sub-transformation – e.g., assign the sub-transformation "sawing" to 15 different workers. In this case, the tasks all have a comparable transformation process (and output) and are said to be of the same type. Based on these considerations, we can now understand de Sitter's "level of functional concentration" (cf. de Sitter 1994, p.43). If it has a maximum value, all operational tasks of the same type are (potentially) coupled to *all* order-types. Moreover, in practice, they are concentrated in specialized (functional) departments (e.g., in a "sales-department," a "planning-department," etc.). In a structure with maximum functional concentration, workers perform tasks related to all possible order-types. Fig. 7.11 presents a structure in which the overall operational sub-transformation has functionally specialized departments, coupled to both types of orders.[8] If the level of functional concentration is minimal, all tasks, necessary for the production of some order-type,

[8] The figure should actually be drawn as a decomposition into parts, i.e. the departments should be "separate" sub-transformations producing output that is the input for the next department – see also section 7.2.1.1.

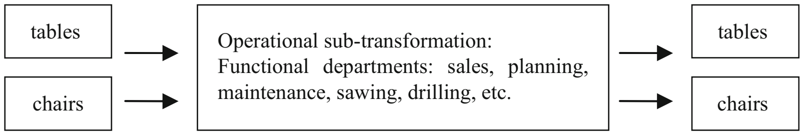


Fig.7.11 Functional concentration: all operational sub-transformations are (potentially) coupled to all (two) order-types

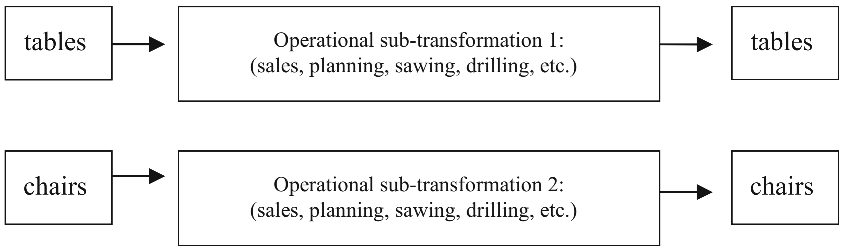


Fig. 7.12 Functional de-concentration: all order-types have their own specific set of operational sub-transformations

are only coupled to this specific order-type – and grouped together into a production-flow. In a structure with minimal functional concentration, so-called (more or less autonomous) "parallel flows" coupled to types of orders may appear.[9] In such a structure, a worker only performs a task related to one order-type – not for others. In Fig. 7.12, a structure with two flows is presented.

Parameter 2: the level of differentiation of operational transformations

The second parameter to describe the production structure is called the level of differentiation of operational transformations. In order to understand this parameter, it is necessary to see that, according to de Sitter (1994) three types of operational sub-transformations can be differentiated: "making," "preparing," and "supporting." Making refers to the actual, direct realization of the output of the transformation (or in the case of organizations: actually producing an order) – for instance, performing the sequence of activities, needed to make a table. Preparation refers to providing the necessary conditions for performing the sequence of "make"-operations (e.g.: scheduling "workers"; providing the necessary raw

[9] It should be noted that many criteria can be used to distinguish "types" of orders, e.g. their physical similarity (tables, chairs), type of customer (e.g. industry, or retail), etc., see also section 7.2.1.1.

materials, tools and equipment for producing a table). Both making and preparing are directly tied to the transformation's specific output (orders in the case of organizations): each time a specific table is made for a *specific* customer, and for this *specific* order, preparations are made. Typical preparation activities in organizations are purchase, sales or planning. Supporting refers to all operational activities that are indirectly tied to realizing the output (producing orders) such as maintenance, human resources planning, or technical service. The level of differentiation of operational transformations is "maximal" if operational sub-transformations are grouped into "make," "prepare," and "support" tasks, and minimal if operational tasks contain make, prepare and support sub-transformations.

Parameter 3: level of specialization of operational transformations

The level of specialization of operational transformations refers to "how much tasks are split up into short (-cycled) sub-tasks." Specialization of operational transformations increases as operational transformations become more specialized and these specialized transformations become separate tasks. For instance, the operational transformation "doing the dishes" can be specialized, i.e., decomposed into the smaller sub-transformations "sorting out," "cleaning," "drying," and "storing" dishes and each of these transformations can become a separate task, realized by an individual worker. Specialization decreases as specialized sub-transformations of a transformation are integrated, and become one task. For instance, "doing the dishes" may be treated as one separate task. It should be noted that specialization may concern make, support as well as preparatory transformations.

Parameter 4: the level of separation between operational and regulatory transformations

Separation between operational and regulatory transformations in tasks is maximal if (1) operational transformations are grouped into tasks that are maximally stripped from their regulatory potential (i.e., they contain very few regulatory sub-transformations) and (2) regulatory transformations are grouped into tasks separated from its operational aspect (as much as possible). In this case, "operational tasks" depend for regulation on separate "regulatory tasks." Separation is minimal if a task consists of both operational sub-transformations *and* the regulatory sub-transformations needed to regulate them. Much separation leads to a separation of the control and production structure – that is, two networks of tasks exist: one devoted to producing orders and a second one, devoted to regulating the first. Little separation leads to one network of tasks, comprising both operational and regulatory sub-transformations.

7.3.1.2 Parameters describing the control structure

Parameter 5: the level of differentiation of regulatory transformations into aspects

The level of differentiation is maximal if the three Ashby-related types of regulation (control or strategic regulation; regulation by design, and operational regulation) are grouped into different tasks. An example of this is an organization where the board deals with strategic regulation, a staff of engineers takes care of regulation by design, and workshop managers take care of operational regulation. The level of differentiation of regulatory tasks is minimal if these three forms of regulation are combined into one task. For instance, J. Jones (a butcher) who makes his own strategic decisions, redesigns his own shop, and regulates the servicing of his clientele to his own standards.

Parameter 6: the level of differentiation of regulatory transformations into parts

With this parameter, De Sitter focuses on the particular decomposition of every regulatory transformation into a "monitoring," "assessing," and "acting" part. According to de Sitter (1994, p.94), every regulatory activity necessarily involves three activities: monitoring, assessing and acting. That is, to deal with the disturbances regarding the operational aspect of a transformation, it is necessary to measure the actual values of variables defining the operational sub-transformation (monitoring). In addition, these values have to be compared with a norm (assessment). If there is a difference between the actual value of one or more variables and the norms, actions reducing the difference between the actual and the norm values of the variables should be selected and performed (acting). The effect of the regulatory actions should be that the relevant variables regarding the operational sub-transformation return within their limits. Because the operational sub-transformation may be interrupted by disturbances at any time, the three regulatory activities should also be conducted continuously. The value of this parameter is maximal if one differentiates regulation into "monitoring," "assessing," and "acting" and assigns these sub-transformations to separate regulatory tasks. The value of this parameter is minimal if these regulatory aspects are integrated into one task.

Parameter 7: the level of specialization of regulatory transformations

As with specialization of the operational transformation, this parameter refers to the level of "splitting up" regulatory transformations into small sub-transformations. The value of this parameter increases as the decomposition of a particular regulatory transformation increases, and as these regulatory sub-transformations become separate tasks. For instance, operational regulation may be decomposed into several regulatory sub-transformations: product quality, efficiency, personnel, etc., and each of these specialized transformations may become a separate task.

Table 7.5 De Sitter's design parameters

Parameter	Parameter type
1. Level of functional concentration (relative to orders)	Parameters to describe production structure
2. Level of differentiation of operational transformations in tasks	
3. Level of specialization of operational transformations in tasks	
4. Level of separation of operational and regulatory transformations in tasks	Parameter to describe the separation between the production and control structure
5. Level of differentiation into aspects of regulatory transformations in tasks	Parameters to describe the control structure
6. Level of differentiation into parts of regulatory transformations in tasks	
7. Level of specialization of regulatory transformation in tasks	

Specialization decreases as sub-transformations of a regulatory transformation are integrated into one task.

Table 7.5 summarizes de Sitter's seven design-parameters for describing and assessing organizational structures. In order to describe and assess an organizational structure, each parameter might be given a "value".[10] This value lies somewhere on the parameter-dimension for which we discussed the minimum and maximum end points. The configuration of seven parameter-values may be taken as a description of the whole organizational structure. In Fig. 7.13 a graph is drawn, tentatively picturing the configuration of parameter-values describing an organizational structure.

De Sitter's contention is that different organizational structures have a different effect on the organization's controllability. More specifically, he holds that these effects can be assessed by means of the configurations of parameter-values. Moreover, an assessment of organizational structures based on configurations of parameter-values can, in turn, be used to (re-) design them. The question, then, becomes: what the effect of specific configurations of parameter-values on controllability is (i.e., on attenuating disturbances and amplifying regulatory potential). Based on this knowledge one can treat these parameters as *design*-parameters – that is, one can select specific values of the parameters and use them as principles, as guidelines to compose an adequate network of related tasks. What designers need to

[10]To hold that these parameters can be given "exact values" on some linear scale would be stretching things. In the text we defined the dimensions on which they can be "scored" and gave their end points. Given these dimensions and end points, a relative (qualitative) valuation may be possible: relative to the end points (e.g. "the value for this parameter lies closer to a maximum value than to a minimum one") and to other parameter-values (e.g. "the value for this parameter seems to be higher than others, given a particular problem in the organizational structure"). A more adequate description of the value of a parameter might be "too high, too low or ok regarding the attenuation or amplification of organizational disturbances". Moreover, the dimensions are not entirely independent. Nevertheless, designers should somehow refer to an (implicit) version of such a parameter-value graph in their design efforts.

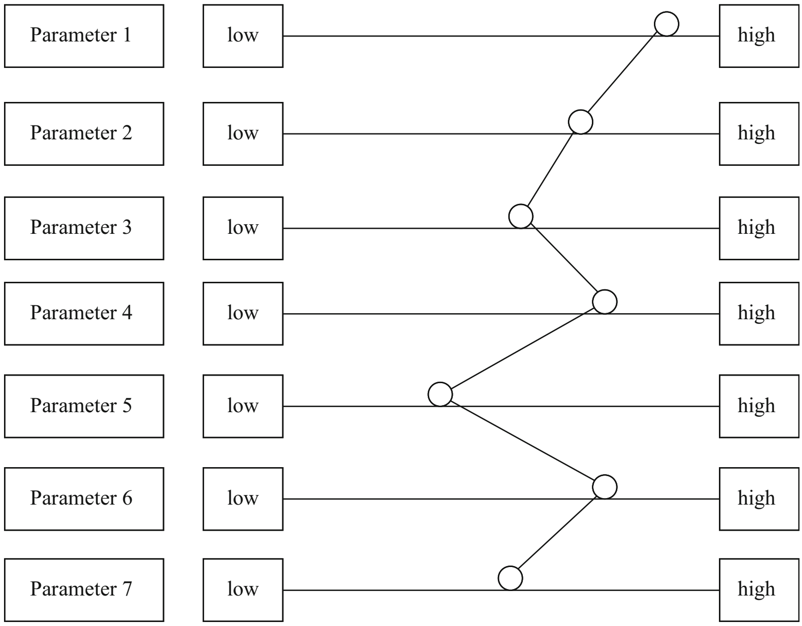


Fig.7.13 A set of parameter-values describing an organizational structure

know, then, is the nature of the impact of different parameter-value configurations – the topic of the next section.

7.3.2 Using Design parameters to Formulate Design Principles

Based on our discussion of the design parameters in the previous section, we now unfold de Sitter's *design principles*. These principles provide the rationale to select a particular organizational structure out of a set of possible ones. They provide de Sitter's answer to what an organizational structure should look like – if it is to attenuate disturbances and amplify regulatory potential. This answer rests on the effect of different parameter-configurations on the structure's controllability – as discussed in the previous section. The design principles allow a designer to make well-reasoned decisions as to the grouping of transformations into tasks and the relations between tasks at every step in the design process. Here, relevant questions are for instance, "what should the production structure at the macro-level of the organization look like given specific environmental requirements?" or "what regulatory tasks should be allocated to a team?"

De Sitter formulates one overall design-principle that provides a heuristic for designing organizational structures. This principle, in essence, reads: "An adequate organizational structure (with optimal controllability) is designed by setting the seven parameter values as low as possible". In formulating this principle, he explicitly chooses for one (and against other) specific configurations of parameter values. To understand that this principle makes sense we have to explore the reasons why setting the parameters as low as possible does, indeed, help in optimizing controllability. We will do so be contrasting two configurations: one in which the parameter-values are set as high as possible and one in which these values are as low as possible. We discuss the effects of these two configurations on the production and the control-structure and thus gain insight into how two (extreme) configurations affect controllability.

7.3.2.1 The Effect of an Organizational Structure with Maximum Parameter Values on Controllability

To understand the impact of organizational structures with high parameter-values on their controllability it is useful to explain what such structures (i.e., their combined production and control structures) look like. To start with, three design parameters are at stake when describing the production structure: the level of functional concentration; the level of differentiation and the level of specialization. If these three parameters have maximum values, they result in a production structure that may be represented by Fig. 7.14

The production structure in Fig. 7.14 contains the three differentiated operational sub-transformations that are specialized into several parts (making and support into four, and preparation into three parts). Moreover, all three order-types (X, Y and Z) pass through all preparation and making sub-transformations. In such structures, there are many relations between support and making; between preparation and making and between orders, preparation and making (see arrows). In practice, organizations with such a structure have departments in which

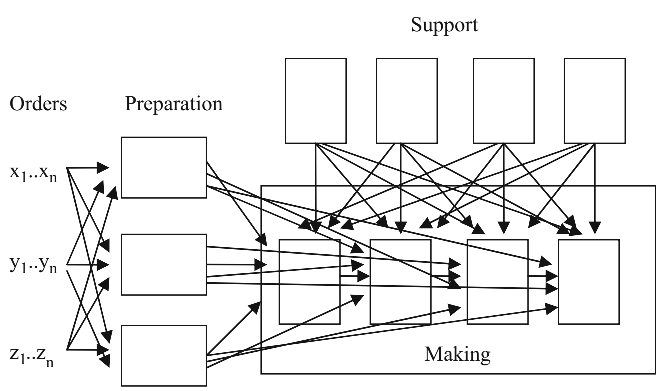


Fig. 7.14 Production structure with a high level of operational differentiation, separation and functional concentration

capacities are devoted to one specific sub-transformation. For example, "preparation"-departments like "planning," "purchase," "sales"; "support"-departments like "personnel," "maintenance," and "making" departments all containing capacities regarding one operational sub-transformation – e.g., a "sawing," "drilling," "painting," and "assembling" department. Although the structure in this figure already gives some idea of the complexity of the resulting network of tasks, it does not – by far – represent the complexity faced in "real-life maximum-parameter-value" organizations, which have far more functional departments (for preparation, support and making), and in which the specialization into small tasks may be taken much further.[11]

In the organizations we are describing in this section, the level of separation between operational and regulatory sub-transformations is high – meaning that the control structure is detached from the production structure: i.e., besides a network of operational tasks, a separate network of regulatory tasks exists. For individual operational tasks, this entails that all four forms of regulation are diminished as far as possible.

To describe the control structure, de Sitter proposes three parameters: the level of differentiation of regulatory transformations into parts and into aspects and the level of their specialization. A control structure with high values on these parameters typically consists of a hierarchy of (small) regulatory tasks. Moreover, in such control structures the three regulatory sub-transformations monitoring, assessment and action-selection are assigned to different tasks, and specific tasks (or sub-networks of tasks) are devoted to strategic regulation, regulation by design, and operational regulation.

Given this description of an organizational structure with high values on all parameters, the question for this section is what effect it has on relevant organizational variables. To evaluate this effect, we use de Sitter's re-specification of Ashby's model of regulation (see Fig. 7.15).

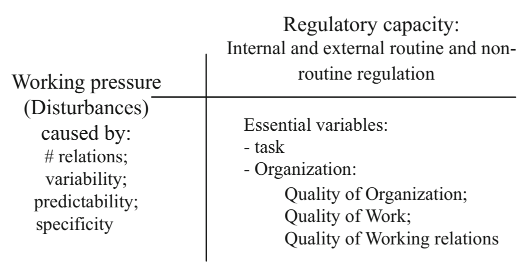


Fig. 7.15 Organizational disturbances and regulatory potential

[11] In the text, production organizations are taken as examples. However, as de Sitter stresses, the same reasoning holds for organizations providing services (e.g. governmental organizations; consultancy, etc.)

As discussed earlier, de Sitter argues that disturbances for a task may be caused by the number of relations a tasks has with its environment, the variability of these relations, the predictability of environmental change, and the specificity of the task. These parameters reflect what de Sitter calls "working pressure." To deal with this pressure a task or a network of tasks requires sufficient regulatory potential (in terms of the four ways of regulation we discussed earlier). A mismatch between regulatory potential and working pressure (i.e., a problematic value of controllability) results, at the level of a task, in the situation that norms of the relevant variables of its operational sub-transformation (its output) are not met. At the level of the network of tasks (the organization) this mismatch results in problematic values of the three "quality-variables." And, in accordance with Ashby, de Sitter holds that by means of organizational design, working pressure may be attenuated and regulatory potential may be amplified.

Based on this model we are now in a position to evaluate the effect of the organizational structure with maximum parameter values on its controllability. We do this by showing how high parameter values of the production and control structure lead to an increased working pressure and, at the same time, to a decreased regulatory potential. Next, we briefly discuss the impact of high working pressure and diminished regulatory potential on realizing tasks and on the three organizational quality variables.

The Effect of High Parameter Values of the Production Structure

The effect of high functional concentration and differentiation (into preparation, make and support transformations) is that (1) all preparation and make transformations are potentially coupled to all order-types, (2) support transformations are potentially coupled to all make transformations, and (3) that all transformations of the same type are clustered into separate departments. High specialization adds to this that these operational transformations are decomposed into a large number of small sub-transformations.

A production structure with these characteristics leads to an increase in working pressure. Differentiation and specialization introduce new sub-transformations and tasks that have to be coupled and thus lead to an increase in relations. As a simple example, suppose that I operate two machines to produce something and that I am also able to maintain both machines as I see fit. For "efficiency purposes," a separate "maintenance" task may be introduced (differentiation). Moreover, this task may be split up into two sub-tasks tailored to maintaining a specific machine (specialization). This introduces new relations, for each time something is wrong with a machine I have to consult one of the maintenance experts. And, as the Sitter argues, every relation adds a possible source of disturbances. These disturbances may be generated by chance (illness, malfunctioning of equipment, etc.) or by the production structure itself. For instance, a shared support-function introduces possible conflict between relations: simultaneous consults have to be scheduled – resulting in delay in the network (see also Galbraith 1973). In general, then, differentiation and specialization lead to an increase in the number of relations in a network.

A high level of functional concentration results in a high level of variability. De Sitter explains why this is the case. Suppose an organization produces two types of products: tables and chairs, and that both tables and chairs come in three different variations. In this case, at any moment in time, the environment can ask one of 2^6–1 combinations of different order-combinations.[12] This means that the production-structure should be capable of "dealing" with all these combinations, which, in turn, entails that the relevant (order-related) network of operational tasks should (because they are, potentially, coupled to all orders) be capable of dealing with these combinations. Every order combination means specific planning (of *this* combination); purchase of raw materials (for *this* combination); preparation of tools and equipment (for the production of *this* combination); sequences of make-operations (tailored to *this* combination), etc. And, because the tasks are related to realize the order-combinations, the relations between the tasks should, potentially, be as variable as the number of order-combinations. De Sitter concludes: the higher the number of products; the higher the number of order-combinations (which increases exponentially with the number of orders see de Sitter 1994, p. 53 ff.) and, hence, the more variable the relations in a functionally concentrated production structure. As we have indicated earlier, a high level of variability leads to an increase in the probability of disturbances. In this case: each new order-combination introduces a new possible source of disturbances, and a high number of order-combinations means that tasks themselves may become more complex (e.g., planning order-routings for six products (63 order-combinations) seems to be more complicated than for five products (31 order-combinations), which means that more can go wrong.

A production structure with high parameter values has a large number of relations and a high level of variability. De Sitter goes on to point out that both effects lead to a lack of predictability. As he puts it: the higher the number of relations and the more variable these relations, the lower the possibility to predict the input to some task (1994, p.144; p.158). Moreover, because sub-transformations of the same type are clustered in different departments the overview over orders is decreased to a considerable degree. De Sitter (1994, p. 43) describes: as soon as orders are placed, they are "made invisible" by "translating them into a sequence of required operational activities and sorting them with regard to the first operational activity" (1994, p. 43). Throughout the production process piles of parts of semi-finished products arrive at an operational department. Next, they are distributed among sub-transformations to undergo some operational activity, and finally they

[12] At any moment, a specific product (table, type 1; table, type 2; table, type 3; chair type 1; chair type 2; chair type 3) can either be ordered or not. This results in 2^6 possible combinations of orders because 6 specific products can be ordered or not. However, the combination in which no products are ordered does not count as an *order*-combination, so the total number of possible order-combination is 2^6–1. This number is even an underestimation, because it does not yet include e.g. possibly differing order quantities.

are grouped together for shipment to the next department. Individual workers realizing a small operational transformation, cannot, as a consequence, recognize the order as order. They only know that a certain activity regarding some piece of material has to be performed. In other words: at the level of an operational task one can only see the effect of the operations belonging to it, one can neither see the order one contributes to nor the rest of the operational process leading up to it.

For someone who has to distribute the work among operational workers, the order is also hard to recognize – most of the time incoming material from some department has to be distributed in order to undergo some operational treatment according to some fixed planning. Throughout the process, then, only piles of material (representing parts of semi-finished products) are visible – not the orders or products themselves. Not being able to recognize orders or the process of producing them decreases the predictability and hence increases the working pressure. De Sitter: the situation in which "workers neither know what is coming, where it is coming from, nor where it is going to" (1994, p. 31) describes low predictability. A lack of overview due to the design of the production structure has many (other) consequences which will be discussed below.

The Effect of High Parameter Values of the Control Structure

Above, we argued that high values of the production structure parameters cause an increase in working pressure. On top of that, a separated control structure with high parameter values further increase working pressure and leads to a decrease in regulatory potential. Separation, differentiation and specialization of regulatory sub-transformations obviously increase the number of regulatory tasks, and therefore, the number of relations in the network. Just as we discussed in the context of specialization and differentiation of the production structure, this increases working pressure. Generally speaking, it leads to a complex network of regulatory tasks each with a small regulatory scope, with regulators lacking insight in and necessarily detached from the operational transformation they are supposed to regulate (de Sitter 1994, Chap. 9). Moreover, it has specific consequences for the regulatory capacity of individual tasks and of the network as a whole. To see this, it is useful to remember the four ways of dealing with disturbances discussed earlier:

1. Internal routine regulation (regulatory activity does not involve other tasks and does not change the task or network of tasks)
2. External routine regulation (regulatory activity involves other tasks but does not change the task itself of the network of tasks)
3. Internal non-routine regulation (the regulatory activity does not involve other tasks but changes the task's infrastructure and/or adds routine regulatory potential)
4. External non-routine regulation (the regulatory activity either changes the task's essential variables and norms or changes the infrastructure of the task's environment).

Separation, differentiation and specialization of the regulatory aspect of a transformation have an effect on all four forms of regulation.

To start with, at the level of individual tasks, it decreases the *internal routine regulatory capacity*. Specialization of tasks in the production structure already resulted in a reduction of the *object* of regulation. For instance, due to specialization I now no longer make a chair or a table but I am supposed to saw pieces of wood. This kind of specialization obviously reduces the object of regulation – for now I deal with disturbances regarding sawing pieces of wood – not regarding making a chair or a table. Therefore, reducing the object of regulation implies a reduction of the (now redundant) internal routine regulatory repertoire. Separation, differentiation and specialization of the regulatory aspect of the task further reduce internal routine regulatory potential: they introduce new regulatory tasks that all realize parts of the internal routine regulatory capacity of the original task. For instance, due to separation I may now have a boss, responsible for monitoring the output, comparing it to some standard and proposing adjustments. And, the more separation, differentiation and specialization, the more the original internal regulation gets dispersed throughout the regulatory network, which, in turn, means a loss of the possibility to react immediately to occurring disturbances.

De Sitter also notes that high parameter values lead to a loss of internal routine regulatory capacity because the number of combinations of regulatory actions decreases. For instance, suppose that, originally, I had a task involving six different operational sub-transformations. If something went wrong while completing the fourth operational sub-transformation, I might have dealt with it by solving the problem with this sub-transformation and performing sub-transformations five and six more quickly. Such "overall regulatory solutions" are combinations of individual regulatory actions. Now, if, by specialization of the production structure my task contains fewer sub-transformations, or if by separation, differentiation or specialization of the control structure my task contains fewer regulatory actions, the number of regulatory combinations, and therefore the internal regulatory potential, also decreases. De Sitter explains that this is an exponential decrease (p. 161). In one task with six regulatory actions, $2^6–1$ regulatory combinations are possible. Removing one regulatory action results in $2^5–1$ regulatory combinations, etc.. Eventually, in tasks with only one regulatory action, internal regulatory combinations can, of course, no longer be formed. To be sure, such combinations may still be possible – but only by means of more complex process-regulation (involving regulators external to the task and with an overview over the process).

In all, high values of the parameters describing both the production and control structure result in tasks with a low internal routine regulatory capacity. They also result in tasks with a low *external routine regulatory capacity*. This capacity means trying to deal with disturbances involving other tasks in the network, without changing the network of tasks. For instance, asking a previous work-station to fix some error. High parameter values of both the production and control structure seriously impair the external routine regulatory capacity. Because of the involvement of other tasks in dealing with disturbances, external regulation requires an overview of the process – i.e., one has to know which other tasks can be involved.

This requires, at least, knowing what the previous tasks in the process are (to enable determining where the input for a task comes from) and what the next tasks in the process are (for determining the tasks dealing with the output of a task). Given this knowledge, external routine regulation also requires communication between tasks to actually deal with the disturbance. High parameter-values of the production structure, however, increase a lack of process overview and make communication virtually impossible. As we argued above, in these structures the process of making orders is made invisible for individual workers, performing operational sub-transformations – they only see the small operational part they are assigned to. As de Sitter puts it: these workers do not know where it [task input] comes from and where it [task output] goes to (see above). Moreover, due to specialization and separation, "communication with other tasks" as regulatory activity is no longer part of a task. To put the matter bluntly: the required communication is impossible because it is no part of the job, and, if even it were a part of the job, it would be useless, because one would not know with whom to communicate.

Even if one knew which other tasks one had to involve to deal with a disturbance externally, the reduced *internal* regulatory capacity of this other task probably would not allow its being involved in dealing with the disturbances of *your* task. This raises an important point: high values of parameters regarding production and control structure diminish the *internal* routine regulatory potential of a task and, at the same time, the *external* routine regulatory potential in terms of its being able to assist in dealing with the disturbances of other tasks is also decreased.

Separation, differentiation, and specialization of the control structure make matters even worse: external routine regulation of an operational task (if it exists at all in such structures) now involves many regulatory tasks (some regulator, who passes on the request to someone else, who passes on the request, . . .) before it eventually reaches relevant other operational tasks. The same holds for dealing with disturbances in the regulatory network itself. In such structures, de Sitter holds, external routine regulation is, because of their structure, nearly impossible, (1994).

Above, we discussed the effect of high value parameter structures on routine regulation. We will now turn our attention to the impact of these structures on the effectiveness of non-routine regulation. *Internal non-routine regulation* at the level of a task means dealing with disturbances by changing the task itself – i.e., by changing its infrastructure. One problem in structures with high parameter values is that, because the task itself is small, the object of infrastructural change is also very small. That is, improvements brought about by internal non-routine regulation only pertain to small parts of the process. Moreover, because of a lack of overview over the whole process one cannot observe the consequences of a change in the task's infrastructure for (the infrastructure of) the whole process – leading to sub-optimization at best. However, in most organizations with high-value parameters, internal non-routine regulation is removed from operational transformations and assigned to other tasks. At the level of these "other tasks" (regulators responsible for "regulation by design") an overview of the whole process is still required – something that is hard to obtain in these structures. At the same time, de Sitter notes that separating and differentiating internal non-routine regulation may lead to regulators who are too detached from the tasks they are supposed to change. It results, as he puts it, in " detached regulation," in which regulators can, because of their lack of involvement in the process of performing operational sub-transformations, neither be really sensitive to nature of its disturbances nor to the effects of the proposed "improvements." They structurally miss relevant "process-information" (de Sitter; p. 329 ff).[13]

External non-routine regulation at the level of a task means dealing with disturbances by changing either the goals of the own task or of other tasks; changing the infrastructure of (parts of) the network of tasks; or changing the environment of the organization. This form of non-routine regulation is greatly impaired by (if not, made nearly impossible in) a structure with high parameter values. If a task has only small specialized operational sub-transformations and does not allow an overview over the process producing orders, the effect of changes to goals or to the network's overall infrastructure cannot be appreciated. To give an example, process innovations (meaning a change in the infrastructure of the network) are very hard to initiate because mostly, disturbances stem from the complexity of the network of tasks and not so much from the complexity of the tasks themselves (de Sitter 1994, p. 50). And it is this network-complexity that cannot be overseen by tasks. Moreover, because of the reduced regulatory aspect of the task, it is impossible to experiment with and learn about effects of "improvements." Suggesting process improvements, then, necessarily becomes a matter of those not directly involved in the processes they want to improve. But this may be problematic – precisely because of their lack of involvement.

To cut a long story short, in structures with high parameter values, internal and external non-routine regulation are separated from operational sub-transformations, and assigned to differentiated, specialized tasks in the network – leading to an increase in the number of tasks and their relations in the network (and therefore to an increase in the probability of disturbances), and to a decrease of the effectiveness of regulation, because disturbances cannot be countered immediately, but should await "interventions" of structurally detached regulatory tasks.

The Effect of High Working Pressure and Reduced Regulatory Potential on Realizing Tasks and on Organizational Quality

Above, we set out to argue that tasks in organizations with high values on all parameters face high working pressure and have low regulatory potential. Needless to say, this affects realizing the task's output. While realizing a task one has to take into account a large number of relations, which are highly variable and rather unpredictable. This increases the probability of disturbances. At the same time the task's decreased regulatory potential impairs dealing with them. To deal with

[13]This is all too often observed in situations where "ICT-departments" offer (impose) "ICT-solutions" that do not support the process they are supposed to improve.

disturbances, one has to rely on a complex network of regulators, which can be a source of disturbances itself, because of its complexity. Being poorly equipped to deal with disturbances, tasks tend to infect other tasks and thus propagate disturbances throughout the network.

Structural consequences at the micro-level (the task-level) thus translate themselves into consequences at a meso-level (groups of tasks – such as departments) and at the macro-level (the level of the organization as a whole). To describe the consequences at a macro level, de Sitter proposes to use the three "quality-variables" we already discussed: quality of organization, quality of work and quality of working relations. Below, we will briefy address consequences for these three variables.

Quality of Organization

To start with, de Sitter states that organizations with high values on all parameters cannot meet the functional requirements with regard to *quality of organization* (see Table 7.4). The complexity of the network, the increased probability of disturbances and the increased lack of adequate regulatory potential necessarily affect production cycle times. Disturbances are not dealt with immediately, but by means of a separated and differentiated regulatory network, which, of course, takes time. Moreover, if disturbances tend to affect other tasks as well, the production process is even more disrupted, and requires more regulation – and hence, more time. De Sitter discusses a particular detrimental effect of maximum functional concentration on production time: the structural accumulation of stock. In functional concentrated structures operational sub-transformations of the same type are clustered together in "functional departments" (e.g., a "sawing department," a "drilling department," an "assembly department" etc.). In such structures departments are separated, and it would be rather tiresome to bring some piece of material to the next department every time an operational activity is finished. A common practice is to schedule work in batches. That is, after a fixed period of time all semi-finished products (output of the department) are collected and delivered to the next department. However, a consequence of this practice is that stock accumulates within departments, leading to longer production cycle times. De Sitter even calculates that the actual production time of a product in an organization with a functionally concentrated structure is often only 5% of the total production cycle time – meaning that a product is "stuck in stock" for 95% of the time it takes an organization to actually finish it (de Sitter 1994, p. 47).

Another functional requirement of the quality of organization is the level of control over order realization. This, too, is problematic in high value structures. Planning of orders is complex and, due to a lack of overview over the process it is very hard to trace individual orders. It is also hard to ensure the effective control of quality of products. One reason is that quality control (in essence: making sure that products meet their required demands) is dispersed throughout the regulatory network. Moreover, because of functional concentration, one structurally lacks insight in the interrelation between the different operational sub-transformations and its effect on the quality of the end-product (ibid. p. 48). This means that one cannot anticipate (during the performance of operational sub-transformations) possible failures of the products. Quality-control, then, is, at best, a reactive activity.

The last functional requirement of the variable "quality of organization" is the potential for innovation. As we discussed earlier, process-innovation is almost impossible, because of a lack of overview over the process. It is necessarily confined to the small sub-processes one is assigned to (if, that is, one has the required possibilities in a task to think of new ways of doing things and to experiment with them). Product-innovation is also difficult – because, as de Sitter argues, it requires a thorough understanding of the whole process of developing, making and selling products and therefore an intricate co-operation between marketing, product-development, production design, production, logistics, sales and service. But, such co-operation is difficult because of differentiation and, again, because of a lack of overview over the whole process.

Quality of Work

De Sitter holds that quality of work mainly depends on (1) the stress associated with task-performance, (2) the (im)possibility to be/feel involved, and (3) the opportunities a task offers to learn and develop. Not surprisingly, de Sitter's contention is that a low level of stress, and opportunities for involvement and learning have a positive effect on the quality of work. However, organizations with high values on the structural parameters fall short to meet these requirements.

To start with, the probability of task-related stress in these organizations is high. As de Sitter points out, feelings of stress with regard to work have to do with "the situation in which you face problems but are unable to solve them" (p. 21).[14] Given this description, de Sitter further operationalizes stress experienced while performing a task as "the possibility one has to deal with disturbances." Now, if the internal regulatory potential in a task is low, one has to depend on the task's environment to deal with disturbances. However, in organizations with high values on the structural parameters, this is problematic because other tasks also have low internal regulatory potential. Therefore, the external regulatory potential is also low. In such organizations, then, the probability of task-related stress is high because tasks face high working pressure (potential problems) and, at the same time, tasks have limited regulatory potential to solve them.[15]

A second problematic consequence of high parameter value organizations for the quality of work is that it is hard to "feel involved." In fact, the probability of alienation of workers is rather high in such organizations. Feelings of involvement

[14] See also Christis 1998, for a more elaborate explanation of the concept of stress based on de Sitter's theory.

[15] De Sitter shows (p.28) how results of a study by Karasek (1979) empirically support his conception of stress.

with regard to work may come about in at least two ways. One may feel "socially" involved because one is – while performing a job – actively engaged in a network of social relations associated with the job. One may also feel "intrinsically" involved because one is able to see and appreciate the point of the task – i.e., its contribution to the process of producing something as well as to the product itself. It is possible to feel intrinsically involved in a passive or more active way. If one can see the point of what one is doing and can attach meaning to one's task, one can be said to be feeling intrinsically involved in a passive way. As de Sitter (and, of course, many sociologists with him) argues: not being able to see the point of what one is doing can be seen as alienation from the task. And, alienation leads to decreased responsibility. Such passive intrinsic involvement requires an overview over the process – which can only come about in organizations that have not split up tasks too much. If one is also actively involved in co-defining or changing the task's goals, the involvement can be said to be active. Both social and active intrinsic involvement require an overview and external regulation. External regulation implies being actively involved in a network of social relations; it is by means of this network that one tries to deal with problems in a routine or non-routine way. Moreover, by definition, external non-routine regulation is (in part) about being able to (re) define goals. Thus, given this view on involvement, it is easy to see that organizations in which external regulatory capacity has been taken away from tasks, workers who are performing such tasks may not feel involved – both socially as well as intrinsically.

The last aspect attached to the quality work that is problematic in high-value parameter organizations concerns the opportunity to learn and develop. Separation, differentiation and specialization tend to make work uninteresting and repetitive. Jobs, then, neither challenge workers to learn new things nor mobilize them to contribute to the best of their ability to the organization's viability. What is more, these jobs make such contributions virtually impossible, for, as we noted earlier, workers are deprived of the possibility to think of new ways of doing their work, experiment with them and learn from such experimentation. Learning is necessarily confined to the small task one is assigned to, which often means that there is not much to learn. Leaning about the process of producing orders and about how to improve one's contribution to this process is, again, virtually impossible, because of a lack of overview.[16]

In all, the quality of work tends to be low in organizations with high values on all parameters. An increased risk of experiencing stress, of feeling non-involved, and few learning opportunities often lead to high levels of absenteeism and personnel turnover. One may conclude that such organizations do not offer workers to live their working lives to their fullest extent – a point taken up in Chap. 11.

[16]These points can be made, based on a more-or-less common sense view of the concept of learning. De Sitter, however, even unfolds his own (functional) model of learning in organizations (1994, chapter 10), in which learning is tied to regulation. He then uses this model to explore the (structural) requirements for learning. In the context of our book, however, it suffices to discuss structural impairments for learning in general.

Quality of working Relations

The last functional requirement to evaluate organizational structures de Sitter uses is the quality of "working relations." This has to do with effective communication and cooperation within the organization. To be effective, it requires knowledge about and involvement in the process one has to communicate about. As we discussed earlier, members in organizations with high values on parameters structurally lack an overview over the process and are often not involved. Therefore, the possibility to deal with disturbances by means of cooperation is structurally diminished. And, at the same time, non-involvement affects shared responsibility, necessary for effective communication and cooperation. As de Sitter puts it, in such organizations communication is often limited to what one "really" has in common – which, unfortunately, often boils down to "the workspace, toilets, coffee, [...]" (1994, p. 52).

7.3.2.2 The Effect of an Organizational Structure with Low Parameter Values on Controllability

What we now have is an overview of how the structure of an organization with high values on parameters leads to high working pressure and low regulatory capacity and how it negatively affects both performing individual tasks and maintaining the viability of the organization as a whole (as described in terms of the three quality-variables). In this section, we discuss de Sitter's alternative to these structures: those with low values on all parameters – and argue how they manage to reduce working-pressure and, at the same time, increase regulatory potential – at the level of individual tasks and at the level of networks of tasks. In many respects this reasoning mirrors the one given in the previous section – the main idea is to show that low parameter values attenuate disturbances (by reducing the number of relations in the network of tasks and the variability of these relations) and, at the same time, amplify all forms of regulation (internal, external, routine and non-routine).

Our strategy showing the effect of low parameter structures is similar to that of the previous section: we first describe their effect on the production and control structure and then move on to the consequences of these structures for realizing tasks and organizational quality (in terms of the three "quality-variables").

Before we can go into the effect of organizational structures with low parameter values, it is a good idea to give a short description of their related production and control structure. A production structure is described by three design parameters: the level of operational differentiation; the level of separation; and of functional concentration. If these parameters have low values, a production structure as given in Fig. 7.16 may result. As compared to a production structure with high parameter-values (see Sect. 7.3.2.1 / Fig. 7.14) order-types are produced in parallel "order-flows" due to a low level of functional concentration. Each order-flow is dedicated to producing one type of orders. Because of a low level of differentiation

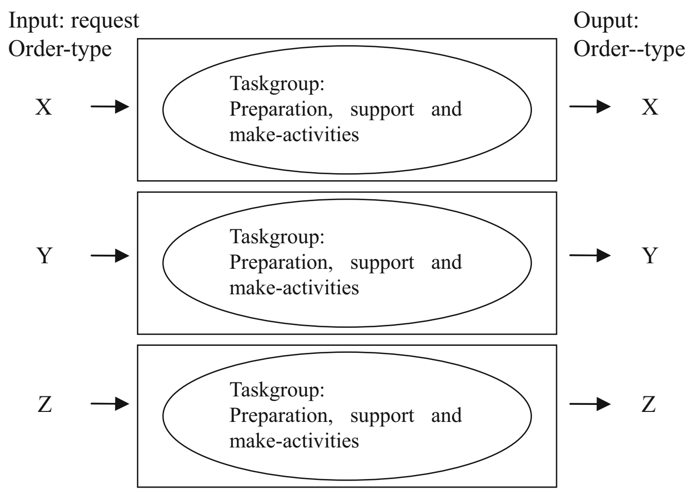


Fig. 7.16 A production structure with a low level of operational differentiation, separation and functional concentration

in each flow, making, support and preparation activities are integrated as much as possible. In addition, because of a low level of specialization, activities are not split up in many sub-transformations. As compared to the production structure with high parameter values, this structure does not have "functional departments" in which specific activities are central (e.g., specific preparation, make or support departments – see previous section). Instead, activities are organized around more or less autonomous parallel flows producing a specific type of output (e.g., a specific type of order – e.g., qua product; type of customer or geographical area). Within these flows, activities, required to produce the output are integrated into so-called "task-groups" or semi-autonomous groups (de Sitter, Chap. 8) see also Fig. 7.16. These groups resemble what Galbraith (1974) called "self-contained tasks," a specific configuration of tasks in which "[...] the subtask grouping [is changed] from resource (input) based to output based categories and [...] each group [...is given...] the resources it needs to supply the output. For example, the functional organization could be changed to product groups. Each group would have its own product engineers, process engineers, fabricating and assembly operations, and marketing activities. In other situations, groups can be created around product lines, geographical areas, projects, client groups, markets, etc." (Galbraith 1974, p. 31).

Ideally, within these order-flows, task-groups contain all possible (prepare, support and make) activities to produce the output they are assigned to (e.g., some product or order-type). Workers to whom such tasks are assigned, can, in principle, carry out all required activities for producing this output of the flow. As compared to the specialized tasks in the high-value structures, tasks in low-value structures contain a "redundancy of functions" so as to be able to quickly react to disturbances (see below).

In practice, however, producing some product often requires a large number of complex activities, and, hence, task-groups can not cover the entire production flow. In this case, de Sitter proposes to introduce (serial or parallel) segments in the flow, and assign task-groups to these segments. Defining segments should be done while keeping in mind that the level of functional concentration should be as low as possible. The idea is to define sub-output within a flow and assign a task group dedicated to producing that output; *not* to define segments containing *activities of the same type* which would re-introduce functional concentration.[17] It should, in other words lead to more-ore-less autonomous units within autonomous units – see next section for an elaboration of this idea.

In a low-value organizational structure, the level of separation between operational and regulatory sub-transformations (parameter 4) is low. This means that the production and control structure are integrated as much as possible. Instead of a separate network of tasks aimed at the regulation of operational tasks of the production structure; relevant regulatory tasks are *part* of operational tasks.

The control-structure of a low-parameter value organization has low values on (1) the level of differentiation of regulatory transformations into parts, (2) the level of differentiation of regulatory transformations into aspects, and (3) the level of specialization of regulatory transformations. Instead of a hierarchy of many small regulatory tasks, all aimed at different aspects of the control cycle and at different levels of regulation (operational, design and strategic) – which is characteristic of the high-value parameter control structure, this control structure is "flat" and consists of "integrated" regulatory tasks, including as many regulatory aspects, parts and levels as possible, covering as much of the production structure as possible.

Ideally, this leads to a structure in which the task-groups not only have the operational means, but also the regulatory means to produce the desired output. However, it should be noted that it is impossible and undesirable for complex production processes to integrate all regulatory tasks into a task-group. In our discussion below, we will assume organizations in which the values of the parameters concerning the control structure are as low as possible – e.g., in which the tasks-groups also have the regulatory capacity needed to produce the output (or even to change the infrastructure in which they produce their output). De Sitter (1994) gives numerous examples of organization in which these parameters have low values. At the end of this chapter we will discuss the issue of "how low" the parameter values of both structures can get; in the mean time it is helpful to keep in mind that we discuss the advantages of lowering the parameter-values as much as possible.

In all, if high value structures can be characterized as complex structures (large complex networks) with simple jobs, low value structures are simple structures (more or less autonomous parallel flows with a built-in control structure) containing

[17] In fact, segments should be defined in such a way that they may contain highly interdependent activities, but the dependence between segments should be as low as possible.

complex jobs (integrated tasks – involving both many operational as well as regulatory sub-transformations).

Given this idea of what an organizational structure with low parameter-values may look like, we will now turn to discussing its effect (1) on the production and control structure and (2) on the essential organizational and tasks variables. The recurrent theme in this discussion is that low parameter-values lead to a structure attenuating working pressure and, at the same time, amplifying regulatory potential. And, in turn, decreased working pressure and increased regulatory potential will lead to higher levels of the organizational quality variables.

The Effect of Low Parameter Values of the Production Structure

In a production structure with a low level of functional concentration parallel "orders-flows" emerge – "units" tied to producing a particular order type. In these flows, due to a low level of differentiation preparation, make and support activities are integrated as much as possible into tasks. A low level of specialization ensures that these tasks are not split up in too many short-cycled sub-tasks. Instead, "integrated" tasks are designed covering a large part of the production process.

As compared to a production structure with high values on these parameters, de Sitter argues that working pressure in low-value structures is far less. To be more precise: in these structures variability decreases dramatically (mainly by means of creating parallel order-flows); and the number of relations decreases (because of integrating operational tasks). To illustrate these effects, let us consider a high-value production structure of a plant producing wooden furniture, as given in Fig. 7.17. This structure has a high level of differentiation, specialization and functional concentration. It produces 8 different products: three types of chairs (x); two types of tables (y), and three types of cupboards (z). Due to a high level of functional concentration, all make and prepare departments are, in principle, coupled to all of these products (see earlier). This leads to a high level of variability:

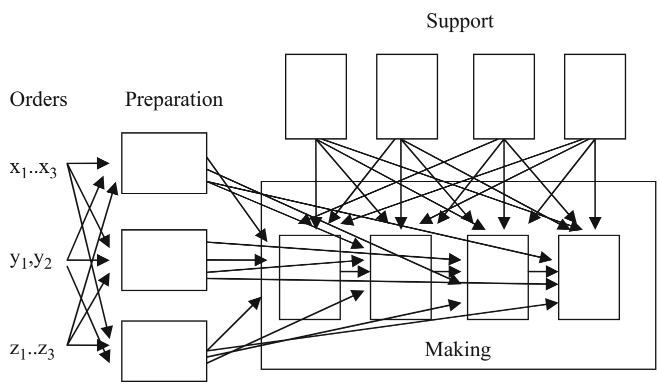


Fig. 7.17 A production structure with high values of design parameters

each prepare and make activity should be able to deal with every possible order-combination (a particular combination of different products, demanded by customers, at some moment in time). A particular order-combination in a particular week may, for instance, be: "2 chairs, type 1, 4 chairs type 2, 18 tables, type 2 and, 24 cupboards, type 3". For the sake of simplicity, we forget about the specific product-type quantities, and just consider the different types of products in an order-combination. In that case, there are 255 different order-combinations (= 2^8–1). As we have argued in the previous section, this production-structure must be able to process each of these combinations. Every combination requires a different way of doing things (e.g., in terms of planning, purchasing raw material, making-activities like sawing, drilling, assembling, etc.). Moreover, the relations in this network should also be able to handle this variety: every order-combination requires a different interaction between operational tasks (in terms of information or material). So, the variability (the variety of the interactions in the network of related tasks) is tied to 255 different order-combinations. It is convenient to use this number of order-combinations to express the variability of the network (and it is still an underestimation, because of the simplifications we've just made).

De Sitter argues that a high level of variability has a negative effect on working pressure. As we discussed before, the higher the variability, the higher the demand on the task processing it. Moreover, this effect increases exponentially with more relations Because the content of the relations between the tasks can differ considerably from time to time (qua information or material), the chance of something going wrong is also substantial (it is considerably higher than in the situation where the content of the relations is always the same).

Now, suppose that the level of functional concentration is decreased in such a way that the production structure contains three parallel substructures – each tied to a particular order-type: chairs, tables and cupboards. Figure 7.18 shows this situation. In this production structure, the variability is reduced enormously. In the substructure dealing with chairs and cupboards, the variability is now 7 (2^3–1), while in the substructure producing tables, the variability is even as low as 3. This is a reduction of complexity in the production structure from 255 to 7. In this new production structure the probability of something going wrong (as a function of the variety passing through the interaction-channels in the network) is, of course, far less. As a result, activities required in the sub-structures become simpler. For example, planning is much easier in the new structure. Also, because each flow only has to deal with one product (in two or three types) production can go faster; for it means that less time is needed to change material or to set up machines (a result that cannot be underestimated in many complex production processes).

The example shows the reduction of variability due to a decrease in functional concentration. De Sitter admits that it may not always be possible (or even efficient) to define such "autonomous sub-structures." However, it should be kept in mind that the example is primarily meant to show the effect of lowering the value of functional concentration – i.e., reduction of variability. At the same time, de Sitter proposes designers to start with looking for autonomous flow-oriented sub-structures (like the ones in the example). If these substructures cannot be found at

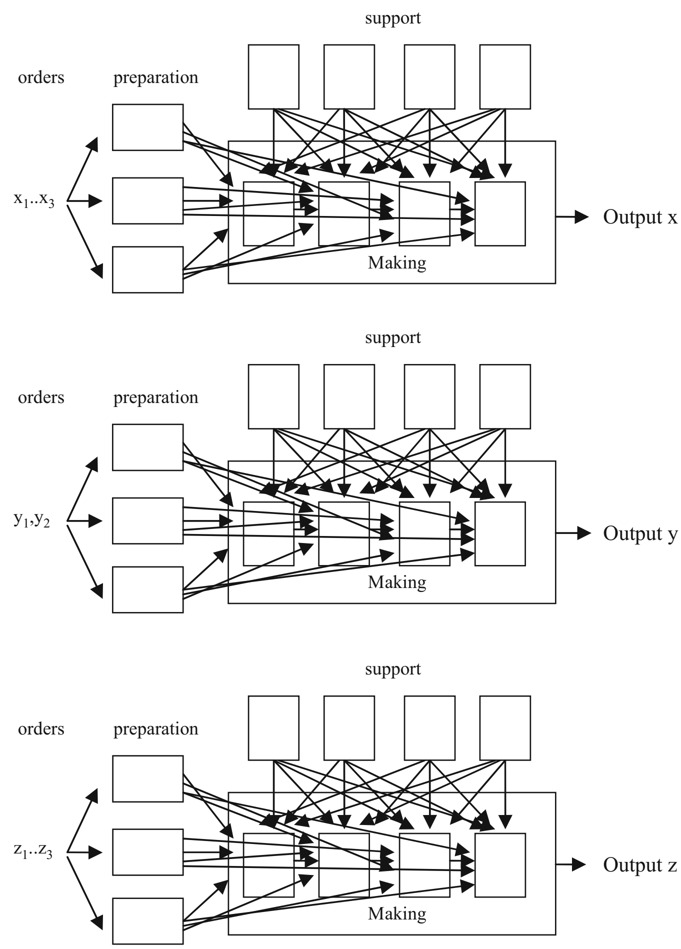


Fig. 7.18 A production structure containing three parallel production flows coupled to order-types

the level of complete orders (or order-combinations), functional concentration should be diminished within the existing structure, by identifying parallel segments in it and assign ("smaller") semi-autonomous sub-structures to these segments (cf. de Sitter 1994, Chap. 7 for the many ways in which this could be accomplished). For example, in some cases he proposes to define autonomous flow-like sub-structures tied to producing product-components that share a particular route. This may lead to different more or less parallelized semi-autonomous teams each producing product-parts that have a different route[18]. The main idea is that one

[18] In this case the "orders," relative to which functional concentration is defined is an "internal" concept; they comprise the requested product-components by the production-process.

should define a complete "in-out"-configuration and cluster tasks in such a way that they are dedicated to that particular output. This can be done at different levels – at the level of complete orders or at the level of parts of the production of an order. It should, however, NOT lead to a configuration of tasks that can, in principle, be tied to every order.

If the levels of specialization and differentiation are also lowered, (1) preparation, make and support activities are part of one task as much as possible (low level of differentiation), and (2) activities are not split up in too many short-cycled tasks, but integrated as much as possible (low level of specialization). For instance, each flow may consist of one or more "task-groups" – groups of workers capable of performing many (if not all) activities required producing the output (see Fig. 7.16 for an example of such a production structure with three different types of products).

The main effect of these parameter settings is that the number of interfaces between tasks is reduced. This has two related results. Firstly, since every additional interface introduces an additional source of disturbances in the production process, task-integration means reduction of disturbances. Secondly, if tasks are not split up, and if different types of tasks are not grouped together in separate departments, the overview over the production process increases, which greatly enhances regulatory potential (see below).

For instance, if a worker in a differentiated structure faces a problem with the machine he operates, he needs to file a request for support. Only after processing this request (which entails passing on the request to the relevant support-unit; planning support activities; and additional communication about the problem) the problem may be resolved. In a non-differentiated structure such problems may be resolved by the worker himself – which saves the time needed to file and process the request for support.

Integrating tasks in terms of de-specialization, also reduces the amount of interfaces between tasks (and hence disturbances). For instance, integrating "make"-tasks (1) means a reduction of disturbances (due to the many interfaces), (2) increases the overview over the production-process, and (3) also often means a reduction of stock. Typically, in a highly specialized production structure, semi-finished products undergo some operational activity and are grouped together for shipment to another department (i.e., they are produced in batches). This has two consequences: (1) every department has its own "stock of semi-finished products," and (2) there is virtually no overview over the production-process as a whole – the overview over the production process is very low. As we said earlier, one cannot see the product one is contributing to nor the rest of the operational process leading up to it. Now, integration of make-tasks into a flow offers the possibility to stop producing in batches (which reduces stock-size dramatically – cf de Sitter 1994, or Womack et al. 1990) and it also means an increased overview, since the new tasks now contains at least two previously separated tasks.

As with functional concentration, de Sitter is well aware of the fact that for many complex production processes, it may not be feasible to integrate all relevant activities into one task (performed by a task-group). In such a case, segments should be identified within a flow. They should be defined in such a way, that a task-group can be defined, consisting of tasks dedicated to produce the output of the segment. Moreover, these segments should – according to Simon's ideas – consist of activities that depend minimally on other segments (but may have a high internal dependency). We do not go into the details of designing segments within flows; we just want to point at the fact that de Sitter's theory addresses the relevance of identifying segments and gives detailed guidelines for doing so. (1994, Chap. 7)

In all, production structures with low parameter-values have (compared to those with high values) less variability (due to a low level of functional concentration) and a fewer number of relations between tasks (due to low levels of differentiation and specialization). This, in turn, leads, to fewer disturbances (as a function of both variability and number of relations in the network of tasks). In this sense such a production-structure *attenuates* disturbances. Moreover, it may lead to less stock and creates an increased overview of the production-process. Finally, the production structure (because it consists of parallel flows with task-groups) creates chances for lowering values of the control structure; the topic of the next section.

The Effect of Low Parameter Values of the Control Structure

As became apparent in the previous section, low values of parameters concerning the production structure *attenuate* disturbances. In this section we will argue that low values of parameters related to the control-structure (1) further *attenuate* disturbances (by decreasing working pressure) and (2) *amplify* regulatory potential.

In general, a low level of separation (i.e., not separating operational from regulatory transformations) results in tasks containing operational *and* relevant regulatory sub-tasks. A control structure with a low level of differentiation integrates different types of regulatory activities, and a low level of "regulatory" specialization results in the integration of small sub-tasks into tasks with a larger regulatory scope. This results, ideally, in regulation that is not detached from but integrated into operational tasks – enabling prompt reactions to disturbances; and a smaller network of (regulatory) tasks, reducing the number of disturbances. Moreover, tasks with a broader regulatory (and operational) scope lead to an increased regulatory potential and to an increased overview over larger parts of the production process, which, according to de Sitter, has several positive consequences.

To appreciate the result of a control structure with low parameter values in some more detail, we will discuss its effect on the different forms of regulation. To start with, a low level of both separation and regulatory specialization counters regulatory problems with internal and external *routine* regulation encountered in structures with high values on these parameters.

De-*separation* entails that operational tasks include relevant routine regulatory capacity. This inclusion entails that disturbances can be dealt with immediately (dealing with them does not require communication with and action from a network of regulators) and effectively (those dealing with the disturbances are no longer detached from the process they intervene in, but, instead, are operationally immersed in it; a condition for appreciating the effects of alternative regulatory solutions to disturbances). De-*specialization* results in the integration of regulatory tasks into operational tasks, covering a larger part of the (operational) process. This only makes sense, of course, if the production structure is also de-specialized. Together, a low level of both separation and (operational and regulatory) specialization create conditions for immediate, relevant regulation covering a larger part of the production process.

De-separation allows for including both internal and external routine regulatory potential. By decreasing regulatory specialization the routine regulatory potential increases, which means that the number of different combinations of responses to problems also increases. De Sitter points out that this increase is exponential (see the previous section in which we discussed the exponential *loss* of this kind of regulatory variety). Including external regulatory routine potential further increases the variety and flexibility in dealing with disturbances. However, as we have just remarked, this requires that workers have some overview over the production process and that the required communication with other tasks is allowed – otherwise they are unable to involve relevant other workers to help them solving or preventing difficulties.

In organizations with low parameter-values the integration of operational tasks in parallel production flows forms a basis for integrating (routine and non-routine) regulatory potential into operational tasks. Typically (or ideally), in "task-groups" team-members can perform many of the required operational activities and also have the required regulatory routine potential to counter most disturbances. Involving other team-members to prevent and deal with disturbances (external routine regulation) is a natural aspect of working in such teams. Moreover, this external routine regulation is possible in these teams because team members possess the required knowledge about the whole process and because "communication with other tasks" is built into the team-task. In such teams the built-in external regulatory potential of members functions as a back-up in case the internal regulatory potential of some member fails.

This kind of external regulatory potential is tied to individual jobs *within* a task-group. One could also define it as internal regulatory potential of the task-group itself, which should be "built into these groups." External regulatory potential at the level groups should also be available, if production flows are segmented (and, hence, consist of more than one task-group). However, in such a case, de Sitter points out that it should be (1) direct (between representatives of task-groups; not via a hierarchy or regulators), (2) reciprocal and (3) symmetric; because the information about specific parts of the operational processes and suggested measures for mutual regulation of it can only be appreciated by the respective representatives of the task-groups dealing with these parts. (de Sitter Chap 9).

On top of low levels of separation and regulatory specialization – which result in the inclusion of regulatory potential into (operational) tasks, a low level of regulatory differentiation entails that (many) tasks also contain non-routine regulatory activities. If non-routine regulatory potential is included in a task, it becomes possible to initiate and carry out changes to the infrastructure of task itself, in order to deal with disturbances. This can range from quite simple changes, such as installing new software to assist in certain activities, or small changes to the sequence of steps realizing the task's output into a more efficient one, to more complicated changes such as installing new technological equipment or introducing a new procedure for accomplishing the task. The advantage of introducing such non-routine internal regulation into a task (as compared to define separate non-routine regulatory tasks) is that it can lead to effective and efficient changes to tasks. It can be more efficient, for instance, because one does not depend on a hierarchy of regulators for changes to the task's internal infrastructure, which saves time and extra communication. It can be more effective, because the changes are initiated and implemented by those who carry out the very task that is to be changed. At the same time, including non-routine regulatory potential into a task is also demanding. It requires an overview over the chain of tasks in order to appreciate the effect of certain task-internal changes to the rest of the network of tasks (otherwise it would lead to sub-optimization). In a similar vein, it requires additional "infrastructural" knowledge on the part of those initiating and carrying out the change; on top of the knowledge required for "just" realizing one's task. One needs to know, for instance, how particular new software can aid in performing one's job, and how it needs to be installed, etc.

To deal with such problems, external non-routine potential should be included into tasks, enabling the possibility to discuss changes to the network of tasks / task-groups. Again, de Sitter stresses that such external regulation should be direct, reciprocal and symmetric.

Another way to show the effect of structures with low parameter-values on the control structure is to emphasize the increased quality of information (cf. de Sitter, Chap. 9). De Sitter argues that timeliness, completeness and relevance of information increases. It results in timely information, because there is less "distance" between disturbance and those interpreting it (it does not need to be interpreted by (a hierarchy of) detached regulators). This also leads to less communication and hence possible distortion of the information. Moreover, since operational and regulatory tasks are integrated, information about disturbances, regulatory actions and their effect is directly relevant in terms of its use in learning about them.

The Effect of Low Working Pressure and Increased Regulatory Potential on Realizing Tasks and on Organizational Quality

A structure with low parameter-values has two main effects on realizing tasks in organizations: it attenuates disturbances and amplifies regulatory potential. In Table 7.6, we summarize these effects for each parameter.

Given these effects, we can also consider the impact on de Sitter's quality variables (quality of organization, quality of work and quality of working relations). Because this impact has already been discussed (implicitly) in the previous section, and because it mirrors the impact of high-value organizations, we will only briefly list the effects.

Table 7.6 Effect of low parameter-values

Parameter	Effect	Attenuation/ Amplification
Functional concentration	Decreased variability	Attenuation
Differentiation of operational transformation	Decreased number of relations	Attenuation
Specialization of operation transformation	Decreased number of relations	Attenuation
Separation	Decreased number of relations, increased regulatory potential	Attenuation Amplification
Differentiation of regulatory transformation (strategic / design / operational)	Decreased number of relations, increased regulatory potential	Attenuation Amplification
Differentiation of regulatory transformation (monitoring / assessment / intervention)	Decreased number of relations, increased regulatory potential	Attenuation Amplification
Specialization of regulatory transformation	Decreased number of relations, increased regulatory potential	Attenuation Amplification

Quality of Organization

1. Flexibility will increase because in low parameter structures stock can be reduced significantly and product-cycle times can be increased.
2. Control over the process is increased because of the emergence of a simpler operational structure and the integration of operational tasks and regulatory tasks – leading to a more reliable production process.
3. Low parameter structures create better chances for product and process innovation. Both type of innovation require an overview over larger parts of the process (including an idea of what the client wants). It also requires the possibility to experiment and learn from experimentation. Low-parameter structures provide this overview and can build the experimental requirements into tasks.

Quality of Work

1. Low parameter structures create chances for reducing stress. Stress, as a function of the ability to deal with disturbances adequately, is reduced when relevant regulatory potential increases.
2. Since, in low parameter structures, operational and regulatory tasks are integrated and cover a considerable part of the production process opportunities for work-related learning are created. One is able to experiment with regulatory actions, see their effect and adjust actions accordingly. Working on complex tasks (as compared to high parameter structures) is more demanding for workers, but at the same time more challenging, offering opportunities to develop.
3. Because of the inclusion of relevant external regulation as part of working in task groups, involvement, both socially and intrinsically seems to be secured.

Quality of Working Relations

Working on complex tasks in task groups, equipped with an overview over a large part of the process and with the regulatory potential to deal with disturbances enables *relevant* work-related communication. Moreover, by defining cohesive segments tied to task groups, the quality of internal communication is bound to improve.

Limits to Lowering Parameter-values

In the structures we have discussed in this section all parameters should be decreased as much as possible. However, as de Sitters admits, there are limits to lowering their values and there are circumstances in which decreasing parameters below some point is not feasible. In this last subsection on low-parameter structures we list some of the criteria indicating such limits.

A first class of criteria has to do with the feasibility of defining parallel production flows. For instance, the defined flows should have a considerable degree of capacity utilization – de Sitter thinks of 80%. Below this percentage, it might not be worthwhile to add the flow. Another criterion for defining flows has to do with the number of shared recourses. If this number increases, the autonomy of the flow may be threatened too much. A third criterion has to do with the costs of building extra flows. Sometimes it may just be too expensive.

Another class of criteria pertains to the complexity of tasks, emerging in low-parameter-value organizations. Tasks may become so complex or demanding, that it may turn out to be hard to find qualified workers for them. And, even if one is able to find them, they are more expensive than those working at less complex tasks. Moreover, training workers to fit the new task-requirements may be more expensive and demanding.

Moreover, some activities cannot be integrated, because of their intrinsic complexity or nature. For some complex tasks specialization is required because otherwise, their performance could be threatened. It does not, for instance, seem to be a good idea to include all kinds of monitoring and nursery activities into the task of a brain-surgeon who needs to be focused on the operation. In a similar vein, the principle of segregation of duties imposes limits to integrating tasks: some activities cannot be integrated into one task, for the sake of financial control or accountability.

More criteria may be given, but it is beyond the scope of this book to go into them in much detail. A final remark we would like to offer is that using cost or capacity related criteria is difficult, for it requires a trade-off between the advantages of lowering parameters (in terms of improvement of the quality variables) and the disadvantages in terms of costs or loss of capacity – a trade-off for which no formulas have yet been designed.

7.4 De Sitter's Organizational Structures and Conducting Experiments

The goal of this book is to describe organizations as social systems conducting experiments and to formulate principles for their design. In part I of the book we used Ashby and Luhmann to describe organizations, thus uncovering their experimental and social *arche*. In this part we formulate principles for designing organizational infrastructures to increase the likelihood of "experimental success." To this end, we started off with Beer's Viable System Model, a model providing us with functional requirements for organizational infrastructures. That is, it delivers a set of desired effects of infrastructures – in terms of five interrelated functions. The current chapter moves beyond Beer, and states de Sitter's principles for designing a specific part of the infrastructure: the division of work.

In fact, it is our aim in this chapter to show that de Sitter's theory can be used to design a division of work that creates chances for conducting experiments with meaningful survival. To this end, we introduced several relevant concepts from his theory and showed how they were related to the design of the division of work (or, organizational structures). In particular, we discussed so-called "high" and "low parameter structures" and discussed their impact on relevant organizational variables (quality of organization, work, and working relations). According to de Sitter, low parameter structures are "cybernetically speaking" favorable structures since they attenuate disturbances for the three classes of essential variables (by design) and the required regulatory potential is built into them.

While discussing de Sitter's theory in the previous sections, it may have already become apparent that his low parameter structures not only benefit the realization of the essential organizational variables de Sitter has in mind; they also create chances for conducting experiments with meaningful survival. In this last section, we want to emphasize this effect of low parameter structures. In order to do this we start with a brief discussion of the relation of de Sitter's theory to the theories we used to describe organizations as social systems conducting experiments (i.e., Ashby's and Luhmann's) and to Beer's VSM. Next, we go into the question how low parameter structures may benefit conducting experiments with meaningful survival.

7.4.1 Relating de Sitter to Ashby, Luhmann and Beer

In his theory on designing divisions of work, de Sitter applies Ashby's theory quite directly. To start with, de Sitter defines his own set of essential variables – the requirements "modern organizations" should meet in order to survive (i.e., the variables related to quality of organization, work, and working relations). By doing so, de Sitter performs Ashby's control-step for organizations in general.

Given these variables, a distinction can be made between disturbances, caused by the infrastructure and "general" disturbances; disturbances that are not caused by the infrastructure. According to de Sitter, designing a division of work should be directed at removing as much as possible the disturbances related to the infrastructure and equip the organization with the potential to deal with the remaining ones. In this way, de Sitter applies Ashby's two basic design heuristics: attenuating disturbances and amplifying regulatory potential. De Sitter points out that only low parameter structures really attenuate disturbances and have the required regulatory potential. These structures, according to de Sitter, enable control, design and operational regulatory activities. In all, it can be seen quite easily that de Sitter's theory is firmly rooted in Ashby's cybernetics. In particular the related notions of attenuation and amplification acted as heuristics for de Sitter to develop his design-principles.

It is less straightforward, however, how his theory relates to Luhmann's. To see that and how his design-principles are related to social systems theory it may be helpful to turn to Chap. 5, in which cybernetics and social systems theory were related. Since de Sitter's theory is a direct application of Ashby's cybernetics to designing divisions of work, it can be linked to social systems theory by means of the argumentation given in that chapter. Luhmann defines organizations as social systems consisting of "decisions producing decisions." Moreover, the production of decisions by decisions is guided by the organization's structure: decision premises. Now, any division of work is the result of a process of decision-making in organizations, and can serve as an infrastructural subject in further decision-making. One could also say that a division of work serves as a decision-premise, since it conditions decision-making in organizations. For example, it structures decision-pathways, it defines what can be expected of members in organizations and it determines who are to communicate with each other. However, these relations between divisions of work and social system theory hold for all kinds of organizational structures. They are not specific for the ones propagated by de Sitter. De Sitter's main contribution consists in principles leading to *specific* divisions of work: his low parameter structures. Although one might argue that these structures allow for swift and relevant communication and lead to a different coherence of decisions (and hence to a different instantiation of an organization as a social system) than high parameter structures, this moves beyond the way Luhmann's reasons about organizations as social systems, for (when defining organizations) he is not interested in specific instantiations of organizations.

As we described earlier, part II of the book deals with principles for designing organizational infrastructures. Beer's VSM was used to derive general principles; i.e., desired effects, infrastructures should meet in order to maintain their viability. Thus, Beer's main contribution is to deliver a functional model of infrastructures. It does not go into how these functions should be *realized by a specific infrastructure* – it does not answer the question what kind of infrastructure creates better chances for viability.

By contrast, de Sitter's does treat this question. It is his contribution to organizational cybernetics that, *given* the requirements for viability, he persistently and systematically directs the attention to the question of how to design organizational structures to meet these requirements. This question remains underexposed in the Viable System Model. Although it specifies the desired effects organizational structures should realize, and in this sense is useful for diagnosing and designing organizational structures, the Viable System Model does not solve the question of how we should go about when grouping transformations into tasks to (1) implement the functions required for viability, (2) attenuate disturbances, (3) and amplify regulatory potential.

In this sense de Sitter goes beyond the Viable System Model. Just as Beer, de Sitter is fascinated by organizational viability. Unlike Beer, de Sitter explicitly directs his attention to unfolding and solving the problem of building organizational structures contributing to realizing viability.

7.4.2 How Do Low Parameter Structures Benefit Organizational Experimentation?

It is our contention that of all possible structures de Sitter's low parameter structures create the best chances for experimenting with rich meaningful survival. To start with, they enable meaningful survival. Based on de Sitter's essential organizational variables as desired effects for many modern organizations, we hope to have shown that low parameter structures attenuate disturbances and amplify regulatory potential – more than other structures.

Moreover, low parameter structures enable experiments in organizations. In fact, they build the capacity to experiment and learn into structures as much as possible. They do so at different levels, i.e., by including non-routine and routine regulatory potential into individual or team-tasks, they allow for strategic, design and operational experimentation. For instance, by adding extra routine regulatory potential into (operational) tasks, the number of possibilities to deal with disturbances increases. In such tasks, workers are required to deal with disturbances as they see fit. This includes experimenting and learning about what works in particular circumstances. By adding non-routine regulatory potential, low parameter structures also include the potential to experiment with the infrastructure and even goals. In low parameter structures, innovation and strategy-formulation is not a matter of detached designers and strategy-makers, instead these structures allow for the participation of those who are involved in operational processes in experimenting with strategic and design issues.

By integrating routine and non-routine regulatory potential into operational tasks, and by creating an overview over the transformation process, low parameter structures have a built-in capacity for experimenting and learning. According to de Sitter, this is essential. Because disturbances cannot be organized away, it is best to equip tasks –at every level, be it individual or team – with the potential to deal with them; and this should include experimenting if one is confronted with new or unfamiliar (combinations of) disturbances. Building in the capacity to experiment at all levels is also essential for innovation: for instance, it allows for swift and relevant changes of the infrastructure governing operational transformations and their regulation by including those who are directly involved in it.

In fact, many of the advantages of low parameter structures, discussed in this chapter can be phrased in terms of their ability to allow and support experimentation at all levels. Conversely, many of the disadvantages of high parameter structures can be phrased in terms of not possessing this experimental ability.

A last point we address in this section is that low parameter structures create chances for *rich* meaningful survival. That is, they create chances for "developing our characteristically human capacities" in organizations. As we pointed out earlier, they allow for learning (skills) and involvement – both intrinsically and socially. Moreover, they enable members at different levels of the organization to see the point of their work, and, hence, to reflect on the (societal) goals the organization contributes to. Although much more can be said about the relation between organizational structures and rich meaningful survival, we are at this point, not yet equipped with the appropriate concepts to identify "rich" meaningful survival. In Chap. 10, we elaborate on the notion of rich survival, and in Chap. 11, we return to discussing its relation with organizational structures.

References

Beer, S. (1995). *The Heart of Enterprise*. Chichester: Wiley.

Christis, J. (1998). *Arbeid, Organisatie en Stress*. Amsterdam: Het Spinhuis.

Galbraith, J. (1973). *Designing complex organizations*. Reading: Addison Wesley.

Karasek, R. (1979). Job demands, job decision latitude and mental strain: Implications for job design. *Administrative Science Quarterly, 24*, 285–307.

Mintzberg, H. (1983). *Structures in fives: Designing effective organizations*. New Jersey: Prentice Hall.

Sitter, L.U. de (1973). A system-theorethical paradigm of social interaction: Towards a qualitative system dynamics. *Annals of System research, 3*, 109–140.

Sitter, L.U. de, (1994). Synergetisch Produceren. Assen: van Gorcum.

Sitter, L.U. de, & den Hertog, J. F. (1997). From complex organizations with simple jobs to simple organizations with complex jobs. *Human Relations, 50*(5), 497–534.

Sitter, L.U. de, den Hertog, J.F., & van Eijnatten, F.M. (1990). Simple organisations, complex jobs: The Dutch sociotechnical approach. In: *Proceedings of the American Academy of Management*.

Thompson, J. D. (1967). *Organizations in Action*. New York: McGraw-Hill.

Womack, P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world*. New York: McMillan.

Chapter 8
Epilogue to Part II: functional and specific design principles

8.1 Introduction

In the previous two chapters, we unfolded Beer's functional and de Sitter's specific design principles. As argued, these principles can be used to diagnose and design organizational (infra)structures supporting experiments with meaningful survival. This means that we have realized the objective set for Part II of the book. In this epilogue, we summarize these principles (8.2) and reflect on their status (8.3). In this reflection, we argue that the design principles are not contingent and risky, like the selections figuring in the experiment, but necessary and certain. Section 8.4 marks the transition to Part III of the book.

8.2 Summary of the Design Principles

To summarize the principles discussed in the previous chapters, we stick to the distinction between functional and specific design principles.

Functional design principles comprise sets of (related) effects that should be produced if a system is to be able to continue its risky experiments with its meaningful survival. Within the class of functional design principles, we distinguish principles specifying: (1) what is desired for effective control in general and (2) effects a system should attain to adapt and realize the goals of its primary activities.

Specific design principles specify which system-specific infrastructural parameters are relevant to realize desired effects specified by the functional design principles and what values these parameters should have if these effects are to be realized. For organizations, there are three relevant classes of parameters: (1) the organizational structure, (2) HR-systems, and (3) technical systems. Because the organizational structure provides the foundation for the other two classes of parameters, we only explored the relevant parameters in this class and their required values.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_8, © Springer-Verlag Berlin Heidelberg 2010

8.2.1 Functional Design Principles

Regarding the functional design principles, we in the first place, summarize those related to effective control in general. These principles comprise Ashby's functional description of regulation and its clues for the design of effective and efficient regulators (Table 8.1).

Table 8.1 Functional design principles related to effective control

Effective control in general
"an essential feature of the good regulator is that it blocks the flow of variety from disturbances to essential variables" (Ashby, 1958, p. 201)
Attenuation: decreasing disturbances as much as possible
Amplification: increasing regulatory potential as much as needed
Control: determining goals
Design: specifying and implementing the infrastructure needed for regulation and performing transformations
Operational regulation: : regulating transformations
Performing transformations: realizing goals

In the second place, the class of functional design principles comprises sets of effects that should be produced if a system is to be able to conduct its experiments with meaningful survival (see Table 8.2).

Table 8.2 Functions supporting ongoing experiments with meaningful survival

Functions needed to conduct experiments with meaningful survival:	
Adaptation of goals	Realization of goals
Policy: setting goals and regulating the discussion between intelligence and control	Control: regulating the cohesion and synergy of the primary functions by: direct commands and reports, resource bargaining, audits, and monitoring coordination
Intelligence: scanning the environment and contributing to the formulation of goals	Coordination: supporting the primary activities by reducing oscillations
Control: contributing to the formulation of goals	Primary activities: realizing the raison d'être of the system

8.2.2 Specific Design Principles

Specific design principles specify relevant system-specific infrastructural parameters and their desired values needed to realize the desired effects specified by the functional design principles. These design principles are "specific," because in different types of systems, different classes of parameters may be relevant. For instance, in multi-cellular organisms or in a societal subsystem like the economy other parameters are relevant than in organizations.

As explained, in the case of organizations, relevant classes of parameters are the organizational structure, HR-systems, and technical systems. De Sitter is one of few

Table 8.3 De Sitter's parameters of the organizational structure and their desired values

Parameters relevant to the organizational structure		Desired value
Parameters relevant to the production structure	1. Level of functional concentration relative to orders 2. Level of differentiation of operational transformations in tasks 3. Level of specialization of operational transformations in tasks	As low as (is economically and technically) possible
Parameter relevant to the relation between the production and control structure	4. Level of separation of operational and regulatory transformations in tasks	
Parameters relevant to the control structure	5. Level of differentiation into aspects of regulatory transformations in tasks 6. Level of differentiation into parts of regulatory transformations in tasks 7. Level of specialization of regulatory transformation in tasks	

authors who both explicitly defines the parameters of the organizational structure and specifies what their value should be in order to realize the functional design principles. De Sitter's specific design principles for organizational structures can be summarized as follows (see Table 8.3).

De Sitter argues that if an organizational structure – described in terms of the abovementioned parameters – satisfies the desired values, it realizes a high:

1. *Quality of organization*, optimally enabling the organization to adapt and realize its primary activities by means of effective and efficient control, design, operational regulation
2. *Quality of work*, creating potentials for worker involvement, work-related learning and reducing the risk of work-related stress
3. *Quality of working relations*, providing conditions for effective communication.

8.3 The Status of the Design Principles

In this book, we conceive of organizations as social systems conducting risky experiments. Selected goals, infrastructural designs, regulatory actions, and primary operations are the "objects" in this experiment. They are contingent and risky hypotheses about what should be done in order to survive.

However, according to Beer, the functional principles are necessary and sufficient. Against this, it could be argued that because they figure in the experiment, they are as contingent and risky as the other "objects." Yet, if this were the case, why devote much attention to them? They would be just as contingent and risky as the "objects" involved in the experiment, and might be outdated tomorrow.

To establish whether the design principles are just a contingent and risky temporary fad or something more, we need to reflect on their status.

To gain insight in the status of the design principles it is useful to specify *how* they are involved in organizational experiments with meaningful survival. On closer examination, two relevant modes of involvement can be distinguished: (1) they are principles ***of*** these experiments and (2) they may be design principles figuring ***in*** these experiments.

8.3.1 Principles of Experiments with Meaningful Survival

The main argument until now can be summarized as follows. Organizations have an experimental and a social "arche." Zooming in on the experimental "arche," we argued that "meaningful survival" is what is at stake and the selections about "objects" involved in the experiments are contingent and risky hypotheses about what should be done in order to survive. Making these selections requires the realization of particular desired effects. Beer's functional design principles specify these effects.

To establish the status of the functional design principles, we depart from the supposition that, in our culture and society, historically contingent phenomena called "organization" exist. Based on reflection, we can say that organizations have an experimental "arche" that cannot be negated of them without negating organizations altogether. That is, if we grant that there are organizations, it can be argued with certainty that this "experimental" feature is a necessary feature of them.

The functional design principles, in turn, specify in terms of desired effects what is implied in the experimental "arche." They result from a reflection on what is necessarily involved in conducting experiments with meaningful survival. As such, the functional design principles specify necessary principles *of* all possible experiments with meaningful survival. They specify with certainty the necessary conditions of their possibility.

This line of reasoning implies that if a particular organization stops to realize one or more of these effects, it loses its potential to conduct experiments with meaningful survival. If this happens, its experimental "arche" is destroyed, destroying the phenomenon called "organization" with it. For this reason, *all* existing organizations realize the desired effects specified by the functional design principles.

However, *not all* organizations do this at the same level. *Some* organizations realize these effects better than others. Using Beer's terminology, one can say that all existing organizations are viable, i.e., they all realize the desired effects specified by the functional design principles, but some of them are more viable than others.

Therefore, the question becomes, what is it that makes one organization more and other organizations less viable? This is where the specific design principles come in.

We know that whether, and if so, how well or how badly, a system realizes the effects specified by the functional design principles – i.e., whether and how viable it is – depends on the design of its infrastructure. For an organization, this means that dependent on the design of its organizational structure, HR-systems, and technical systems, it is more or less able to conduct its experiments. The infrastructure "embodies" the organization's ability to conduct these experiments.

The specific design principles specify requirements to the organization's infrastructure if it is to realize the desired effects specified by functional design principles. In this sense, the specific design principles are (system-specific) principles of experiments with meaningful survival.

To the extent that these principles are deduced from either functional design principles or other "archai" of organizations (e.g., the social "arche"), they are necessary and certain as well. For instance, de Sitter deduces the parameters of the organizational structure and their required values from (1) decision premises involved in the social "arche" of organizations (goal program, task, communication pathway, etc.) and (2) Ashby's theory of regulation (his functional model, attenuation, and amplification). To this extent, his specific design principles can be called necessary and certain.

To conclude, functional design principles specify in terms of desired effects what is necessarily implied in all possible experiments with meaningful survival. They are necessary – and not contingent – principles that can be established with certainty by means of a reflection on the functional conditions of the possibility of the experiments constituting the experimental "arche." To the extent that specific design principles can be deduced from both the social "arche" and functional design principles, they share in the status of the functional design principles (see Fig. 8.1).

8.3.2 Design Principles Figuring in Experiments with Meaningful Survival

Above, we concluded that we have a set of principles specifying with certainty necessary conditions of all possible organizational experiments with meaningful survival. Based on Chaps. 6 and 7, we know that these principles can be used to guide selections with regard to the design of organizational infrastructures. The principles *of* all possible experiments with meaningful survival can thus figure as design principles *in* these experiments.

If this happens, necessary and certain principles are applied in concrete organizations to design concrete infrastructures. For instance, based on de Sitter's approach, we know that to improve the quality of organization, the level of functional concentration should be low. This principle holds for all organizations. However, in

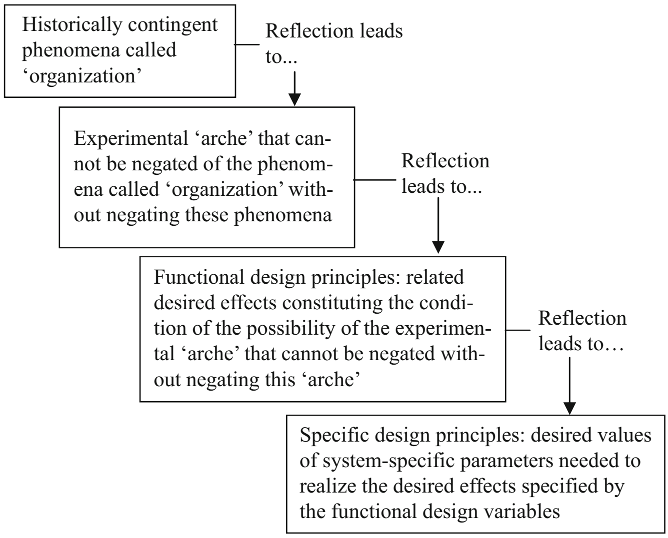


Fig. 8.1 From phenomena to necessary conditions

order to design the actual organizational structure of a particular hospital or energy company, this principle has to be applied to the particulars of these organizations.

Once design principles figure in the experiment, there seems to be a "gap" between them and the concrete situation to which they are applied. The design principles are necessary and hold for all possible organizational experiments. In this sense, they are *general*. The situations in which they are applied are concrete, i.e., they are applied to design this or that organization.

To "bridge" this gap between general principle and concrete situation, a "designer" seems to be needed. This designer (1) knows the design principles, (2) is sensitive to the particulars of concrete design problems, and (3) has the deliberative capacity and experience to apply these principles to the particulars of the problem. This designer is able to "judge" whether particular infrastructures are more or less able to realize the desired effects specified by the functional design principles.

Please note that these judgments are *not* necessary and certain; they do not *guarantee* success with necessity and certainty. Because they involve concrete, complex, and changing particulars, they are contingent and risky. However, because they also involve knowledge of design principles, experience, and deliberative capacity, and apply these to the particulars of the design problem, they can be said to be "better" and more "informed" than the judgments of those who do not have the aforementioned knowledge and competencies.

Designer's judgments bridging the gap between the necessary design principles and the concrete particulars of a design problem do not guarantee success with necessity and certainty. They increase the probability of success; they increase the probability that the organization to which they are applied is able to experiment. For this reason, selections regarding the organization's infrastructure, even if they are based on necessary and certain design principles, remain contingent and risky as was argued in Chap. 1.

8.4 Transition to Part III: Poor and Rich Survival

In Chaps. 6 and 7, we discussed design principles that, if they are applied by able designers, support organizational experiments with meaningful survival. However, in Chap. 1, we distinguished two modalities of meaningful survival: poor and rich survival.

We defined "poor" survival as maintaining a separate meaningful existence by selecting *whatever* goals, for *whatever* reason, and realizing these goals in *whatever* way. In other words, poor survival means that as long as organizational goals and means to realize them meet whatever criteria considered meaningful for whatever reason, meaningfulness is ensured, whatever the possible detrimental consequences for organizational members or the society the organization is a part of.

Rich survival was defined as maintaining a separate and meaningful existence by selecting and realizing goals that contribute to the creation of societal conditions enabling human beings to develop and realize their humanity. Rich survival makes organizational contributions to society into a serious issue. It is about organizations contributing to the improvement of society by improving conditions for its citizens to flourish.

Developing a rich concept of meaningful organizational survival is especially relevant because of the type of society we live in. In spite of all kinds of other characterizations (e.g., network society, risk society, world society, information society), our society can still be called an "organization society." By this we mean that it is a society in which organizing and organizations are vital for realizing societal values such as "health," "justice," "education," "safety," in fact, for a lot of the things we both value and take for granted in our daily lives. At the same time, we know that organizations, their functioning and its consequences, constitute a threat to all of the aforementioned values: they also contribute to huge societal problems, undermining conditions enabling us to live a fulfilled life. And, finally, to counter these problems, we still need the complexity-resolving potential of organizing and organizations.

Therefore, to attenuate the probability of organizations causing societal problems in the first place and to amplify the organizational potential to cope with societal problems, it seems to be required that organizations learn to act in a socially responsible way, contributing to societal conditions enabling human beings to live a meaningful human life. This is why we think it is worthwhile to develop the idea of organizations as social systems conducting experiments with *rich* meaningful survival, for contributing to a society enabling its members to live a fulfilled life is exactly what "rich meaningful survival" means.

This importance of socially responsible behavior seems to be increasingly acknowledged in society, in organizations (albeit sometimes reluctantly), and in management literature. Forced by public opinion, important stakeholders, or by law, organizations introduce "codes of conduct" (as instances of compliance-based approaches to business ethics) and/or encourage the development of "professional virtues" (as instances of integrity-based approaches). Here, we do not go into the question whether "codes" and "virtues" are a sign of real concern or only are there to placate important stakeholders. We just want to point at the growing acknowledgement of the importance of improving the societal record of organizations.

However, discussions and literature about "codes" and "virtues" seem to be limited to what is wanted of organizational members and how this contributes to socially responsible behavior. They pay less attention to infrastructural conditions needed to enable rich survival.

In Part III of the book, we want to address this question. We not only want to know what rich survival entails, we also want to know which infrastructural conditions can enable it. Given our focus in this book on organizational structure, we limit this question to the organizational structures that can support rich survival.

This means that it is our purpose in Part III of the book to unfold what we mean by rich survival and explain which organizational structures can support it. To this purpose, we organized this part into three chapters.

In Chap. 9, we start by presenting a vivid example of *poor* survival. We do this, to create an awareness of both the destructive potentials of organizations and the cybernetic and social systemic principles underpinning them. Given an understanding of poor survival, we develop a counter-model to it (Chap. 10). Central to this counter-model, i.e., central to the model of rich survival, is the idea of human beings "living a fulfilled human life." To explore what this idea entails, we turn to Aristotle's ethics. Finally, in Chap. 11 we discuss specific principles for the design of organizational structures supporting rich survival.

Chapter 9
Poor Survival: Disciplining Organizational Behavior

9.1 Introduction

In this chapter, we want to discuss poor survival. Because there are many possible instances of poor survival, we selected an especially vivid example, i.e., an example that illustrates everything that is possibly worrying about it. We take this example from Foucault's book *Surveiller et Punir* (1975, 1977).

In this book, Foucault discusses the emergence in the eighteenth century of what he calls the "disciplines." According to Foucault, the disciplines constituted a new way of subjecting human behavior. It was the aim of the disciplines to make human behavior both "*productive*" and "*controllable*" and to do this in a *scientific*, *deliberate*, and *methodical* way. As such, the disciplines were applied in all kinds of societal domains and in all kind of institutions such as factories, schools, asylums, hospitals, barracks, or prisons.

Although the disciplines came into being quite some time ago, we will show that the methods constituting them (1) are applications of the cybernetic and social systemic principles we discussed in previous chapters and (2) can still be found in the organizations of today. This is rather *disquieting*, because attached to the disciplines are features that are destructive to the persons subjected to them and may be destructive to the societies they are a part of.

To show how it is possible that the disciplines and the management of modern organizations share the same cybernetic and social systemic principles and to show how the disciplines and their principles can still be found in organizations of today, we proceed step by step (see Fig. 9.1).

In Sect. 9.2, we begin by introducing the disciplines as they were described by Foucault (see Fig. 9.1, bottom left). Given this introduction, we analyze their cybernetic and social systemic underpinnings. This analysis is the subject matter of Sect. 9.3.

In that section, we first analyze the disciplines as described by Foucault from a *cybernetic perspective* (Fig. 9.1, cybernetic generalization). As a result of this analysis, we find the set of general "cybernetic" activities underpinning instantiations

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_9, © Springer-Verlag Berlin Heidelberg 2010

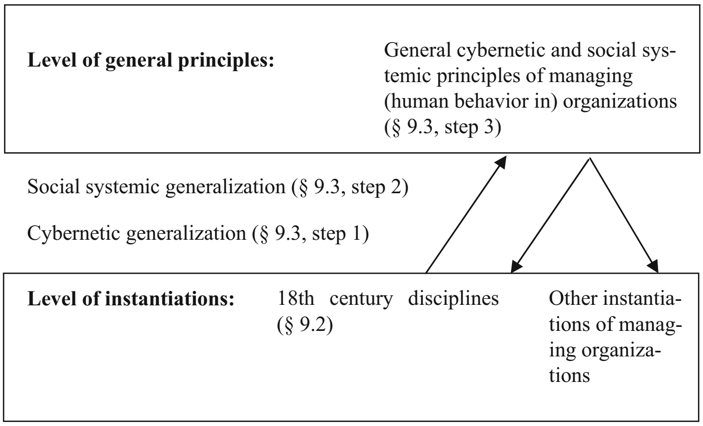


Fig. 9.1 Outline of the discussion of poor survival in this chapter

of disciplining human behavior in both eighteen century and today's organizations. Second, we analyze the disciplines from a *social systemic perspective* (Fig. 9.1, social systemic generalization). Once again the aim of the analysis is finding general principles, yet this time general social systemic principles, underpinning the management of human behavior in organizations. We show that there is a similarity in the way the disciplines manage human behavior and the way this is accomplished in organizations. Third, we combine the results from both analyses by showing how the related cybernetic and the social systemic principles allow for the management of behavior in eighteenth century disciplines and today's organizations (Fig. 9.1, bottom left and right: eighteenth century disciplines and other instantiations of managing organizations).

Figure 9.1 captures the argument starting at the level of one instantiation – eighteenth century disciplines –, then proceeding by generalization to arrive at general principles cybernetic and social systemic principles that are at the root of both the disciplines and all other instantiations of managing behavior in organizations.

Section 9.4 is devoted to disquieting features of the disciplines, i.e., to the features that make the disciplines into a good example of poor survival. First, we list and describe these features. Second, we elaborate on the generalizations made in Sect. 9.3. If the disciplines are such a good example of poor survival, and if their principles are the general cybernetic and social systemic principles underpinning the management of whatever organization, then the question arises whether *all* organizations share the disquieting features attached to the disciplines, i.e., whether all organizations only can achieve poor survival, making rich survival into an utopist project.

To anticipate on the results of this discussion in the next chapters, we think that rich survival is not unachievable. We think that it is possible to avoid the disquieting features attached to the disciplines. Or, to phrase it in more positive terms, we think that some of the instantiations of the general cybernetic and social systemic principles underpinning the management of organizations are desirable, both challenging and supporting organizational members to develop themselves. These instantiations constitute examples of what we call rich survival. Other instantiations – like the disciplines – exemplify poor survival. Some of these instantiations deny organizational members, organizations, and even society the chance to develop themselves to the best of their ability. In the most evil cases, they may systematically degrade the humanity of the "subjects" entering their domain, and even become vehicles for corruption and death.

We can summarize this possibility of rich to extremely poor instantiations of the general cybernetic and social systemic principles of managing organizations by stating that these principles are "*pharmakon*." Just as knowledge of the principles of pharmaceutics can be used to cure or kill someone, insight in the general cybernetic and social systemic principles underlying the management of behavior in organizations can be used to instantiate rich to poor forms of organizational survival. Because these principles are *pharmakon*, it is important to be able to distinguish between rich and poor survival. As indicated, in Sect. 9.4 we start to make this distinction by describing the disciplines as a vivid example of poor survival, outlining their disquieting features. In contrast with this inventory of disquieting features, we develop the concept of rich survival and the principles of its infrastructural conditions in Chaps. 10 and 11.

9.2 Foucault: The Disciplines in the Eighteenth Century

To begin our exposition of what we mean by poor survival, we turn to Foucault's book *"Surveiller et Punir*" (1977). In part two of this book, Foucault describes the development and structure of methods, new in the history of mankind, to make human behavior both controllable and useful. "These methods, which made possible the meticulous control of the operations of the body, which assured the constant subjection of its forces and imposed upon them a relation of docility-utility, might be called "disciplines'" (Foucault 1977, p. 137).

Of course, methods with a similar purpose existed long before the eighteenth century. Earlier centuries had brought, for instance, slavery, domestication, and the ascetic life as ways to control human behavior and to make it useful. However, what, according to Foucault, is new in the eighteenth century is that the disciplines make the control and usefulness of human behavior into an explicit object of *scientific research* linked *to deliberate design,* enabling *methodical* control. Never before, had this particular link between research, design, and method occurred in such a systematic way for the sole purpose of subjecting behavior.

In "Surveiller et Punir" (1977), Foucault follows the traces of the development and structure of the disciplines in the infrastructures of eighteenth century schools, hospitals, prisons, armies, and factories. To explain what the disciplines are and how they work, we abstain from discussing Foucault's historical analyses and go straight to the heart of their structure.

According to Foucault, the disciplines consist of seven related "activities." These activities are: (1) the analysis of space, (2) the analysis of bodily operations, (3) the analysis of the process of production, and (4) the synthesis of space, bodily operations, and production processes into an infrastructure creating the conditions for the production of controllable and productive behavior. The actual behavior of people involved in this infrastructure is managed by (5) hierarchical surveillance and (6) the use of normalizing sanctions (rewards and punishments). Finally, by means of (7) examination of the operations of the bodies entering the sphere of the disciplines, the infrastructure, and thereby the control over and the usefulness of behavior, is continuously improved. Let us discuss these activities one by one.

9.2.1 The Analysis of Space

According to Foucault, the analysis of space evolves in two steps. First, space is separated into "outside" and "inside": outside and inside the school, the hospital, the prison, the barracks, or the factory. This first separation constitutes a difference between a heterogeneous world "outside" and a homogenized space "inside" devoted to a particular purpose (teaching, health, re-education, training, or production of goods or services). Then, within the demarcated space, fenced off from its environment, space is further analyzed into parts. This internal subdivision of space aims to enable the effectiveness and efficiency of production processes, hierarchical surveillance, and examination.

9.2.2 The Analysis of Bodily Operations

To make human behavior productive and to make production efficient, the operations of the human body are studied and analyzed into "basic operations".[1] In turn, these "basic operations," can be standardized, trained, and synthesized into larger complexes. Thus, "basic operations" can be used as building blocks for the design of more complex operations. For instance, using a set of "basic" operations of the body resulting from analysis, soldiers can be trained to execute more complex processes such as "forming a line," "presenting arms," or "aiming, firing, and reloading arms" with high predictability and precision.

[1] The production process and its analysis provide the context for determining what can be "basic" operations. For instance, at a production line for mobile phones, "basic operations" may be defined at the level of minute arm, hand, and finger movements that must be carried out with high precision in very short cycle times. "Gardening" may require quite different "basic operations." Or, in modern "knowledge oriented" organizations "basic operations" may be translated into systems of more or less specified competencies enabling, for instance, problem solving, presiding meetings, negotiating, etc.

9.2.3 The Analysis of the Process of Production

To make “basic operations” productive, they need to be linked to processes of production. The way to do this is to analyze or decompose the entire process of production into increasingly smaller tasks that, in the end, can be linked to (complexes of) “basic operations.” For instance, the process of “fighting a battle” is decomposed into smaller tasks such as “forming a line,” ”forming a column,” “aiming, firing, and reloading arms,” etc. These tasks, then, are coupled to complexes of basic operations that thereby become productive.

9.2.4 The Synthesis of Space, Bodily Operations and the Process of Production

The analysis of space, bodily operations, and production processe, provide the building blocks that can be combined or synthesized into infrastructures enabling productive behavior and its control. By linking space, basic operations, and tasks, an infrastructure is put into place, allowing for the performance of complex production processes. The people involved in this infrastructure, only need to master and perform sequences of basic operations allotted to them. It is the designer who has to take care of the larger picture, arranging the basic operations to realize the overall goals. Just as a clock may be built by combining small parts, the “parts” resulting from analysis can be combined to build a “production machine” that is as predictable as a clock. In this way, confining human behavior to make it productive becomes an “Ars Combinatoria,” a skill of combining basic parts into larger wholes enabling the production of controllable behavior that can be put to any use.

These first four elements of the disciplines are “architectural.” They presuppose the specification of goals and they involve the analysis of space, bodily operations, and production processes into parts as well as the synthesis of these “parts” into an infrastructure enabling the production of controllable and productive behavior, realizing the established goals.

However, the infrastructure resulting from the first four activities is just a prerequisite for actually producing productive behavior. The activities we describe below ensure that the productive potentials built into the infrastructure are realized. They constitute the cybernetic motor running smoothly at the heart of the disciplines. They are what Foucault calls “the means of correct training” (1977, p. 170). They are “hierarchical surveillance” and “normalizing sanctions.”

9.2.5 Hierarchical Surveillance

By “hierarchical surveillance,” Foucault refers to *monitoring* the behavior of the people put to work in a particular structure. Monitoring behavior is needed, in the first place, to "normalize" it, i.e., to make and keep the "operations of the body" according to the specifications of the design. Second, monitoring is needed for the purpose of the reproduction and maintenance of the design of the structure.

To make surveillance of many workers possible, a *hierarchy* of observed and observers is put into place. Workers are observed by foremen, who, in turn, are observed by their bosses, etc., right to the observers at the top of this pyramid (who, if there are more, observe each other). Foucault describes at length how the spatial design of schools, factories, and hospitals is geared to the purpose of hierarchical surveillance. A particularly strong example of such a design is the so-called "Panopticon."

Originally, the panopticon – an invention of the philosopher Jeremy Bentham – enabled hierarchical surveillance in prisons, but, as we shall see later on, its principle has universal application in organizations (Foucault 1977, p. 195ff.). In a panoptic prison, the prisoners are locked up in cells arranged in a circle around a central tower. The guards reside in the tower. Between the prisoners in the cells and the guards in the tower is an asymmetric observation and power relation. From the tower, the guards can observe the behavior of the prisoners. Thus, this structure functions as a "platform for visibility" of the behavior of the prisoners entering its space.

Dependent on their behavior, the guards have the power to punish or reward the prisoners. However, from the cells it is impossible to see the guards in the tower and the prisoners do not have the power to punish or reward the guards. The prisoners know that their behavior may *possibly* be monitored and punished or rewarded by the guards in the tower. But, they do not know whether their behavior is *actually* monitored. They do not even know whether there are any guards in the tower.

The panopticon is such a strong design because, after an initial training period, the prisoners, knowing that their behavior *may* be observed and punished or rewarded, behave *as if* their behavior is *actually* observed. Even if there are no guards in the tower, the prisoners do not want to run the risk of punishment and continue to behave *as if* they are observed. In this way, the prisoners are disciplined to "normalize" their own behavior; they are disciplined to discipline themselves. They are part of an infrastructure that makes them into the guards of their own behavior.[2] Thus, the panopticon becomes a "platform for normalization" of behavior.

According to Foucault (1977, p. 195ff.), the panoptic principle of the asymmetric visibility and power relation between the observed and the observer can be applied in many different contexts, such as factories, hospitals, barracks, and schools. Moreover, it is still used abundantly in modern organizations. For instance, the automation of production processes may generate management information that can be used to reward or punish individual workers. In this way, information

[2] A number of presuppositions have to be met if the disciplines are to work. For instance, prisoners must once in a while witness the punishment of (other) prisoners. In such cases, they must be able to connect the reason for punishment (the undesired behavior) both to the punishment and to their own behavior. It is not opportune here to explore all these presuppositions.

technology may function as a "one-way mirror" allowing managers to observe the behavior of their subordinates who know *that* they may be monitored without knowing *when*. Below, we will argue that there is an even deeper relation between the panopticon and organizations than Foucault imagined.

9.2.6 Normalizing Sanctions

In spite of the functioning of the panoptic principle, hierarchical surveillance in isolation cannot yet produce controllable and useful behavior. It needs to be linked to the application of normalizing sanctions. By the term normalizing sanctions, Foucault refers to a system of punishments and rewards that, in principle, is as detailed as, and linked to the "basic operations" they are supposed to govern. Given both hierarchical surveillance and the system of normalizing sanctions, even the smallest positive or negative deviations from the desired "basic operations" may be monitored, assessed, and rewarded or punished. Combined with hierarchical surveillance in the form of the panoptic principle, the application of normalizing sanctions makes that the observed – the prisoners, the soldiers, the workers – discipline themselves to perform even the smallest operations exactly as desired by the structure of the processes they are involved in.

It is characteristic of normalizing sanctions that the punishments and rewards *themselves* are designed with the purpose of normalizing behavior. Organizational members are not punished, for instance, by a beating or by physical confinement. These are punishments that do not necessarily make their behavior more controllable or productive. In the case of the disciplines, punishment is designed to achieve the type of behavior that is wanted. Preferred punishments therefore include:

- Corrective training to re-enforce useful skills and obedience
- Demotion to re-enforce obedience and hierarchy
- Marginalization or even exclusion to separate hopeless cases, bad examples, disturbing influences, or agitators from the rest to prevent them from setting a bad example.

Preferred rewards include:

- Education and training to re-enforce useful new skills and obedience
- Small promotions to re-enforce obedience and hierarchy.

Together, hierarchical surveillance and normalizing sanctions produce normalized behavior, i.e., behavior that is controllable and useful. Negative deviations from what is desired are corrected by carefully selected punishments. Normal behavior or positive deviations may be rewarded (either by withholding punishment or by "normalizing" rewards). Thus, they re-enforce the structure that makes them possible, govern the production of controllable and useful behavior, and regulate the actual processes of production.

9.2.7 Examination

One of the characterizing elements of the disciplines is what Foucault calls the examination. Once individuals have entered the space of the disciplines, their behavior becomes the object of systematic scientific observation. They become "cases." Their operations are recorded and analyzed; they are compared to standards and to each other in order to search for possible improvements. On the basis of this research, bodily operations, and production processes may be synthesized into new and more productive ways, leading to the design of improved infrastructures. In addition, research may be used to improve the methods of hierarchical surveillance and the system of normalizing sanctions, adding to the potential to both produce productive behavior and regulate actual production processes. And finally, examination may turn upon itself, and improve the methods used to research human operations.

By means of the examination, the disciplines acquire the capacity for the continuous transformation of their own particular instantiation. In the course of time, the specifics of the design of organizational structures, of the methods for surveillance, of the system of normalizing sanctions, or of the methods for the examination may change, but the seven elements underpinning this process of transformation remain the same. In this way, the spaces governed by the disciplines become a kind of laboratories in which the transformation and improvement of production and control are allied to the emerging behavioral sciences. Foucault – not without a touch of drama – even maintains that the birthplace of the human sciences as we know them today is "probably to be found in these "ignoble" archives, where the modern play of coercion over bodies, gestures and behavior has its beginnings" (1977, p. 191).

Together, the seven activities constitute an efficient "machine" for realizing the different types of goals served by the disciplines. For the purpose of the generalizations in Sect. 9.3, it is useful to distinguish two types of goals here.

The first type is the goal of the regulation of human behavior in organizations as such. Foucault defines it as "permitting the minute control of the operations of the body, assuring the continual submission of their forces and imposing upon them a relation of docility and utility." However, this formulation is tied to Foucault's eighteenth century description of the disciplines. For the purpose of generalization, we propose a more general goal: "permitting the production of controllable and useful behavior." This general goal can be further specified dependent on what "controllable" and "useful" are taken to mean. For example, in the eighteenth century disciplines as described by Foucault, "controllable" and "useful" are interpreted as the "continuous submission of the forces of the operations of the body imposing upon them a relation of docility and utility." In modern organizations, "controllable" and "useful" may be interpreted as "professionally steering one's own behavior in complex unpredictable environments." Because what is considered as "controllable" and "useful" may change over time, the goal of regulating behavior in organizations is not fixed.

The second type of goals consists of the goals of the production processes for which behavior is to be made controllable and useful. For instance, in a school, this goal may be "educating children" or in a factory "producing good X." This type of goals can vary in different situations and can be specified at different levels of recursion. For example, at the same time different schools may have different educational goals and one and the same school may have different educational goals at different moments in time. An example of different goals specified at different levels of recursion is that the goal of a factory may be defined as "producing energy" and that the goal of one of its divisions may be specified as "producing wind energy."

In the disciplines the two types of goals stand in a circular relation. The first type of goal, i.e., permitting the production of controllable and useful behavior, is realized by realizing the second type of goal, i.e., by realizing the goals of the production processes. For, it is by the disciplined realization of the goals of the production processes that behavior is made controllable and useful. The second type of goal, i.e., realizing the goals of the production processes, presupposes the first type of goal. For, to realize the goals of the production processes, the controllable and useful behavior of organizational members is required.

The two types of goals are presupposed by the disciplines. Foucault does not describe "goal setting activities" as a part of the disciplines. Relative to already established goals the disciplines are a "machine" to realize them. Given whatever goals with respect to production processes and controllability and usefulness, the disciplines provide the underlying structure to realize them. Relative to the disciplines the goals of the production processes are contingent (i.e., they can be anything).

To wind up our discussion of the disciplines, it must be admitted that Foucault in "*Surveiller et Punir*" reconstructs these activities using historical accounts of eighteenth century practices and textbooks from that period prescribing the organization of schools, prisons, hospitals, factories, etc. Thus, the instantiations of the disciplines described by Foucault are tied to a particular period in history. Therefore, it may be hard to find an exact copy of them in "knowledge-based," "resource based," "post-modern," or even most "modern" organizations (however, "sweat shops" are still abundant). Members of these organizations would probably find it difficult to recognize themselves in Foucault's bleak picture of "docile bodies," "submitting their forces" to "purposes" they do not need to comprehend. It is even probable that most of them will reject the particular instantiation of the disciplines described by Foucault, both as not really productive and as morally flawed. In spite of the many differences between the methods and techniques of the disciplines as described by Foucault and current methods and techniques for managing organizations, at a closer look they appear to be instantiations of the same set of principles. In the next section, we set out to uncover them.

9.3 A Cybernetic and Social Systemic Generalization of the Disciplines

In this section, we analyze the disciplines to charter the principles underlying both eighteenth century discipline and management in today's organizations. As a result, we get a general cybernetic and social systemic answer to the question how organizations – being the particular type of social systems they are – discipline organizational members to discipline their own behavior.

9.3.1 Cybernetic Analysis of the Disciplines

In this step, we analyze the disciplines as described by Foucault from a *cybernetic* perspective. The aim of this analysis is to uncover the invariant cybernetic principles governing both the eighteenth century disciplines and today's management of organizations.

More in particular, by using the conceptual framework provided by cybernetics, we show that Foucault's disciplines can be modeled as consisting of two related feedback loops. The first loop is devoted to the *realization* of the two types of goals mentioned above and the reproduction of the infrastructure constituting the disciplines. The second loop is devoted to the *adaptation* of the disciplines. Together, the loops contribute both to the production of controllable and useful behavior and to the realization of the goals of the production processes. Moreover, they provide the capacity for self-transformation, supporting the development of both the disciplines and the many instantiations of management active in today's organizations.

To uncover the cybernetic principles governing the disciplines, we start by distinguishing two types of categories of activities.

1. The *first* category comprises activities *presupposed* by the disciplines. It consists of goal setting activities. More in particular, it comprises setting goals with respect to what is considered "controllable" and "useful" behavior and defining goals with respect to production processes (see Fig. 9.2, light grey area).
2. The *second* category comprises activities *constituting* the disciplines (see Fig. 9.2, dark grey area). There are two subcategories of activities here.
   – *Architectural activities*: the analysis of space, bodily operations, production processes, the synthesis of the results of analysis into an infrastructure conditioning the realization of the goals, and the examination to improve the infrastructure
   – *Regulatory activities*: hierarchical surveillance and applying normalizing sanctions.

Given this classification, it is not difficult to recognize the three basic cybernetic functions discussed in Sect. 2.3.2 of the chapter on Ashby: control, design, and operational regulation.

The *control function* sets the goal. In the case of the disciplines it consists of the goal setting activities allowing for the adaptation of what is considered controllable and useful behavior and of the goals of the production processes.

The *design function* sets the scene for the realization of the goals specified by control. In the case of the disciplines as described by Foucault, design consists of the analysis of space, bodily operations, and production processes, the synthesis of

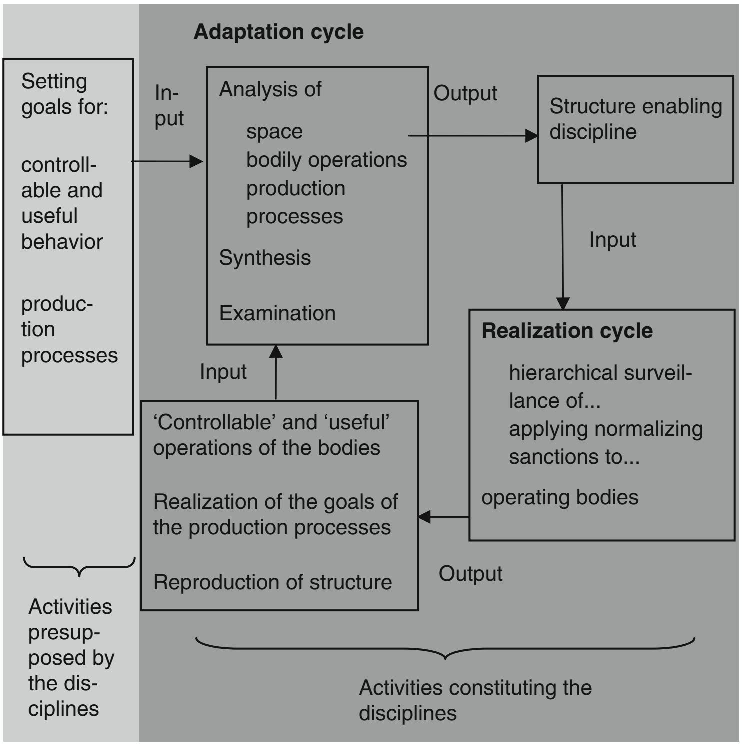


Fig. 9.2 Eighteenth century disciplines: adaptation and realization

the results from analysis into an infrastructure enabling discipline, and the examination. The examination provides "analysis" and "synthesis" with information about the processes that may be redesigned.

Given the infrastructure, *operational regulation* secures the realization of the two types of goals. In the case of the disciplines, hierarchical surveillance and the application of normalizing sanctions constitute operational regulation. Their object is the operations of the bodies, ultimately realizing the goals of the disciplines.

This reinterpretation of the activities constituting the disciplines in terms of control, design, and regulation, also exposes the discipline's two connected cycles. The "*realization cycle*" takes care of the realization of the goals and the reproduction of the structure enabling discipline. The "*adaptation cycle*" defines and adapts the structure enabling discipline, conditioning the processes in the realization cycle.

The *realization* cycle is regulated by hierarchical surveillance and applying normalizing sanctions. Their object is: the operations of the bodies. Given the goals and the infrastructure specified in the adaptation cycle, the operations of the bodies performing the actual production processes are regulated by hierarchical surveillance and the application of normalizing sanctions. If the operations of the bodies deviate from the specified goals or infrastructure, it is picked up by hierarchical surveillance. By means of applying normalizing sanctions, these deviations are then corrected. Thus, hierarchical surveillance and normalizing sanctions *regulate* the operations of the bodies. By means of regulation, the bodies entering the disciplines realize the goals of the production processes, reproduce the infrastructure enabling discipline, and behave according to what is considered "controllable" and "useful" behavior. In this way, the operations of the bodies, surveillance, and applying sanctions allow for both realizing the goals and reproducing the infrastructure specified in the adaptation cycle.

The *adaptation* cycle takes care of the specification and adjustment of the infrastructure enabling discipline. To this purpose, examination researches the operations of the bodies. Using preset goals as criteria, it can analyze and evaluate the data resulting from its research. Dependent on evaluation, the current infrastructure enabling discipline can be adjusted on the basis of analysis and synthesis. Analysis, synthesis, and examination are linked to allow for the transformation of the infrastructure enabling the realization cycle.

In the disciplines, the adaptation and realization cycles are connected. Given contingent goals, the adaptation cycle specifies the conditions for the realization cycle. As argued, analysis, synthesis, and examination are the main activities in this cycle. The "output" of these activities consists of the infrastructure enabling the activities constituting the realization cycle.

The infrastructure resulting from analysis and synthesis, functions as "input" to the realization cycle, i.e., it specifies how the goals are to be realized. In the realization cycle, the operations of the bodies are regulated to both realize the goals and reproduce the infrastructure specified by the adaptation cycle. The regulated operations of the bodies are the "output" of the realization cycle. This "output" functions as "input" for the activities of the adaptation cycle. The regulated operations of the bodies are researched by the examination and the data resulting from this research are used to adapt the structure.

Together, the activities constituting the disciplines as described by Foucault constitute two interdependent cybernetic cycles adapting and producing the controllable and useful operations of the bodies entering their domain.

It is important to understand that in the case of the disciplines the task of confining the range of possible behaviors to a subset of desired (controllable and useful) behaviors is realized by a system of activities capable of explicitly, deliberately, and systematically changing its own structures in the course of time. Given this self-transforming capacity, we can understand that the disciplines as described by Foucault are only *one, historically first, instantiation* of a system developing itself into the huge variety of structures, systems, and methods for the management of organizations as we know them today. It is in this sense that the disciplines are "polymorphic" as Foucault (1977, p. 195ff.) argues.

The polymorphic character of the disciplines may, for instance, mean that in the course of time, the particular methods of hierarchical surveillance and applying normalizing sanctions develop into more sophisticated techniques that are less directive. For instance, in the organizations of today coaches may use techniques such as 360 degrees feedback as a method of monitoring and assessment. Something similar may be said about the methods and techniques underlying the analysis of space, bodily operations, and production processes and their synthesis into a structure.

In this sense, Fig. 9.2 is not yet a depiction of the general cybernetic structure of "discipline" in organizations. It rather depicts the cybernetic structure of eighteenth century disciplines as described by Foucault.

To formulate a general cybernetic model holding for both Foucault's description of eighteenth century disciplines and modern organizations, we need to abstract from particular methods and techniques used in the eighteenth century and concentrate on the underlying cybernetic functions involved. Moreover, we must redefine the object of these functions.

As argued the underlying cybernetic functions are control, design, and operational regulation and their object is the behavior of organizational members. Given these functions and this object, it is possible to redraw Fig. 9.2 and come up with a figure depicting a general cybernetic model of managing behavior in organizations.

Figure 9.3 accommodates both Foucault's description of the disciplines and the manifold of management systems in modern organizations.

To regulate the behavior of organizational members, control sets and adjusts the goals. It defines what controllable and useful mean as well as what the goals of the production processes are. Both the methods used to realize the control function and the goals resulting from it may change over time. For instance, in the case of the instantiation of the disciplines described by Foucault, almost "machine-like" operations of "bodies" geared to more or less complex but predictable production processes seemed to be desired. In the case of modern instantiations of discipline, for instance, in "knowledge intensive organizations," "creative," "self-steering,"

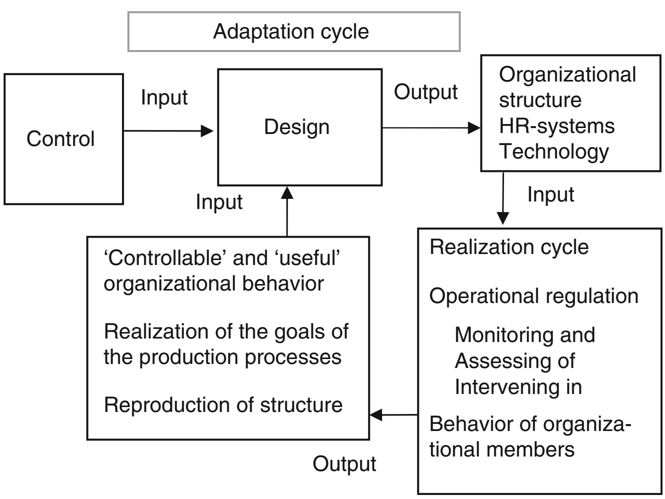


Fig. 9.3 General cybernetic model of confining behavior in organizations

"professional," or "multi-skilled" behavior geared to complex and less predictable production processes seem to function as interpretations of what controllable and useful mean.

Given the definition of these goals, design defines the organization's infrastructure. This infrastructure enables both the realization of the goals of the production processes and its regulation (control, design, and operational regulation). Once again, the methods by which infrastructures are designed as well as the infrastructures resulting from the application of these methods are contingent and may develop over time. For instance, in the case of the disciplines as described by Foucault, they have to facilitate hierarchical surveillance and the application of normalizing sanctions. In modern organizations, the structure may be socio-technical, designed to "mobilize" human resources, creating chances for their participation in and contribution to the adaptation and realization of organizational goals. Together, control and design define the conditions for the realization of the production processes and their regulation.

Operational regulation regulates both the realization of the goals specified by control and the reproduction of the infrastructure defined by design. It takes care of the production of controllable and useful behavior by means of monitoring, assessing, and intervening in behavior. Just as in the case of control and design both the methods and the results of regulation may vary. In the case of Foucault's disciplines, hierarchical surveillance and applying normalizing sanctions are considered pivotal. In modern organizations, we encounter all kinds of systems for monitoring, assessing, and intervening in the behavior of organizational members. For instance, these systems may depart from principles stemming from scientific outlooks as far apart as behaviorism and constructivism. They may be based on the presupposition of man as a basically opportunistic or as an intrinsically social being. And they may cover the entire spectrum of management styles ranging from highly "directive" to "enabling" or "coaching."

With Fig. 9.3 and its explanation, we have a set of general cybernetic principles of managing behavior in organizations. These principles underpin both the eighteenth century disciplines as described by Foucault and current systems for management in organizations. Moreover, they explain how it is possible that instantiations of "managing organizations" change and develop over time.

Viewed from a cybernetic perspective, organizations are managed by means of an underlying invariant structure. In this structure, the control, design, and operational regulation function are related to make behavior controllable and useful. The relation between these functions is such that, in the first place, they allow for realizing the goals for the behavior of organizational members set by control and reproducing the infrastructure specified by design. In the second place, the relation between these functions allows for transforming – and possibly improving – these goals and infrastructure. As such, the three related functions contribute to an understanding of the transformation of the disciplines, giving rise to the variety of systems for managing organizations we know today.

9.3.2 A Social Systemic Analysis of the Disciplines

On the basis of the previous step, we have gained an insight in the general cybernetic principles underpinning the management of behavior in both the eighteenth century disciplines and the organizations we know today. However, these principles do not explain how organizations *as a particular type of social systems* manage the behavior of their members. In this section, we answer this question by generalizing the disciplines as described by Foucault in terms of social systems theory. More in particular, we show that just as the panopticon has the function to discipline the behavior of the bodies entering its domain, decisions and decision premises have the function to *structure* behavior of organizational members.

To understand the functional similarity between the panopticon and the elements and structures of organizations, it is useful to revisit Foucault's explanation of the panoptic principle.

Given a set of desired behaviors (the goal of normalization) and an initial training, the prisoners are incarcerated in their cells, entering the "platform for visibility." Because they know that their behavior can *possibly* be monitored and punished or rewarded by the guards in the tower and because they do not know whether their behavior is actually observed or not, the prisoners behave *as if* they are *actually* observed, even if there are no guards present in the tower. The prisoners become the guards of their own behavior. They are disciplined to discipline their own behavior. Thus, combining "visibility" with "normalization," the panopticon functions as a "platform for discipline."

Now, to show why organizations as social systems have a functional dimension that is quite similar to that of the panopticon, we have to go back to the discussion of decisions and decision premises in the chapter on Luhmann's theory of social systems (Chap. 4).

There, we argued that decisions are a particular type of communications. Decisions are communications communicating a selection as a selection. Each decision communicates that from a range of options, one is selected and the others are rejected. Now, decisions may be attributed to organizational members who may be held "accountable" for selecting this option and rejecting the other ones. Given this responsibility, organizational members may be punished or rewarded for making the decisions attributed to them.

Moreover, we argued in Chap. 4 that in organizations – in principle – *all* behavior can be treated and attributed *as* decisions irrespective of the question whether it was actually intended as such. Whether behavior *counts* as a decision depends not so much on the behavior or its intentions, but rather on other communications treating it *as* a decision. Even if I as an organizational member act routinely, this action may be treated in the organization *as if* it were a decision. Once this happens, it can be attributed to me, and I may be held accountable for it.

As organizational members, we all know this. We all implicitly or explicitly anticipate on the possible treatment of our actions as decisions for which we may be held accountable and for which we may be punished or rewarded. As a rule, this anticipation tacitly guides the selection of our actions. Given my tasks and roles in the organization, I do what I expect that is expected of me. However, sometimes, for instance, if I am about to overstep an organizational guideline or if roles conflict, I become acutely aware of the possibility that my behavior may be treated as a decision for which I may be held accountable. If this happens, it may, for instance, withhold me from violating the guideline or, if it does not, I may try to cover my back, for instance, by decreasing the organizational visibility of my action or by inventing a justification for my behavior.

In this sense, we as organizational members are like the prisoners in the panopticon. Just as the prisoners know that their behavior may *possibly* be observed and therefore act *as if* their behavior is *actually* observed, we as organizational members know that our behavior may *possibly* be treated as a decision and therefore act *as if* our behavior will *actually* be treated as a decision. Because in organizations behavior *may* be treated as decisions that *may* be attributed to their members who *may* be held accountable for their choices, organizations function as a kind of "platform for visibility."

That organizations also function as a kind of "platform for normalization" can be argued if we take their decision premises into account. In the panopticon, the prisoners discipline themselves to discipline their own behavior. To this purpose, they at least have to know what counts as relevant behavior and more specifically what relevant behavior is desired of them. Given this knowledge, the possibility of being observed and punished or rewarded suffices to discipline the prisoners to discipline their own behavior.

Once again, organizations seem to have a functional dimension comparable to that of the panopticon. In Chap. 4, we argued that decision premises in organizations have two main functions. In the first place, they condition what can count as a decision in an organization. For instance, in organizations, "giving orders" is decision behavior that is tied to decision premises such as membership and particular communication pathways. These decision premises function as a heuristic for delimiting circumstances and sets of behaviors as "organizationally relevant," i.e., as behavior that may be treated as a decision.

The second main function of decision premises is that of enhancing the probability of particular organizationally desired behavior. As we explained in Chap. 4, decision premises neither "cause" nor "determine" behavior. They just increase the probability of the occurrence of desired behavior. They do this by focusing communication in organizations on the question whether behavior conforms to or deviates from the "desiderata" specified by decision premises.

For instance, the goals of the department I work for function as decision premises in my organization. They "prepare" a focus for departmental communication on the question whether my behavior conforms to or deviates from these goals. As I anticipate on this focus, the probability increases that my behavior takes these goals – i.e., takes the desiderata specified by these decision premises – into account. Of course, I may still attempt to undermine the realization of these goals, yet I know that this may be communicatively treated as "sabotage"; and as a decision for which I may be held accountable.

By focusing organizational communication on the question whether actual behavior conforms to or deviates from the particular desiderata specified by them, decision premises enhance the probability that future behavior takes into account what is desired. They condition organizational members to look at their own behavior and to that of the others with the eyes of what is organizationally desired.

Just as the prisoners in their cells look at their own behavior with the "eyes" – i.e., the norms for desired behavior – of the (possible) guards in the tower, organizational members look at their own behavior with the "eyes" – i.e., the decision premises defining desired behavior – of the organization. Thus, organizations, by means of their decision premises have a function that is quite similar to the "platform for normalization" instantiated by the panopticon.

So, we can learn that organizations, because of the decisions they consist of, have a functional dimension that is comparable to the *platform for visibility* instantiated by the panopticon. They can in principle treat any behavior as a decision that can be attributed to someone who can be held accountable and punished or rewarded for it.

Moreover, decision premises allow organizations to focus communication on the conformity or non-conformity of behavior to particular sets of desired behaviors. In this way, they increase the probability that organizational members look at their own behavior with the "eyes" of the organization and confine it to what is specified as desired. Thus, organizations have a functional dimension comparable to the *platform for normalization* incorporated by the panopticon.

Just as the panopticon disciplines the prisoners to discipline their behavior, decisions and decision premises structure the behavior of organizational members. Given the decisions and decision premises organizations consist of, organizational members reflexively monitor their own behavior (1) as if it can be treated as decisions for which they may be held accountable, and (2) focus on its conformity or deviance from established decision premises.

Now, it is important to highlight that in spite of the functional similarity between the panopticon and organizations, organizations are not the same as the panopticon.

As has been explained, the panopticon is a *particular* infrastructure designed to discipline the bodies entering it. As such, it has a central tower, peripheral prison cells, prisoners, guards, etc. Even if this infrastructure is generalized, e.g., in the form of managers monitoring the behavior of workers by means of camera's or workflow software, it remains one type out of many possible types of infrastructures. Organizations, on the contrary, are a *general* type of social systems allowing for many different types of infrastructures, including that of the panopticon.

Therefore, if we say that organizations have a functional dimension comparable to that of the panopticon, we make the quite general point that organizations structure the behavior of their members by the possible attribution of their behavior as decisions (platform for visibility) and the focus on conformity or deviance of attributed decisions from pre-established decision premises (platform for normalization). Given this possibility of structuring the behavior of organizational members, organizations can have different types of infrastructures. Therefore, although organizations (as social systems) have a functional dimension comparable to that of the panopticon and although every panopticon is an organization, not all organizations are a panopticon.

9.3.3 Cybernetic and Social Systemic Principles Underpinning Discipline in Organizations

From the analyses made above, we can learn that organizations manage the behaviors of their members by means of (1) cybernetic principles: control, design, and operational regulation, and (2) social systemic principles: decisions and decision premises making organizations into platforms for visibility and normalization.

However, in this book, we consider organizations as social experiments with meaningful survival. In them, the cybernetic principles underpinning the experimental "arche" and the social systemic principles underpinning the social "arche" are related.

Therefore, to finalize the generalization of the disciplines into a model explaining how organizations as social systems conducting experiments manage the behavior of their members, we need to relate the cybernetic to the social systemic principles. This is the topic of this section.

Organizations as social systems have decisions as their elements. Because in principle all behavior can be treated as if it is a decision and can be attributed to organizational members who can be held accountable, organizations have a functional dimension comparable to the *platform for visibility* instantiated by the panopticon. Decision premises specify what behavior is considered organizationally relevant and what behavior is desired. By focusing communication on *what* behavior is relevant and *what* behavior is desired, decision premises decrease the probability of undesired and increase the probability of desired behavior. As such, decision premises function as a platform for "normalization" of behavior. Together, decisions and decision premises provide organizations with the possibility of structuring the behavior of their members.

Decisions and decision premises also provide the connection between the social systemic and the cybernetic dimension of confining behavior in organizations. The cybernetic dimension entails that that in organizations, behavior is confined to a region of desired behaviors by means of control, design, and regulation.

The control activity defines the goals. It decides what type of behavior is desired as controllable and useful and it defines the goals of the production processes. As we know from the chapter on Luhmann's social systems theory, goals are a particular type of decision premises, focusing communication on what is considered relevant and desired behavior. In these terms, the "cybernetic" activity of control produces decisions about goals that function as decision premises structuring organizational behavior. In organizations, there will be decision premises for the activity of goal setting as well. Therefore we get as decision premises: premises governing the activity of goal setting and the goals set by the control activity.

The design specifies the organization's infrastructure. This infrastructure introduces a set of decision premises conditioning production processes and their regulation. For instance, it defines how production processes ought to be executed, or who can be held accountable for what, or it establishes ways to monitor, assess, and correct behavior of organizational members. In organizations, there also will be decision premises governing the design activity (e.g., BPR, Lean Production, or de Sitter's socio-technical approach can be regarded as sets of decision premises for designing organization structures, cf. de Sitter, 1994). As a result of the design activity, we get infrastructures that function as decision premises for organizational decisions.

Operational regulation produces decisions realizing its sub functions: "monitoring" and "assessing" decisions. Moreover, it produces "intervention decisions" to "normalize" organizational behavior. To this purpose, regulation needs the decision premises specified by the control and design function. It uses goals, organization structures, and systems as decision premises for monitoring, assessing, and normalizing behavior. Regulation can be regarded as deciding to actually intervene in the behavior of organizational members in order to reinforce or adapt decision premises that, in turn, condition further decisions. For instance, Peter's recent demotion re-enforced existing decision premises concerning departmental goals, influencing the decisions of his colleagues.

In this way, the decisions constituting organizations and the decision premises structuring the production of decisions become the vehicle for the variables, goals, structures, and systems produced by the activities constituting the (generalized version of the) disciplines. If we want to know how organizations as social systems conducting experiments manage the behavior of their members we can point to the decisions instantiating the control, design, and regulation activities, the decision premises guiding and resulting from these functions, and the decisions conditioned by these decision premises.

9.4 Why are the Disciplines Disquieting?

In the introduction to this chapter, we suggested that the disciplines are a vivid example of poor survival. However, we were too busy uncovering the general cybernetic and social systemic principles underpinning the management of behavior in organizations to pay attention to the question *why* the disciplines as described by Foucault exemplify poor survival and what implications this has for realizing the rich modality of meaningful survival in modern organizations. In this section, we answer these questions.

By uncovering the disquieting features of the disciplines, we also prepare the ground for the next chapter. For, if the disciplines are so disquieting, and if the disciplines are one of the many possible instantiations of the general cybernetic and social systemic principles underpinning the management of human resources in organizations, it can be asked whether there also are less disturbing instantiations of these principles and what these instantiations look like. Unless we can find such instantiations, or more general, unless we can find the criteria to distinguish disquieting "discipline-like" instantiations of these principles from less disquieting or even beneficial instantiations, we run the risk of designing infrastructures systematically undermining the viability both of organizations and their members.

In this section, we discuss six disquieting features.

9.4.1 Trivialization

To explore why the disciplines are so disquieting, Foucault's description of what the disciplines *do* may be helpful. As stated before, according to Foucault, the disciplines are the methods making possible the meticulous control of the operations of the body, assuring the constant subjection of its forces, and imposing upon them a relation of docility-utility.

For example, space is separated into compartments allowing for easy surveillance. Bodily operations are broken down into small standardized operations that are easy to learn and can be combined into more complex wholes. Production processes are decomposed into smaller processes that can be performed by standardized bodily operations. Surveillance and sanctions are geared to the "blind" execution of standardized operations that are only a tiny part of standardized production processes.

If we phrase this goal of the disciplines in terms of Von Foerster's distinction between trivial and non-trivial machines (see Chap. 3, Sect. 3), it can be argued that the disciplines aim at trivializing the principally non-trivial behavior of the human beings entering their domain. From Chap. 3 we know that it is one of the characteristic features of trivial machines that they are analytically determinable. This means that an external observer can determine their input/output relation and predict their output given a particular input. This "ideal" of the predictability of input/output behavior is exactly the aim of the disciplines. The bodies subject to the disciplines – pupils, soldiers, prisoners, workers, etc. – are trained to perform particular operations (output) given particular commands (input). These predictable operations, in turn, are a part of a larger production system generating its output with predictability.

In short, producing docility and usefulness means producing trivial behavior, i.e., given a command, a particular bodily operation always follows, This trivial behavior is part of larger and more complex trivial processes (e.g., educational, rehabilitation, or production processes) that, in turn, further trivialize the operations of the bodies entering their domain. A first disquieting feature of the disciplines, then, is that they seem to shape us into something we neither are nor want to be: bodies operating as trivial machines, indifferent "cogs" in a larger production process producing us as indifferent "cogs."

Three additional disquieting features of the disciplines come to the fore if we look at *how* they manage to trivialize behavior. Generally spoken, they do this by means of infrastructures enabling hierarchical surveillance, the application of normalizing sanction, and examination.

9.4.2 Self-Trivialization

If we concentrate on hierarchical surveillance and applying normalizing sanctions – particularly in the form of the panopticon – we can say that the disciplines discipline people to discipline themselves.

This means that the persons subject to the disciplines, because of the ever-present possibility of surveillance and sanction, monitor their own operations with the "eyes" of their guards. These persons shape the perception and assessment of their own operations according to the standards required for docility and usefulness. Thus, the process of disciplining does not stop at the "outside." It rather reaches into the "inside." This process shapes the "minds" (and bodies, for that matter) of the people entering its domain. It engraves into their minds the standards for docility and usefulness as categories for the reflexive monitoring of their operations. This "engraving" takes place by means of a process very similar to operand conditioning, i.e., by the repeated application of a stimulus (command) followed by punishment or reward (surveillance and normalizing sanctions) dependent on the response (docile and useful behavior). It is by this kind of conditioning that the disciplines discipline the operations of bodies by reaching into the minds of the people that are their subject. That is why Foucault – turning around the Platonic dictum of the body as the prison of the soul – argues that the soul is the prison of the body. The disciplines reach into the minds of their subjects shaping the categories and norms for the reflexive monitoring of the operations of the bodies in such a way that they operate as trivial machines. This is the second disquieting feature of the disciplines. They not only trivialize the operations of bodies, but they do this by assuming possession of their subjects' minds shaping their categories and norms for reflexive monitoring. They condition their subjects to trivialize themselves.

9.4.3 Fear as a Prime Motivational Factor

Hierarchical surveillance and applying normalizing sanctions – and more in particular the panopticon – harbor the next disquieting feature. As argued, the prisoners in the panopticon behave as if they are monitored because they do not want to run the risk of actually being monitored. "Not wanting to run the risk" is a nice way of saying that the prisoners *fear* the consequences of being caught while deviating from standards set for their behavior. In this sense, the functioning of hierarchical surveillance and applying normalizing sanctions in the form of the panopticon is based on the emotion of *fear*. In organizations, this emotion may be exploited for the purpose of managing human resources in more or less subtle ways (e.g., threatening to withhold particular rewards). As organizations incorporate panoptic forms of management into their HR-tools, they become "systems of fear," introducing this emotion as a prime source of motivation.

9.4.4 A Science of Discipline

The examination introduces another disquieting feature of the disciplines. Above, we saw that the examination is a mechanism built into the disciplines contributing to their continuous sophistication and self-transformation. Aside from the disquieting point that "scientific research" is used for the sole purpose of producing docile and useful bodies, the examination contributes to the formation of a "body of knowledge" concerning both human behavior and the techniques for trivializing it. Based on this body of knowledge, goals, structures, and methods for trivializing minds and bodies may be both improved and transferred to new domains of application. So, what is disquieting about the disciplines is that trivializing human behavior is not a stationary process, remaining at the same level of sophistication or staying within particular domains of application. By means of the accumulation of knowledge resulting from the examination, its techniques may become more adapted, colonizing new societal domains.

9.4.5 Detachment and Lack of Involvement

Yet another disquieting feature of the disciplines is that its success involves "bodies" performing small and simple tasks that can be learned easily. To phrase this in the words of the chapter on organizational structures, the disciplines involve complex organizations with simple jobs. In these complex organizations, each member contributes its bit to the realization of the two types of goals distinguished earlier: producing docile and useful bodies and realizing the goals of the production processes.

Because the contribution of each of the members only involves a tiny part of the total production process, members run the risk of becoming "detached" from the larger goals they contribute to. As the "distance" between these goals and the tasks performed by each individual member grows, the chance that the "bodies" performing these tasks feel both involved with and responsible for these goals and their realization diminishes (see for, instance, Bauman 2005, p. 155ff., on inhumanity as a function of social distance).[3] In some cases, one may even stop

[3] Bauman quotes Milgram (1974): "Any force or event that is placed between the subject and the consequences of shocking the victim, will lead to a reduction of strain on the participant and thus lessen disobedience. In modern society others often stand between us and the final destructive act to which we contribute", and he then adds, "Indeed, mediating the action, splitting the action between stages delineated and set apart by the hierarchy of authority, and cutting the action across through functional specialization is one of the most salient and proudly advertised achievements of our rational society" (Bauman 2005, p. 155).

seeing the relation between one's contribution and overall goals. In less extreme cases, the experience of "being responsible": may be disseminated over organizational members to the extent that it becomes very hard to attribute and even seems to evaporate.[4] It is therefore a disquieting feature of the disciplines that they loosen the relation between overall goals and contributions to overall goals, undermining the individual's experience of involvement with and responsibility for these overall goals. As such, the disciplines may become an efficient vehicle for realizing even the most evil goals appearing to have nothing to do with individual actions.

9.4.6 Contingent and Minimal Goals

Perhaps the most important disquieting feature is the way the disciplines relate to goals. There are two things disquieting here. First, the disciplines are indifferent to the goals of production processes. Second, the goal of the disciplines themselves – controllable and useful behavior – is a quite minimal goal.

Above, we discussed that in the case of the disciplines, we need to distinguish two types of goals: the goals of the production processes and the goal of the disciplines: producing controllable and useful behavior. With respect to these two types of goals, we noticed that there is a circular relation between them. Controllable and useful behavior is needed to realize the goals of the production processes and realizing the goals of the production processes is needed to produce controllable and useful behavior. Thus, the disciplines are a kind of "self-carrying" construct in which means become goals, and goals become means. However, it remains unclear whether this construct serves an overall goal and what its value is. This puts us on the track of the indifference to and minimalism of goals implied in the disciplines.

As the production of means becomes the goal and vice versa, the disciplines not only loosen the relation between overall goals and individual feelings of involvement and responsibility, they also blot out talk about and reflection on the goals of the production processes. As long as these processes contribute as a means to producing docile and useful bodies, it does not matter what their goals are. Thus, in principle, the disciplines are indifferent to the goals of the production processes, i.e., these goals are *contingent* however base or criminal they may be.

And if we look at the goal of producing docile and useful bodies, we see that the disciplines produce trivialized beings, realizing contingent goals of production processes that, in turn, contribute to their further trivialization. This, at least to us, appears to be a far cry from what we as human beings experience ourselves to be or hope to

[4] Bauman calls this "free-floating responsibility," "We may surmise that the overall effect of such a continuous and ubiquitous responsibility shifting would be a *free-floating responsibility,* a situation in which each and every member of the organization is convinced, and would say so if asked, that he has been at some else's beck and call, but the members pointed to by others as the bearers of responsibility would pass the buck to someone else again. One can say that *the organization as a whole is an instrument to obliterate responsibility*" (2005, p. 163).

become. In this sense, the disciplines offer only a *minimal* picture of our development as human beings and fail to pay attention to the question what its value can be.

So, what we have now is a list of features that make the disciplines a rather disquieting instance of poor survival. For the purpose of clarity we list the disquieting features below, albeit in a different order.

1. The disciplines are indifferent to goals: contingency, minimalism
2. The disciplines reduce feelings of involvement and responsibility
3. The disciplines rely on fear as a prime motivator of action
4. The disciplines aim at trivializing – fundamentally non-trivial behavior – of their subjects
5. The disciplines "invade" the minds of their subjects by shaping their categories and norms for reflexive monitoring behavior by means of operand conditioning
6. The disciplines invest in research aimed at developing methods for trivializing behavior.

Although it is perhaps possible to extend this list of disquieting features, there is a problem that seems to be the most worrisome of all.

Until now, we spoke of the disciplines as a particular historical phenomenon, including a number of disquieting features involving the degradation of the humanity of those entering their domain. Of course, these features are bad enough by themselves. But things may get worse.

In the previous section we showed that if we examine the principles underlying the disciplines, we find cybernetic and social systemic principles that seem to hold for *all* the management of human behavior in organizations. Dependent on the *particular* goals, structures, and systems selected to realize these principles, we get organizations that more or less resemble the disciplines including the six disquieting features mentioned above. However, on the basis of our analysis, we did not find criteria to assess and design organizations that do *not* show the disquieting features listed above.

What is disquieting here is that we found cybernetic and social systemic principles that *possibly* are instantiated by infrastructures more or less resembling Foucault's description of the disciplines (including all their disquieting features). And what is more, if we look around us in the organizations we work for, we may find that to some extent such infrastructures actually are in place. So the question urges itself whether these cybernetic and social systemic principles *necessarily* imply "discipline-like" instantiations. For, if this is the case, the management of modern organizations has nothing more to offer than the bleak picture of the disciplines including the negative features mentioned above.

This is why there is a *seventh feature* that is most disquieting of all. Unlike the six previous features, it is not about the disciplines as a particular and historically contingent instantiation of the cybernetic and social systemic principles underpinning the management of human resources in organizations. It rather is about these principles themselves and it asks whether they *necessarily* lead to infrastructures resembling the disciplines.

If they do, this means that at the heart of *all* infrastructures there are principles that necessarily trivialize behavior, invade minds by shaping categories for reflexive monitoring, feed on fear, colonize new parts of society in ever more sophisticated forms, loosen the bond between overall goals and individual feelings of involvement and responsibility, and are indifferent with respect to overall goals.

Therefore, the question becomes: "Do the cybernetic and social systemic principles underpinning the management of human resources in organizations necessarily lead to infrastructures more or less resembling the disciplines or are other infrastructures possible that do not have the disquieting features of the disciplines?"

In the next two chapters, we hope to show that this is not the case. We hope to argue that these principles are "*pharmakon*." This means that they – just like pharmaceutics – can be used to either harm or benefit their users. Moreover, we hope to provide a perspective on the design of organizational structures that do not have the disquieting features of the disciplines, but rather enable the viability of organizations and their members.

References

Bauman, Z. (2005). *Modernity and the Holocaust*. Cambridge, UK: Polity Press.
Foucault, M. (1975). *Surveiller et punir: Naissance de la prison*. Paris: Gallimard.
Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Sheridan, A. (Trans.). London: Allan Lane.
de Sitter, L. U. (1994). *Synergetisch produceren*. Assen: Van Gorcum.

Chapter 10
Towards Rich Survival: Aristotle

10.1 Introduction

The previous chapter was about poor survival. To be more precise, it was about Foucault's description of the disciplines. We argued that attached to them there are six disquieting features, all related to the instrumental use of trivialized behavior to realize contingent goals of organizational production processes. By analyzing the disciplines, we arrived at a set of general cybernetic and social systemic principles underpinning all management of organizational behavior. The question arose whether the application of these principles in organizations *necessarily* leads to "discipline-like" forms of management. This question is disquieting because an affirmative answer would mean that *all* organizations trivialize the behavior of their members, using it as an instrument to their contingent and possibly evil ends.

Against this unattractive picture of organizing and organizations, we think that rich survival is possible. This means that organizations do not necessarily trivialize behavior and that the ends of their primary processes are not necessarily contingent.

Of course, we do acknowledge that management in organizations involves technical rationality and deliberate design to realize explicitly selected goals. We also acknowledge that technical rationality and deliberate design can and often do lead to cynical structures and methods for managing human resources, "de-humanizing" the unfortunate entering their domains. We only need to open our eyes to the organizations we work for to recognize examples of the disquieting features described in the last section of the previous chapter. Therefore, we do not deny that poor survival – and extreme cases of poor survival for that matter – really exists.

However, we do not admit that the cybernetic and social systemic principles *necessarily* lead to the disquieting effects of the disciplines. As indicated, we rather think that these principles are *pharmakon*. That is, we think that they *can* be instantiated by infrastructures supporting poor survival. However, we also think that it is possible to design "alternative" infrastructures, supporting what we have called rich survival. This chapter and the next one are devoted to finding out what organizational structures can support it.

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_10, © Springer-Verlag Berlin Heidelberg 2010

To find this out, we first need a better description of "rich survival." Until now, we have contented ourselves with the definition presented in Chap. 1. Roughly, rich survival entails that organizations contribute to society enabling human beings to live a fulfilled life, both inside and outside the organizations they work for. In order to specify what this contribution looks like and how organizational structures can support it we first need to explore what it means to live a fulfilled life. This is the aim of the present chapter.

An easy way to start this exploration is to have another look at the disquieting features listed in the previous chapter (see Sect. 9.4). Each one of these features in the first place may be regarded – in some way or another – as an *effect* of the organizational infrastructure instantiating the disciplines. Some of them are effects the disciplines *aim* to produce. Others are *side effects* factually produced by them. In the second place, given that they are "disquieting," each of these effects may be regarded as *un*desired. This means that organizational infrastructures functioning as alternative to the disciplines:

1. *Should not* be indifferent to goals (goals should make a difference)
2. *Should not* reduce feelings of involvement and responsibility (feelings of involvement and responsibility should be mobilized and cultivated)
3. *Should not* rely on fear as a prime motivator of action (there should be other motivators of action)
4. *Should not* aim at trivializing – basically non-trivial – behavior (should capitalize on the learning history of organizational members)
5. *Should not* shape our categories and norms for monitoring behavior by operand conditioning (other ways of learning should be envisaged)
6. *Should not* invest in tools for further trivializing behavior (management research should capitalize on the non-triviality of behavior)

By stating what alternative organizational infrastructures *should not* do, we advance only a small step. For instance, if they should not introduce indifference to goals, how, then, should goals make a difference? Or, if fear should not be a prime motivational force, what, then, should motivate organizational members? Or, if operand conditioning should not be the way to develop categories and norms to reflexively monitor behavior, how, then, should we develop such categories? In short, by negatively formulating requirements, we only demarcate the area of effects we *do not* want. To specify *positively* desired effects to be realized by alternative infrastructures, we need a perspective on human behavior allowing for the articulation of what it, for instance, means to pay attention to goals, to introduce feelings of involvement and responsibility, or to develop categories for reflexive monitoring by other means than operand conditioning.

In a way, we do not have to search very far to find such a perspective. We can start by looking at ourselves and at our own lives. Most of us will find the features attached to the disciplines disquieting. Maybe we do not have the conceptual tools and criteria to determine exactly what is wrong with them, but we very probably are aware that they negate aspects of our lives we find valuable.

For instance, we all "know" that goals matter. We all have the experience of asking ourselves whether pursuing this or that course of action is worthwhile; not only relative to some limited goal, but also in the light of our life with others. We all have the experience of wanting to be involved in and responsible for things we find worth pursuing and we all have the experience of the motivational energy flowing from this.

Perhaps we also have the experience of being reduced to a "case," of being reduced to a quasi object of operand conditioning, or of being treated as a kind of trivial machine, producing the "right" output given some input independent of one's own learning history or personal development. For the purpose of the argument we want to unfold here, each and every of these experiences is worthwhile, because it functions as a kind of starting (and end) point for a reflection on what we find valuable in our life and, included in it, in our work.

Given this starting point, we need a framework doing justice to the kind of experiences mentioned above and allowing for their conceptual articulation. This framework, then, can function as a background for the formulation of what it means to live a fulfilled human life.

We have found this framework in Aristotle's ethics. In the *Nicomachean* and *Eudemian Ethics*, Aristotle reflects on the fulfilled human life and its relation to human capacities (such as reason and desire) and action. His ideas allow us to take the experiences we just talked about seriously. For Aristotle, goals do matter. In fact, his ethics is a reflection on the highest goal, the highest good, for man. Moreover, related to the question of the highest good, Aristotle pays attention to relevant dimensions of human behavior and their corresponding types of rationality. He points at the role of emotional dispositions in action, and comes up with a model for human learning and development that can function as an alternative to the operand conditioning implied by the disciplines. In this chapter, we will take Aristotle's *Nicomachean Ethics* as our text of reference (1991).[1]

To explain how Aristotle's theory of the fulfilled life provides a "counter-model" to the disquieting features attached to the disciplines, we organized the rest of this chapter into three sections.

Section 10.2 is devoted to Aristotle's conception of the "highest good for man." According to Aristotle, this "highest good" is: "living a fulfilled life." To explain what this means, Aristotle focuses on characteristically human capacities and their development into so-called virtues. Living a fulfilled life, he argues, implies (1) developing our characteristically human capacities into virtues and (2) exercising them in the best possible way.

Section 10.3 discusses the virtues related to the practical dimension of our existence. Three virtues are important here: moral virtue, practical wisdom, and skill. Particularly moral virtue and practical wisdom are needed to live a fulfilled practical life. In Sect. 10.3, we discuss these virtues, their development, and the way they are related to a fulfilled life. Finally, in Sect. 10.4, we show how Aristotle's

[1] In this chapter, we take our citations from the Oxford translation edited by Barnes (Barnes, 1991).

concept of living a fulfilled life can function as a "counter-model" to the disquieting features attached to the disciplines.

10.2 The Highest Good for Man: Basic Distinctions

In this section, Aristotle's conception of the highest good, i.e., "living a fulfilled life" – or *eudaimonia*, as he calls it – is at issue. In Sect. 10.2.1, we explain what Aristotle means by this. In the course of this explanation, it will appear that living a fulfilled life depends on the way we develop and use capacities that are characteristically human (such as our capacity for reasoning). In Sect. 10.2.2 we discuss these capacities, their development, and their relation to the fulfilled life. In this discussion, we show that to be able to live a fulfilled life, we need to develop our characteristically human capacities into so-called virtues. These virtues and their relation to *eudaimonia* are at issue in Sect. 10.2.3.

10.2.1 The Highest Good for Man: Living a Fulfilled Life

The *leitmotiv* in Aristotle's ethics is the question of the highest good for man. To explore what this highest good is, Aristotle distinguishes between things we do for the sake of something else and things we do for their own sake (see Book I of the Nicomachean Ethics). He argues that things we do for the sake of something else cannot be the highest good for man because we do these things because of the end they produce. Because this end motivates us to act as we do, we apparently value this end as a higher good. Now, with regard to this end, we can again ask, whether we want it for the sake of another "higher" end or for the sake of itself. Until we find an end that is an "end in itself," i.e., an end that we do not want for the sake of something else, we have not found the highest good.

Of course, this description does not yet tell us *what* the highest good for man is. In a next step, Aristotle tries to find possible candidates for what can be an end in itself. To this purpose, he discusses ends like "wealth," "pleasure," and "honor" considered valuable in his time (as a matter of fact, most of these ends still are appealing today).

For different reasons he dismisses each one of these as the highest good, as an end in itself. "Wealth" is not an end in itself, for it is what we can *do* with it that we actually desire. A life of "pleasure," in any case, its more "base" forms, makes us to the slave of our emotions and desires. And we only value "honor," if we are honored because we or our actions are considered as valuable (see also Hughes 2001, p. 24–25).

Instead, Aristotle comes up with his own candidate: "*eudaimonia,"* which according to commentators can be translated as "living a fulfilled life" (Hughes 2001, p. 22; Nussbaum 1997; Wolf 2002, p. 30ff). According to Aristotle, everything we do is ultimately directed at this goal. If we continue to ask someone, "What is it that motivates you to act in this way?" the ultimate answer will be, "because I think it enriches my life as a human being, because it adds to the fulfilled human life".

Now, if we accept *eudaimonia* as the highest good, the question becomes, what does it mean?

To explain this, take the example of a coffee machine. You may ask, what is the highest good for this machine? Probably the answer would be "to function well." So, you ask, what does it mean for the coffee machine to function well? The answer would, then, probably be, to do what it according to the design of its internal organization is set out to do: to produce well-made coffee. This is its "fulfilled life." Everything else it "does" (e.g., boiling water or percolating it), it does for this "reason."

Although experts on Aristotle probably will grit their teeth reading this example, a similar argument can be made for the class of living, and therefore, for human beings.

Aristotle regards living beings as beings that are internally organized to live their life in their own particular way. Based on their internal organization, different kinds of living beings have different "capacities" (capacity is "*dynamis*" in Greek). These capacities, enable them to live their lives as they do. For instance, based on their internal organization, plants have the capacity to reproduce, to feed, and to grow, disposing them to a "plant-like life." Based on their internal organization, animals have the capacity to reproduce, to feed, to grow, to move, to sense, and more or less, to have emotions, and to desire. These capacities dispose animals to lead a life we call "typically" animal-like.

Aristotle calls the internal organization of living beings and the capacities resulting from this their "soul" (*psychē*). For him, a soul is not some strange "ghost in the machine," "a separate thing which inhabits my body" (Hughes 2001, p.35). "Soul" rather refers to the way a living being is internally organized and the capacities it has, based on this organization.

Given its internal organization and the capacities based on it, i.e., given its soul, a living being is disposed to live its life in a particular way. Aristotle calls this the "*ergon*," the "work" or "function" of that being. As indicated, given their internal organization and their capacities, plants, for instance, are disposed to live a characteristically plant-like life and animals are disposed to live a typically animal life; this is their particular "*ergon*" or function.

Implied in this function, is a goal (*telos*). This goal is nothing less than living the life one is disposed to live given the capacities of one's soul in the most fulfilled way. Whether this goal is actually realized depends on the way a living being lives its life. Some plants, animals, or human beings will be able to realize the life to which they are disposed more fully than other beings of that type. In any case, the highest good for each of these living beings is to live the life it is disposed to live – given the capacities of its soul – in the best possible way.

If one would ask Aristotle, "What is the highest good for a living being?", "What is relative to that being an end in itself?" he would answer, "Living a fulfilled life:

eudaimonia". If one would then ask, "What does living a fulfilled life mean?" he would probably say, "To live in the best possible way the life to which one is disposed given the capacities of one's soul". This is the end, to which all other ends are ordered: the highest good.

This line of reasoning implies that the fulfilled life of for instance, an ant is quite different from that of a human being, for ants have capacities that are very different from those of human beings disposing them to an ant-like, instead of a human life. Given these differences, the fulfilled ant-life will look quite different from the fulfilled human life. Therefore, if we want to answer the question what the fulfilled life for human beings entails, we have to look at the capacities they have based on their internal organization. This is exactly what Aristotle does.

10.2.2 Characteristically Human Capacities and the Fulfilled Life

In his *Nicomachean Ethics* (as well as in other works), Aristotle lists quite a number of capacities based on our internal organization, such as the capacity to reproduce, to feed, to grow, to move, to sense, to imagine, to have emotions, to desire, to reason, to judge and to grasp some insight as true.

The question now arises, "Which of them are crucial for living a fulfilled *human* life?" Aristotle's answer is quite straightforward, "Those capacities which if they are realized in the best possible way are most characteristic of the humanity of human beings". For instance, realizing to the best of our ability the capacity for growth is not living a characteristically human fulfilled life. Plants may also do this.

Therefore, to get an insight in the fulfilled life for human beings, we have to know which of the capacities of the human soul are characteristically human.

According to Aristotle, these capacities are: (1) "*reason*," i.e., the capacity disposing us to know truth and (2) – as will be explained later, in a derivative sense – "*desire*," the capacity disposing us to "reach out" for something that is perceived as good. In his view, particularly "reason" is the capacity that sets of the best developed human being from the best developed animal.

For human beings, living a fulfilled life, therefore, means exercising in the best possible way our capacities for reason and desire, thereby living the life we are disposed to live given our characteristically human capacities.

To explain what it means to exercise our capacities for reason and desire in the best possible way, we need to make a distinction between two types of dispositions: *natural* and *habitual* dispositions.

As argued, based on their internal organization beings have capacities disposing them to behave in a particular way. For instance, given its internal organization sugar has the capacity to dissolve in hot water. This disposition is already there, given with its internal organization. It is what can be called a *natural disposition* of sugar. Sugar does not need to "learn" to dissolve in water. Based on its internal organization, it is naturally disposed to do so.

Now, if we turn our attention to our capacities for reason and desire, we can say that, according to Aristotle, we as human beings are *naturally disposed* to know truth and desire what "appears as good." However, if we look more closely, we can see that we, in the course of our lives, can develop our capacity for knowing truth and desiring what appears as good. For instance, dependent on one's upbringing, one can develop a desire for unhealthy food, disposing one to choose fast food over fruit. This desire for unhealthy food is not a natural disposition. It is not a disposition that is naturally given with one's internal organization as a human being. It is rather an *acquired* or *habitual* disposition, a disposition one develops by eating and enjoying unhealthy food. According to Aristotle, reason and desire are capacities that naturally dispose us to develop habitual dispositions. They are capacities that allow us to develop particular habitual dispositions dependent on the way we actually "exercise" – both in the sense of "use" and "train" – them.

Given the distinction between natural and habitual dispositions, we can say that given our capacities for reason and desire, we are naturally disposed to develop habitual dispositions. Moreover, dependent on *how* we use and train our capacities we develop more or less good (virtuous) or bad (vicious) habitual dispositions.

Now, the natural goal of our capacity of reason (including the ability to grasp first principles, to reason and to make judgments) is to know the truth about things. By exercising this capacity according to this goal, we become habitually disposed to know the truth, i.e., we become disposed to judge that something is the case, if it actually is the case, and to judge that something is not the case, if it is not. However, we may also neglect the exercise of this natural capacity or train it incorrectly. In this case, we become habitually disposed to make erroneous judgments about what is the case or not.

In a similar way, our "desire" may be trained. The natural goal of our desire is to aim for something that is perceived as "good." By using and training our desire both according to this goal and in accordance with what a well-developed reason judges as "good," we become habitually disposed to desire what is actually good, and to avoid what is not. However, we also may train our desire badly. In this case, we become habitually disposed to desire as good what is actually not good and to shun as bad, what is actually good.

So, according to Aristotle, both reason and desire may be cultivated in accordance with their respective natural goals, i.e., they may be cultivated in such a way that we become habitually disposed to realize these goals in the best possible way. However, we may also acquire "bad" habitual dispositions that stand in the way of realizing these goals.

This possibility of developing "good" and "bad" habitual dispositions enables Aristotle to define the concept of *virtue*. He defines a virtue as a habitual disposition, disposing us to realize in the best possible way the goal connected to the function of a capacity. Virtues are of central importance to Aristotle's ethics.

The highest "good for man," "*eudaimonia,*" "living a fulfilled life," or "functioning in the best possible way," for human beings, then, means (1) *developing* our virtues, i.e., developing habitual dispositions in the best possible way, disposing us

Table 10.1 Natural capacities, habitual dispositions, and virtues

The highest and most characteristic human natural capacities:
reason and desire (in a derivative sense)
. . . are naturally disposed to:
Knowing truth and desiring what is perceived as good
. . . and allow human beings to develop the habitual disposition to
know as true what is actually true and desire as good what is actually is good
. . . thus establishing virtues, being habitual dispositions, disposing human beings to realize in the best possible way the goal of their characteristic capacities by actually:
knowing truth and desiring what is perceived as good

to realize what our characteristic capacities are set out to do, and (2) *operating* according to the thus developed virtues. Table 10.1 summarizes the distinctions made thus far.

10.2.3 Virtues Involved in Eudaimonia

To provide a general overview of the virtues involved in the fulfilled life, it is relevant to inspect the different capacities of the soul more closely.

To this purpose, Aristotle distinguishes two "parts" (or functions) of the human soul: the "irrational" and the "rational" part.

The irrational part consists of two "subparts": the "vegetative" and the "desiderative" part.

The vegetative part allows for reproduction, nutrition, and growth. This part we have in common with all other forms of life. Because it is not distinctive of being human, and because *eudaimonia* for humans is a kind of operation involving the development of distinctive *human* capacities, Aristotle does not as such include the vegetative part of the soul into his account of *eudaimonia*.[2]

The desiderative part of the soul allows for the appetites, emotions, and desires. As a principle of appetite and desire, it "moves us from within towards the object of desire". According to Aristotle, the desiderative part is not rational by itself. Still, it can "participate" in the rationality of the rational part in the sense that it can be "controlled, persuaded, and shaped by" it (Sherman 1989). Although not rational by itself, the desiderative part of the soul can be cultivated to desire what is right by means of habituation in accordance with reason. Because of this relation to reason and because desire is needed for action, Aristotle includes the desiderative part of the soul in his account of *eudaimonia*. He calls the virtues connected to the desiderative part of the soul the moral virtues.

[2] Of course, this does not mean that the vegetative has nothing to do with *eudaimonia*. For instance, health and nutrition can contribute to living a fulfilled life. However, by themselves they do not constitute it.

The second part of the soul is the rational part. It also consists of two "subparts": the "calculative" and the "theoretical" part.

The calculative part allows for knowing things insofar as they are variable. It allows us to deliberate and judge about things within our capacity to change and covers the domains of human "making" and "acting."

- Deliberation and judgment about *making* entails finding the most suitable means to produce some product (e.g., finding the means to build a ship). Aristotle calls the virtue related to deliberation and judgment about making "skill" (technē).
- Deliberation and judgment about *action* entails finding an appropriate instantiation, here and now, of what it means to live a fulfilled life (e.g., acting courageously in the face of the enemy). The virtue related to action is practical wisdom (phronēsis). Someone who is practically wise has acquired the ability to deliberate and judge well about what actions contribute to *eudaimonia*.

The theoretical part allows us to know things insofar as they cannot be otherwise. It aims at certain and necessary knowledge. This knowledge comprises:

- basic principles for demonstration (for instance, the principle of non-contradiction). "Understanding" (nous) is the virtue allowing us to grasp these first principles.
- conclusions from highest principles within particular regions of being (e.g., of living things, or of inanimate things). The virtue of knowing these conclusions is "science" (ēpistemē). As there are different regions of being, there are also different sciences (e.g., biology, or physics).
- the highest principles of being as such and propositions demonstrable from these principles. Aristotle calls the virtuous disposition related to this knowledge "wisdom" (sophia). According to Aristotle, the virtue of wisdom has a direct relation to *eudaimonia*.

Table 10.2 captures the distinctions made thus far. In the table, we find five virtuous dispositions: moral virtue, skill, practical wisdom, science, understanding, and wisdom. However, only moral virtue, practical wisdom, and wisdom have a direct relevance for *eudaimonia*. The reason for this is that *eudaimonia* means realizing the *highest* of our human capacities in the best possible way. The best possible realization of our desiderative capacity is moral virtue. Practical wisdom – and not skill – is the highest realization of the calculative part of the soul. For, as will be explained later, practical wisdom should guide the application of skill. Wisdom is the highest realization of the theoretical part of the soul. In a sense that not will be explained here, it comprises both science and understanding.

Therefore, applied to the domain of human practice, *eudaimonia*, in the first place, means developing the calculative part of our rationality in such a way that we are disposed to deliberate and judge well about what it means to live a fulfilled life in the particular situations we find ourselves. This means developing practical wisdom. In the second place, it means, developing our emotions and desires in accordance with practical wisdom. In this way, we are disposed to desire what is right (because it is right). Once we develop moral virtue and practical wisdom, we

Table 10.2 Aristotle's basic distinctions regarding the human soul

'Part' of the soul	Function (ergon)		Virtuous disposition
Vegetative	Reproduction, nutrition, growth		–
Desiderative	Moving the agent from within towards an apparent good		Moral virtues (aretē)
Calculative	Knowing truth about things in sofar as they are variable and within our capacity to change applied to the operation of...	making (poiēsis)	Skill (technē)
		acting (praxis)	Practical wisdom (phronēsis)
Theoretical	Knowing truth about things in sofar they cannot be otherwise based on...	Grasping first principles	Understanding (nous)
		Demonstration from first principles in particular regions of being	Science(s) (ēpistemē)
		Grasping and demonstration from first principles of being	Wisdom (sophia)

are disposed to "act well" and "acting well" (*eupraxia*) is the culmination point of living a fulfilled life in the practical dimension of our existence. It is the first constituent of *eudaimonia* (see Table 10.3).

Applied to the domain of human knowing *eudaimonia* means acquiring the most "deep" and "inclusive" knowledge. According to Aristotle, this means knowing the first principles ("deep" knowledge) of being ("inclusive" knowledge). The virtue corresponding to this "deep" and "inclusive" knowledge is wisdom. Once we develop wisdom, we are disposed to contemplation (*theoria)*, which is the culmination point of living a fulfilled life in the theoretical dimension of our existence. It is the second constituent of *eudaimonia* (see Table 10.3).

Given these considerations, we get the three virtues connected to living a fulfilled life. Moral virtue and practical wisdom relate to living a fulfilled life in the practical dimension of human existence. Their end or culmination point is acting well ("*eupraxia*"). Wisdom relates to *eudaimonia* in the theoretical dimension of human existence. Its end is contemplation ("*theoria*"). Because this chapter is about human behavior in organizational contexts, we limit our discussion of the highest good for man to the practical dimension of our lives and its highest good "*eupraxia*." "*Eupraxia*" and its requirements are the topic of the next section.

Table 10.3 Virtues related directly to *eudaimonia*

Part of the human soul	Desiderative	Calculative	Theoretical
Principle of...	moving the agent from within towards the apparent good	knowing truth about things insofar as they are variable and within our capacity to change applied to the domain of human acting (praxis)	knowing truth about things insofar as they cannot be otherwise (are necessary) based on grasping of and demonstration from first principles of being
Virtuous disposition	Moral virtue(s)	Practical wisdom	Wisdom
Constituent of *eudaimonia*	Acting well (*eupraxia*)		Contemplation (*theoria*)

10.3 Eupraxia: Moral Virtue, Practical Wisdom, and Choice

According to Aristotle, "*eupraxia"* means to "act well." One might be tempted to say that acting well means "doing what is right" or "doing the right thing in the given circumstances." According to Aristotle, this is insufficient. For, suppose that someone does the right thing purely by chance, without choosing it. This, we would surely not praise as an instance of acting well. Somehow, actors should not "stumble" across the right thing. To act well they must choose it. Therefore, choice has to be involved in the act if it or the actor is to be called virtuous.

However, mere choice is still insufficient. For it may well be that someone chooses to do the right thing, yet chooses the act for the wrong reasons (e.g., being kind to reap financial benefits). In such cases, we would call neither the act nor the actor virtuous. Therefore, "acting well," does not only mean choosing to do the right thing, but also *choosing it for the right reasons*.

However, even choosing to act well and choosing the act for the right reasons is still insufficient for *eupraxia*. For, suppose that someone acts virtuously sometimes and viciously at other times. We would still not call such a person "virtuous." Apparently, some form of constancy is required. One must not only factually act well in some situations, but one must be *habitually disposed* to act well in all situations. Therefore, to act well, one must do the right things, one must choose to do them, choose them for the right reasons, and choose them from a morally virtuous habitual disposition or character (1105^{a}30–1105^{b}1).

A closer examination of *"eupraxia"* shows that "choice" is central to it. To live a fulfilled practical life one must *choose* well, *choose* well for the right reasons, and be habitually disposed to *choose* well for the right reasons.

Aristotle defines choice as a deliberate desire for some apparent good, involving both the desiderative and the rational part of the soul. For a choice to be right, one must: (1) be habitually disposed to *desire* to do the right thing, i.e., one must be *morally virtuous* and (2) be able to deliberate and judge well about what is the right thing in a given situation, i.e., one must be *practically wise*. It appears that "*eupraxia*" has three constituents:

1. the cultivation of our desires in accordance with reason to acquire the *morally virtuous disposition* or character for acting well;
2. the development of *practical wisdom*, i.e., of our capacity for reasoning and judging about contingent truths related to action to be able to choose the right things and to choose them for the right reasons;
3. the combination of moral virtue and practical wisdom in the *virtuous choices* needed for the virtuous acts constituting "*eupraxia*."

Therefore, to explain Aristotle's ideas on "*eupraxia*" we need to explain what these virtues entail, how they develop, and how they cooperate in virtuous choice. To this purpose, we first discuss Aristotle's notion of moral virtue. Second, we discuss his conception of practical wisdom. Third, we show how moral virtue and practical wisdom cooperate in the virtuous choices constituting "*eupraxia*." Finally, we go into the question of how moral virtue and practical wisdom, according to Aristotle, can develop and reach perfection in the course of our social lives.

10.3.1 Moral Virtue

Aristotle calls the perfection of the desiderative part of the soul moral virtue. He defines moral virtue as, "a habit concerned with choice, lying in a mean relative to us, this being determined by reason and in the way the man of practical wisdom would determine it" (1106b36–1107a2). To understand what this definition means, we follow Hughes (2001) and address its different "parts."

Aristotle defines moral virtue as "concerned with choice." Above, we argued how important the concept of "choice" is for Aristotle. We cannot hope to live a fulfilled active life if we do not *choose* our actions from a virtuously developed disposition or character. Moral virtue is precisely that. It consists in the perfection of the desiderative part of the soul, disposing one to desire what is the "right" or "appropriate" thing to do. Moral virtue, therefore, is concerned with choice because, as a desire for doing what is the right thing, it is one of its prerequisites. It is the "character" from which we are disposed to desire what is right, and "desiring what is right," in turn, is one of the "ingredients" of the choices that make our life a fulfilled life.

Aristotle states that moral virtue is a *habit* concerned with choice. To explain what he means by this, we can refer back to the discussion of capacities and habitual dispositions. As human beings, we all have the capacity to feel and desire. Moreover, we all have the ability to cultivate this capacity. Dependent on our training and behavior, our feelings and desires can become more subtle and refined. They can become more sophisticated and adapted to the complexities of living with others in a community. According to Aristotle, this process of cultivation is guided by the intellect. In the course of this process, we develop a "character," i.e., a more or less coherent web of habitual dispositions, conditioning our emotional responses to and desires in particular situations. Dependent on our upbringing, the choices we make, the things we do, and the way we reflect upon them, this character may be either more or less vicious or more or less virtuous. According to Aristotle, moral virtue consists in a web of perfectly developed habits coming together in a coherent and stable character disposing someone to desire/choose what is right because it is the right thing to do. Thus, moral virtue is a *habit* concerned with choice.

Until now, we presupposed *that* our emotions and desires can be cultivated by reason, but we did not yet explain *how* this is possible. Yet, understanding this is important, for moral virtue is a habit (concerned with choice) that must be in accordance with reason ("…this being determined by reason and in the way the man of practical wisdom would determine it"). Now, the condition of the possibility of the cultivation of the emotions and desires by reason is that, according to Aristotle, emotions and desires have a "cognitive structure" involving intentional objects and beliefs.

For instance, fear always implies fearing something we believe to be fearsome or happiness is always happiness about something we believe to be pleasurable. In a similar vein, desire is always desire for something we believe to be desirable. Because these cognitive correlates are connected to beliefs of the person experiencing the emotion or desire, and because these beliefs may be more or less rational, our emotions and desires can be more or less rational (e.g., someone may develop an "irrational" fear for open spaces). Developing moral virtue then means, learning to appreciate emotionally the moral salience of concrete situations in an "appropriate" way by developing the "right" beliefs about, for instance, what is fearsome, pleasurable, or desirable. Given, this "rational dimension," moral virtue involves the cultivation of the belief-system governing our emotions and desires to the point that we are disposed to emotionally respond in concrete situations as "reason" or "the man of practical wisdom" would determine it.

We now know that moral virtue entails being able to emotionally respond appropriately in concrete situations. We also know that an appropriate emotional response is one "as reason or practical wisdom would determine it." To explain further what this means, Aristotle calls moral virtue a habit concerned with choice, "lying in a mean relative to us". To understand what he means by this, it is useful to see that our emotional responses and desires allow for *gradation*. For instance, dependent on the particulars of a situation, it may be appropriate to get more or less angry. Now, by the phrase "lying in a mean relative to us", Aristotle provides a formal description of what "appropriate" means. An emotion or desire is appropriate, if it – given the particulars of the situation – is neither an excess (over-reaction, for instance, becoming too angry) nor a defect (under-reaction, remaining too friendly). The morally virtuous person has the ability to hit in a given situation the mean between emotional excess and defect.

To be sure, the mean Aristotle talks about is not an arithmetical mean leading to "average" emotional responses lying between two extremes irrespective of a given situation. If this were the case, for instance, it would be impossible to become very angry if the situation required this. Instead, Aristotle's mean is a proportional mean, a mean relative to the situation we find ourselves in. As such, it is not an average but rather the optimal, best possible, response over the two worst possible responses of under and over-reaction. In this way, it is indeed possible to become very angry if the situation requires this.

The mean of the morally virtuous neither is a "sociological mean." It is not the response of the "average person" in some group. As the best possible emotional response in a given situation, it is the response of the best. By "the best," Aristotle once more does not refer to the best in whatever group. He rather refers to those who developed their capacity for reasoning about practical matters to the extent that they have become practically wise. Thus, Aristotle defines the criterion for hitting the mean in terms of reason and practical wisdom, i.e., as "determined by reason and in the way the man of practical wisdom would determine it".

Until now, we talked about moral virtue as if it is "one thing." In a sense, this is correct. For moral virtue refers to the habitual disposition of the desiderative part of the soul disposing us to desire what is right and appropriate, as the man of practical wisdom would do. However, as human beings, we find ourselves in different and complex situations, forcing us to deal with different and multiple emotions and desires.

Proportional to this multiplicity, Aristotle distinguishes a plurality of moral virtues. Each of them relates to particular dimensions of our emotional, desiderative, or social life and consists in a virtuous optimum (the mean) over two vicious maxima (over and under-reaction). To give some examples; with respect to the pleasure/pain dimension of our existence, Aristotle defines the moral virtue of "temperance" relative to the two vices of "insensibility" and "self-indulgence." Related to the emotion of anger he distinguishes "patience/good temper" as the optimum over "irascibility" and "lack of spirit," or related to the dimension of "honor" he defines "magnanimity" as the optimum over "vanity" and "pusillanimity" (for a complete overview see Wolf 2002, p. 79–80). Thus, moral virtue consists in a plurality of virtues connected to the many dimensions of the complex emotional, desiderative, and social life of human beings.

Now, with respect to these many moral virtues Aristotle argues that it is impossible to subsume them under or translate them into one "overriding" virtue. They truly are a plurality adapted to the complexities of our lives. Yet, in spite of this plurality, there still is a sense in which the moral virtues are one, for they are intrinsically related. This means that, according to Aristotle, it is impossible to be morally virtuous in one dimension, without being morally virtuous in all the other dimensions. Unlike other habitual dispositions that can exist independently of each other (e.g., excellence in sports and excellence in arithmetic), the moral virtues depend on and presuppose each other. To be virtuous in one means to be virtuous in all. Thus, the many virtues are *not* a *mere* plurality – each independent of the other. They rather presuppose each other, coming together in a moral "character" that is both constant and consistent. Persons of moral virtue, in concrete situations, can rely on this character. It encompasses the different moral virtues, enabling them to be sensitive to the contextually relevant emotions and desires, and their proportional mean, allowing them to desire the optimum for each of them, thus taking into account the plurality of dimensions characterizing both our humanity and the problems we as humans have to meet.

Given these considerations, we are in a position to understand Aristotle's definition of moral virtue as "a habit concerned with choice, lying in a mean relative to us, this being determined by reason and in the way the man of practical wisdom would determine it" ($1106^{b}36$–$1107^{a}2$). It is the habitual disposition of the desiderative part of the soul enabling us to have the emotional responses and desires allowing us to appropriately deal with the complexities of our social existence by disposing us to desire what the practically wise would understand as the right thing to do.

10.3.2 Practical Wisdom

Moral virtue is the first "ingredient" of the virtuous choices needed for *eupraxia*. Practical wisdom is the second. According to Aristotle, practical wisdom entails the ability to deliberate and judge well about action, i.e., about what operations contribute to the fulfilled life of the actor in a given situation. To show what this entails we intend to do two things. First, we explain the difference between "making" and "acting." Second, we explain their respective virtues, "skill" and "practical wisdom."

10.3.2.1 Making and Acting

To explain what practical wisdom entails, it is relevant to explain the difference between making and acting. We can see why this is so by going back to Table 10.2. In that table, we can see that the proper objects of the calculative part of the soul are things insofar as they are variable and within our power to change. Within this domain, Aristotle distinguishes two types of operation, each with its own type of deliberation/judgment and virtue. These two types of operation are "*making*" and "*acting,*" the types of deliberation related to these operations are *technical* and *practical deliberation*, and the two related virtues are "*skill*" and "*practical wisdom.*"

To explain what making and acting entail, we discuss their differences. First, we pay attention to their different ends, respectively some product and living a fulfilled life. Second, we discuss the difference between what contributes to the ends of making and acting; respectively means to realize the end and instantiations of the end. Third, we discuss differences between making and acting with regard to judgment about their success, respectively the quality of the product and the disposition and motives of the actor.

"Making" (*poiēsis*) aims at producing something external to the maker (e.g., constructing a house, building a ship, making someone healthy). "Acting" (*praxis*), on the contrary, aims at the perfection of the actor as a human being. Unlike the end of making, the end of action is not the perfection of something external to the actor (some product), but rather the perfection, the *eudaimonia*, the fulfilled life, of the actor her or himself.

This implies that, for making, there is, in principle, nothing to constrain us from making whatever is in our power to make. From the perspective of the maker, it does not matter what the product is. It is a contingent selection limited only by what is technically possible. In principle, each product is as "good" as any other. Acting is quite different. The goal of acting is the fulfilled life of the actor. To this purpose, acts have to be instantiations of what it means to live a fulfilled life. This, in turn, means that not every act that is possible technically is also desirable. Given the particular circumstances, the act must be an instantiation of what it means to live a fulfilled life. This means that the act must fit both the particular circumstances of a given situation and the end of living a fulfilled life here and now. The act still is a contingent selection, yet it is not only constrained by technical possibility, but also by the end: the fulfilled life of the actor.

The distinction between the end of making (some product external to the maker) and the end of acting (the perfection of the humanity of the actor), also introduces a distinction between the operations aiming at these ends.

The "making-operations" leading up to the finished product are not themselves the finished product, i.e., they do not have their end in themselves. For instance, when building a ship, the operations producing the finished ship are not themselves the finished ship. One could also say that, when making something, the end is *external* to the operations leading up to it. This does not hold for acting. Here, the end is *eudaimonia*, the perfection of the actor as a human being. Now, living a fulfilled life is not a "product" that can be separated from the actions constituting it. It is something instantiated in these actions. It is by acting virtuously, that the actor lives a fulfilled life. This means that acts have the end – living a fulfilled life – within them. It is impossible to separate the act and the end. Thus, if making is concerned, the end (the finished product) is something external to the process of making. In the case of acting, the end (living a fulfilled life) is something internal to the act, i.e., the act itself is an instantiation of living a fulfilled life.

These distinctions between making and acting reflect themselves also in the way we judge the skill of makers and the virtue of actors. More in particular, we judge the skill of the maker by the finished product, for this is the end of making. If we want to know how good a maker is, we are not interested in the disposition of the maker or in what the maker was thinking when he or she made the product. We do not ask what his or her motives were, or whether these motives were virtuous or vicious. We rather judge the product. In the case of acting, the motives and the disposition of the actor are crucial for judging the virtue of the actor. As argued above, doing the right thing is not sufficient for an act or an actor to be virtuous. Someone may do the right thing for the wrong reasons. Thus, if we want to judge the virtue of actors, it is insufficient to look at the act in isolation. We also need to take into account the character and the motives of the actor. Thus, for acting, it is impossible to separate the quality of acts from the disposition and motives of actors.

10.3.2.2 Skill and Practical Wisdom

Apparently, making and acting are quite different dimensions of our active life. According to Aristotle, these differences are also reflected in their respective virtues: skill and practical wisdom. Strictly speaking, we only need to explain practical wisdom here, for this is the virtue needed for *eupraxia,* which is the "practical constituent" of *eudaimonia*. Still, we choose to explain both skill and practical wisdom. There are two reasons for this. First, just as making and acting, skill and practical wisdom are contrasting concepts we can use to elucidate their respective meaning. Second, below we argue that to function as an alternative to the disciplines, organizational tasks should enable the development of both skill and practical wisdom. Therefore, to define alternative functional requirements to organizational structures, we need to understand the meaning of both concepts. To explain what skill and practical wisdom entail, we describe them in terms of the knowledge and experience required for them. Moreover, we describe deliberation and judgment involved in them.

Skill: Knowledge and Experience

A skilled maker requires knowledge and experience. Knowledge is required about (1) the product, and (2) about the causes producing this product. The required experience refers to what produces this product in a specific situation.

To start with, a maker needs to have knowledge about the end (the product). No one would call a doctor skilled, if he or she had no idea of the end of the art of medicine "health" or its opposite "sickness." This knowledge of the end can be more or less articulated. For instance, as the art of medicine develops, the sophistication of the conception of "health" increases, allowing doctors to better specify the end of their interventions.

Moreover, specialization seems to be possible. For instance, a doctor may specialize in producing healthy kidneys, and the knowledge required for this. However, note that in spite of specialization, it belongs to the skill of the doctor to see the connection between the health of the kidney and the health of its owner.

Skill-related knowledge of the end abstracts from the question whether this end is good or not. Doctors relying on their skill-related knowledge alone cannot judge whether it is a good thing to produce health or not. For this judgment, they need to see "producing health here and now" as an instantiation of what it means to live a fulfilled life. This insight does not result from skill-related knowledge, but from practical wisdom.

A skilled maker also requires knowledge of "what contributes to the end." To make something, a maker needs knowledge of causes that can be used to realize the ends (the desired effects) of his or her art. Once, again, we would not call a doctor skilled if she has a clear conception of what health is, but does not know what causes produce health. According to Aristotle, this knowledge of causes resembles scientific knowledge.

Just as scientific knowledge is knowledge about causes explaining why phenomena are as they are and cannot be otherwise, the knowledge about causes in the case of making allows for understanding why this effect is produced and not another. Notice that in the case of making, this knowledge of causes is *pharmakon*, i.e., it can be used to either virtuous or vicious purposes. For instance, a skilled doctor may use her knowledge of herbs to produce sickness or even death instead of health.

In sum, as makers become more skilled, they will have better knowledge of the causes that can be used to produce the end of their art, allowing for an understanding of both that and why this end is produced.

Besides knowledge a skilled maker needs experience. Experience is relevant for producing this or that effect in this or that particular situation. For instance, if a doctor heals Peter: this individual, here and now, with this particular condition and history.

To be able to determine what is the case with Peter, what medicine to use, and how much of it – not too much and not too little –, mere scientific insight in the causes of health is insufficient. What is needed in addition is experience.

In the context of making, experience is a kind of aggregated knowledge about what worked and what did not, given relevant particulars of concrete situations. Experience provides the maker with the ability to "judge" what is required in a given situation to realize some effect, without necessarily knowing why this is required, i.e., without knowing the causes producing the effect.

Because experience is acquired by doing and example, it is connected to one's learning history and therefore fundamentally personal (although one may – to some extent – generalize and transfer one's experience to others).

Aristotle values experience highly as a constituent element of skill. Without it, the general "scientific" knowledge about causes is impossible to apply with accuracy and swiftness in concrete instances. Who would choose a doctor who only has a lot of scientific knowledge of the causes of health, but no experience in applying this knowledge in concrete situations? Therefore, as makers become skilled, they acquire experience, enabling them to produce desired effects belonging to their art, taking into account the salient features of the particular situation.

Skill: Deliberation and Judgment

Given knowledge of the end, knowledge of causes, and experience, skilled makers must be able to deliberate well. According to Aristotle, deliberation in general is an inquiry into what contributes to an end, either as a means (making) or as its instantiation (acting). As such, each deliberation presupposes the end as a starting point. It is for this reason that Aristotle says that deliberation is not about the end (Book VI).

Given the end, deliberation proceeds as an inquiry into what contributes to it, and if this is more than one thing, which of them is the most suitable contribution. It stops if one finds a possible and suitable contribution to the end, fitting the particular circumstances.

Applied to making, deliberation requires that given the end, the maker inquires into the possible, and if there are more, the most suitable causes that can be used as means to produce it. In this inquiry, experience with making similar products and knowledge about causes that may be used and the maker's ability for deliberation come together. Drawing from both experience and knowledge, the skilled maker can both swiftly and accurately trace and weigh possible means and find the ones that best fit the production of the end.

Finally, a skilled maker needs judgment. The maker must judge that the operations suggested by deliberation in this particular case constitute suitable means to realize the desired effect. This judgment may be an "implied judgment," i.e., not an explicit judgment as "I affirm that Socrates is a human being," but rather a "seeing as." The skilled maker is able to "see" these particular operations "as" possible and suitable means. For instance, a skilled gardener is able to see this amount of nutrient dissolved in this amount of water, and applied in this way as both a possible and suitable means to restore the health of this rose, planted in this kind of soil, receiving this amount of sunlight. To this purpose, the gardener needs a continuously developing sensitivity for the "condition" of a garden and experience with restoring the health of roses in different gardens and under different circumstances.

Practical Wisdom

The practically wise are able to deliberate and judge well about concrete actions instantiating the fulfilled life. Just as in the case of skill, this requires knowledge and experience as prerequisites, and deliberation and judgment as constituent elements. However, the knowledge, experience, deliberation, and judgment involved are quite different.

Practical Wisdom: Knowledge and Experience

In the case of acting, the end is the fulfilled life of the actor. Unlike the end of making, this is not just any end or product. It rather is the end at which we as human beings are directed, the fulfillment of what we, given the best of our natural capacities, can become.

Moreover, unlike the end of making, the end of acting can be specified only in outline. It is striking that Aristotle never gives a detailed description of what living a fulfilled life entails. He rather argues the other way around. As cultivated morally virtuous adults with more or less sophisticated experience, and a more or less developed ability for reasoning about practical matters, we can both point at acts that are examples of what it means to live a fulfilled life and motivate why we think that these acts are exemplary.

So, without being able to specify in detail what living a fulfilled life entails, we can "recognize" good acts when we see them. According to Aristotle, persons lacking this capacity for "recognition" will never understand his ethics (or any ethics for that matter) and should not even begin studying it. Thus, unlike the end of making that, in principle, can be specified in detail, the end of acting can only be specified in outline, and it depends on the virtue of the actor to find in each concrete situation an action instantiating this goal. Consequently, to be practically wise, the actor must know in outline what it means to live a fulfilled life. To this purpose, the actor must be morally virtuous, i.e., must be disposed to desire what is right based on training and education, and the actor must have a broad experience with virtuous and vicious action. With each virtuous act, both the actor's moral virtue and knowledge of the end will become richer, adding to his or her practical wisdom. However, such a "rich" conception of the end of acting still is different from the "detailed" specification of the end of making.

Knowledge of the end is only a necessary condition for practical wisdom. For instance, someone may know that it is virtuous to act courageously, but this does not yet mean that he or she knows what a courageous act in a given situation is. This is what practical wisdom is about: being able to find what is needed to live a fulfilled life in the situation one finds oneself in.

According to Aristotle, the knowledge required for this is different from the scientific cause-effect knowledge implied in skills. We already explained that in the case of making, knowledge is required about causes that can be used as means to produce the end. In the case of acting, this semantics of means and ends does not apply. Acts are not means to the end of living a fulfilled life. For instance, if someone acts courageously, the courageous act is not a "means" to bring about the fulfilled life as a "product" that can be separated from the act. As argued above, the act *itself* is an instantiation of what it means to live a fulfilled life. It is an instantiation *of* this end. Thus, action is not a means to the goal of living a fulfilled life, as if living a fulfilled life is a kind of "product" that can be separated from the acts constituting it. It is rather a – more or less, excellent – instantiation *of* the goal of living a fulfilled life. Consequently, the knowledge required for finding acts that are here and now realizations of living a fulfilled life is not scientific knowledge about cause-effect complexes that can be used as means to some end as is the case with the skills.

The kind of knowledge that is required is knowledge implied in moral virtue and experience by acting in a community. Moral virtue means: being disposed to desire what is virtuous – the mean between two vices – in a given situation. This disposition entails a cultivated sensitivity to respond adequately to the salient particulars of a given situation. We build up this sensitivity by training, experience, and reflection on both our actions and those of others. Because of the importance of experience in practical wisdom, Aristotle argues that practical wisdom comes with age, or more precise, that it comes by living an active life, involved in a community, allowing the development of a rich base of experiences with virtuous and vicious acts and the circumstances under which these acts were performed. In this respect, practical wisdom is quite unlike sciences such as mathematics. Such sciences primarily rely on knowledge of principles, requiring only little experience. For this reason, the young can easily learn them. Therefore, to be practically wise, i.e., to be able to deliberate and judge well about what contributes to living a fulfilled life, one needs experience evolving by action and reflection upon it.

Practical Wisdom: Deliberation and Judgment

The deliberation involved in practical wisdom also differs from that involved in the skills. Acting virtuously means acting appropriately in a given situation because the act is seen as the appropriate thing to do. This kind of deliberation involves finding a particular act fitting the requirements of both the end (live a fulfilled life) and the situation in which the actor finds him or herself. Thus, in contrast to making, the problem of deliberation about action is not finding causes functioning as means to some end, but about finding the most fitting instantiation of living a fulfilled life here and now, taking into account the particular circumstances of the situation. To this purpose, the actor needs to be:

1. sensitive to the context of the action and take into account the features relevant for it;
2. sensitive to values that may be at stake in this context and to weigh them;.
3. able to think of possible courses of action and weigh them in the light of both the specifics of the situation and the values at stake (given the requirements of the situation).

Furthermore, the actor needs to:

4. learn from past actions to increase the quality of the body of experience needed for further action.

Thus, instead of ordering causes as more or less efficient and effective means to ends in technical deliberation, practical deliberation "ranks" acts as more or less good instantiations of living a fulfilled life. To this purpose, the conception of the end must be right, the reasoning must be correct, and the deliberation must neither proceed to slow nor to quickly given the requirements of the circumstances (1142^b11–1142^b33).

Deliberation is an inquiry to find an act that may be a fitting instantiation of living a fulfilled life in a given situation. As such, deliberation is insufficient for practical wisdom. What is required in addition is "judgment." Someone who is practically wise must be able to judge that an action found by inquiry is or is not a suitable instantiation of what it means to live a fulfilled life in the given situation. As in the case of the skills, this judgment may be an "implied judgment" that can be compared to "seeing as." To explain this kind of judgment, think of a chess player who is able to "read" a position on a chessboard and to "see" whether a particular move strengthens or weakens it. In a similar vein, someone who is practically wise is able to "read" a situation aright and to "see" whether a particular act is a fitting instantiation of living a fulfilled life. Aristotle compares the "seeing as" related to practical wisdom to seeing three lines (e.g., drawn in the sand) as triangle. By this comparison, he touches on several dimensions of judging that this act is an appropriate instantiation of living a fulfilled life.

First, the seeing involved is not a mere perception of qualities (e.g., shades, colors, etc.). It is not "seeing" but rather "seeing as"; not moving a chess piece, but seeing the move as an improvement of one's position.

Second, this "seeing as" involves taking a manifold, the three lines, as a whole, the triangle. As such, the "seeing as" takes its point of departure in perception (three concrete lines in the sand). However, it ends in an act of synthesis by the rational part of the soul (the judgment: "these lines constitute a triangle"). Something comparable occurs when we judge that in this particular situation, this particular act is an instantiation of living a fulfilled life. In this case, we experience a manifold, the situation and the act, with their many features, as fitting together in a meaningful whole, i.e., as an instance of living a fulfilled life. Therefore, someone who is practically wise can judge that a particular situation and a particular act fit together in a meaningful whole, i.e., into an instance of what it means to live a fulfilled life (see also Wolf 2002, p. 153).

Third, the "seeing as" also involves taking something "particular" – these three lines in the sand – as something "general," a triangle. Something similar occurs when we judge that this particular act is an appropriate instantiation of living a fulfilled life. All action is about particulars. Being sensitive to what is salient, and judging that doing this, here, now is an instance of what it means to live a fulfilled life, involves judging that something particular (this act) is an instance of something "general" (our conception of the fulfilled life). However, note that the "general" in question, is not something fixed. It develops as we gather experience, deliberate about, and reflect upon virtuous action. By deliberating about action and judging that this act "counts as" an instance of living a fulfilled life, we both find what contributes *to* the end as an instantiation of it and, in this way, further enrich our conception *of* the end.

Fourth, the "seeing as" is a kind of "grasping" something ultimate. For instance, suppose someone does not see the lines in the sand as a triangle. Of course, we may explain what a triangle is and say, "Don't you see it now?" We may draw other triangles and point at the similarities with the lines in the sand. In the end, however, the person in question either "grasps" the lines as a triangle or not. In this sense, the "seeing as" of practical wisdom can be construed as a counterpart of what Aristotle calls "understanding." In the case of understanding, we grasp the principles from which scientific demonstration proceeds. These principles are ultimate in the sense that they are the basis of all demonstration. As such, it is impossible to prove them by demonstration. In the case of judging that some act is an appropriate instantiation of living a fulfilled life, we also are dealing with an ultimate. However, this ultimate is not a general principle of demonstration, but a particular, this action, seen as a proper instantiation of what it means to live a fulfilled life.

Now, the "seeing as" involved in practical wisdom, must not be taken as an inborn ability for ethical intuition. According to Aristotle, this ability develops as we acquire experience. Just as the skilled carpenter learns to see what is good or wrong with a particular construction, the practically wise have learned to judge particular situations aright. This process of learning involves (1) a flexible and continuously developing concept of what it means to live a fulfilled life, (2) experiences with virtuous or vicious action (both of oneself and of others), and (3) reflection on the action and its motives feeding back into the concept of what it means to live a fulfilled life. Therefore, by means of acting virtuously and reflection on our acts, we develop our ability for judging new situations in terms of our developing concept of what it means to live a fulfilled life.

Comparing Skill and Practical Wisdom

In practical wisdom, good practical deliberation and good judgment come together. They dispose the practically wise to (1) search for acts fitting the particulars of the situation and instantiating what it means to live a fulfilled life and (2) judge that this act indeed should be done because it is an instantiation of what it means to live a fulfilled life. Table 10.4 summarizes the main findings regarding the requirements for and the constitutive elements of respectively skill and practical wisdom.

Reflecting on these differences between skill and practical wisdom it is perhaps possible to describe them as different modes of reasoning, highlighting two dimensions of purposeful human behavior.

It is possible to describe skill as an *isolating mode* of reasoning about purposeful behavior. We focus on this against that product, taking them as "contrary" opposites, standing indifferently and contiguously in a manifold of products. We do not consider the operations needed to produce them as more or less perfect expressions of what it means to live a fulfilled life. We neither consider these operations as contributing to one life with others in a community. When reasoning about making, we isolate the product and the operations to produce them from the "condition," the life, and the community of the maker. We isolate causes producing isolated desired effects and, in turn, isolate these cause-effect complexes from possibly disturbing circumstances. In short, by reasoning about purposeful behavior in terms of making, we concentrate on finding and ordering causes to realize whatever contingent desired effect considered in isolation from what it means to live a fulfilled life.

Relative to this isolating mode of reasoning, Aristotle seems to propose an *inclusive mode* of reasoning about purposeful behavior. In this mode, we speak of acts as more or less substantially virtuous, thus relating them to what it means to live a fulfilled life as end in itself. By speaking in terms of acts, we include the humanity, the emotions, desires, and rationality of the actor. We also include his or her "condition," moral development, and social life, and we take into account the circumstances of the act – not as possible disturbances – but as salient features

Table 10.4 Skill and practical wisdom: requirements and constitutive elements

	Skill	Practical wisdom
Knowledge of end	The end is the desired effect of the skill in question. There are as many ends as there are skills. The end can be articulated in detail in advance. It can be decomposed into sub-goals (specialization).	The end is living a fulfilled life. This is the one and ultimate end of both moral virtue and practical wisdom. Developing both moral virtue and practical wisdom, implies developing experience-based, tacit knowledge about the end. By means of reflection, one may articulate knowledge of the end.
Knowledge of what contributes to the end	Resembles scientific knowledge, involving knowledge of causes that may realize the end the skill	There is no knowledge resembling scientific knowledge about cause-effect complexes that may serve to find means (the semantics of goals and means does not apply to the case of practical wisdom).
Experience	Experience is needed to both find means and to apply knowledge to realize in each particular situation the desired effects related to the skill	Experience is needed to find in concrete situations actions instantiating what it means to live a fulfilled life (the semantics of goals and means does not apply in the case of practical wisdom).
Deliberation	The calculative process of finding in a particular situation the possible and suitable means to realize a desired effect (product) belonging to the skill (deliberation is based on experience and knowledge of the relevant cause-effect complexes).	The deliberative process of searching, in a particular situation, a possible and suitable action instantiating what it means to live a fulfilled life (deliberation is based on a morally virtuous character and experience).
Judgment	Seeing particular causes as possible and suitable means to realize in a concrete situation the desired effect (product) belonging to the skill (judgment is based on experience).	Seeing this action as a possible and suitable instantiation of living a fulfilled life in this particular situation (judgment is based on moral virtue and experience).

relevant to its virtue. In short, by reasoning about purposeful behavior in terms of acts we concentrate on finding and ordering actions pursued for their own sake because they contribute to living a fulfilled life. Thus, we concentrate on what makes purposeful behavior of value to our lives with others in our community.

Although the dimensions are logically distinct, they are dimensions of the same operation. One can view an operation as a cause to a desired effect in the isolating dimension and as an instantiation of the unqualified end of living a fulfilled life in the inclusive dimension. For instance, reasoning in the inclusive dimensions, a doctor judges that treating this patient here and now in this particular way is the virtuous thing to do and for this reason chooses to act accordingly. This makes the act virtuous in the inclusive dimension. However, the judgment on the treatment of this patient may also involve reasoning in the isolating dimension, arranging the most suitable causes as means to treat this patient given his or her condition. In this case, by producing something external to the maker (the health of the patient), the maker/actor realizes his or her humanity (living a fulfilled life).

The inclusive mode of reasoning about purposeful behavior and its related virtue practical wisdom are intrinsically related to the isolating mode and its virtue: skill. Making always involves an end that is considered as a value. This means that making gets its real significance in the context of acting. Now, reasoning in the inclusive mode and its related virtue practical wisdom are about deliberating and judging about values in the context of action. So, if making, reasoning in the isolating mode, and skill are to contribute to living a fulfilled life, they presuppose acting, inclusive reasoning, and practical wisdom. By means of skill alone, we can never fathom the real significance of our behavior for our development as a human being. To that purpose, we need practical wisdom in everything we do and make. Reasoning in the isolating and inclusive mode are two distinguishable dimensions of human purposeful behavior that are related in the sense that, if our purposeful behavior is to contribute to living a fulfilled life, making follows acting, reasoning in the isolated mode follows reasoning in the inclusive mode, and skill follows practical wisdom. For this reason practical wisdom and not skill is the excellence of the deliberative part of the intellectual part of the intellect.

10.3.2.3 Choice: Combining Moral Virtue and Practical Wisdom

Until now, we discussed moral virtue and practical wisdom in isolation. We did not yet pay attention to their combination in choice. Still, this combination is important, for as we argued, according to Aristotle, *eupraxia* as part of living a fulfilled active life requires that acts are chosen, chosen for their own sake, and chosen from a morally virtuous character.

According to Aristotle, choice is a (voluntary) deliberate desire for particular ends that lie within our power. As *deliberate* desire, choice involves deliberation and judgment, which are operations of the rational – and more specific, calculative – part of the soul. As deliberate *desire*, choice involves the operation of the soul's desiderative part.

To explain how the desiderative and the calculative part of the soul contribute to choice, we begin with the role of the desiderative part. According to Aristotle, the desiderative part of the soul is the principle of the emotions, appetites, and desires. As such, it is what moves one from within to act, "The mover, then, is one, the desiring part" (433^{b}13–28). Desire is the capacity inducing movement/action from within the agent towards the end. According to Aristotle, this end is always an apparent good. Dependent on the virtuous or vicious habituation of the desiderative part of the soul, this apparent good is either an end in itself or a relative good (a good in relation to, for instance, pleasure or wealth). The virtuously or viciously habituated desire for an apparent good is the first constitutive element of choice.

However, the desire for an apparent good is not sufficient for action. For instance, Peter may desire to act courageously. Yet, this desire alone does not tell Peter how to act in a particular situation (e.g., a situation in which an old person is harassed by hooligans). There still is a gap between the desire for an apparent good and the desire to act in this or that way in a particular situation.

The operation of the calculative part of the soul in the form of deliberation and judgment is required to "fill this gap" between the desire to "do the right thing" and choosing to act in this or that way. Given the end provided by desire (in the example, the desire to act courageously), deliberation is the process of finding out what is an appropriate instantiation of this end in the situation one finds oneself in. By means of deliberation and judgment, Peter may find a particular act that is in his power and can be affirmed as fitting both the desire to be courageous and the salient features of the situation he has to deal with.

Still, intellectually affirming this particular act as an appropriate instantiation of what it means to act courageously is not yet sufficient. To become a choice, the result of the deliberation must be taken up by desire. Given the initial desire for an apparent good, desire must take up what is intellectually affirmed as an appropriate "here and now" instantiation of this good, and make it into its object. Once this happens, a (voluntary) deliberative desire – a choice – comes into existence for a particular end that is an appropriate and feasible instantiation of the initially desired apparent good.[3]

Given this structure of choice, we can go through the different elements once more and explain what is needed for the *virtuous* choices needed for eupraxia constituting the practical dimension of living a fulfilled life.

In the substantially virtuous choice, desire is the principle of action. According to Aristotle, desire by nature is directed at the apparent good. Of course, we may be mistaken about this apparent good. What appears as good may be so only in a qualified sense (such as wealth or pleasure). If this is the case, we may desire the "wrong" end as an apparent good. To be able to desire as good what actually is good, desire requires the guidance of the intellectual (calculative) part of the soul aiming at truth. By being habituated in accordance with practical wisdom, the morally virtuous person has developed the habitual disposition to desire as valuable what is actually of value for living a fulfilled life.

In the substantially virtuous choice, the calculative part of the rational part of the soul is the principle of truth. The practically wise person is disposed to deliberate and judge well about what contributes to the desired end and thus hit the truth about what pertains to the end of living a fulfilled life in a given situation. Given the desire for the right end, practical wisdom is the virtuous disposition allowing for finding

[3] Some authors arrange the constitutive elements of choice in the form of a 'syllogism' (e.g. Kenny, 1979 who is also critical of this arrangement; Sherman, 1989; Hughes, 2001). Kenny (1979) calls syllogisms pertaining to choices of ends sought for their own sake 'ethical syllogisms' (as opposed to 'technical syllogisms'). The desire to do the apparently good thing is the 'major' of this syllogism, which is also called the 'premise of the good'. The minor of this syllogism is the particular act that instantiates doing what is apparently good in the given circumstances. As this act must be in our power to realize, the minor is also called the 'premise of the possible'. This premise is the result of deliberation and judgment. The 'conclusion', then, is the desire to act in this particular way in the given circumstances.

an act that, given the circumstances, is a fitting instantiation of what it means to live a fulfilled life.

However, knowing the virtuous thing to do in a given situation is insufficient. What is known as virtuous must be realized in action. To become an (intention to) act, what deliberation and judgment affirmed as the virtuous thing to do, must be taken up by morally virtuous desire. Once this happens, we have a virtuous deliberate desire for a concrete action that is seen as the right thing and desired because it is seen as the right thing.

Thus, in a substantially virtuous choice, moral virtue and practical wisdom come together. We need a well-disposed desire in the form of moral virtue to provide the right disposition towards – a still generally conceived – end of living a fulfilled life. Given this disposition, we need a well-disposed intellect in the form of practical wisdom to find the truth about what *here and now* contributes to the desired end of living a fulfilled life. Finally, we again need well-disposed desire in the form of moral virtue to take up the result of deliberation and make it into an intention to act in this or that way. Together, well-disposed desire and a well-disposed calculative intellect provide the necessary and sufficient conditions for the virtuous choices needed for *eupraxia*.

10.3.2.4 Developing Moral Virtue and Practical Wisdom

Virtuous choices require moral virtue and practical wisdom. However, these virtues are not natural capacities like, for instance, our ability to reason. Based on natural capacities, they may (or may not) develop in the course of our life. The question, therefore, becomes *how* moral virtue and practical wisdom develop. This question becomes more pressing once we realize ourselves that both moral virtue and practical wisdom are habits, cultivated by action. More in particular, by bad action, we cultivate bad habits and by virtuous action, we cultivate virtuous habits.

This relation between habits and action seems to imply a circle. Virtuous actions require virtuous habits and virtuous habits require virtuous actions. To this circle, a second is added, for, on the one hand, Aristotle tells us that moral virtue requires practical wisdom. Yet, elsewhere in his ethics, he argues that practical wisdom requires moral virtue. To be morally virtuous, we need practical wisdom and to be practically wise, we need moral virtue (see Fig. 10.1).

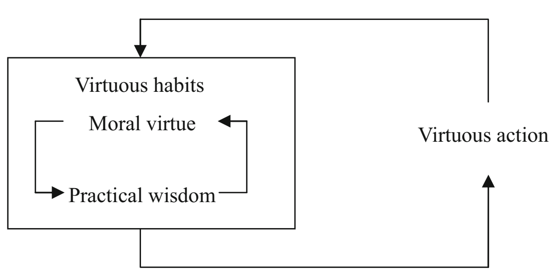


Fig. 10.1 Two circles implied in virtue and its development

It is central to Aristotle's ideas on the development of moral virtue and practical wisdom that he does not try to escape these circles. He rather embraces them to show how virtuous action, moral virtue, and practical wisdom develop together in the course of our life "with and for others and in, just institutions" (Ricoeur 1994 p. 172). This development starts with acts one could call "*accidentally virtuous*," i.e., acts that only appear virtuous to an external observer, yet lack everything making them truly virtuous acts. It ends with acts one could call *"substantially virtuous*," i.e., acts that have everything making them into virtuous acts. Without going into the details of this developmental process and its related problems, it is possible to highlight some of its characteristics (see for an elaborate analysis Burnyeat 1999; Sherman 1989; Vasiliou 1996). To discuss these characteristics, we highlight (1) the goal of and (2) requirements for the developmental process, and by relating these two, we discuss (3) its dynamics.

The ultimate goal of the process of moral development is to acquire the ability to act substantially virtuous, i.e., to choose with the required swiftness, accuracy, and pleasure what is right in a wide variety of situations, to choose it for its own sake, and from a firm and stable character. This entails that *desire* is habituated to see the moral relevance of situations, to be sensitive to their morally salient features and relevant value dimensions, and to respond correctly to these features by being directed at the virtuous middle of the relevant value dimensions. Moreover, it entails that the *calculative part* of the intellect is habituated to deliberate and judge well in a wide variety of situations about what contributes to living a fulfilled life. As argued, this requires rich, experience-based, cultivated, tacit and reflective knowledge of what "fulfillment" entails. Moreover, it requires vast experience with different situations and actions contributing to fulfillment, enabling the virtuous person to deliberate about and see both that and why actions contribute to fulfillment.

According to Aristotle, developing moral virtue and practical wisdom requires processes of habituation, i.e., of developing the capacities involved by practicing/ exercising them. To outline the dynamics of habituation, we add Fig. 10.2.

In the figure, we listed feedback relations (arrows) and natural capacities (in bold) involved in the process of moral habituation (see left side of Fig. 10.2). Moreover, we listed habituation's most important learning devices (italics, in the middle of Fig. 10.2). At the right side of the figure, we indicated the first and last stage of habituation: respectively accidentally and substantially virtuous acts.

Zooming in on the capacities and feedback relations, we see that desire and the calculative part of the intellect are prerequisites for "acting." Together, they are involved in "choices" we make and the acts we perform. By "acting" in a particular situation, we gain "experiences," shaping our emotional responses to situations. By reflecting on these experiences, we develop our ability for deliberation and judgment. Note that "experience based on action" is not the same as a mindless aggregate of sense data. It involves emotional and intelligent processing of actions that, in turn, allows for increasingly sophisticated emotions, concepts, and ways of seeing things, ultimately resulting in increasingly sophisticated habitual dispositions, allowing for "better" responses in more complex situations.

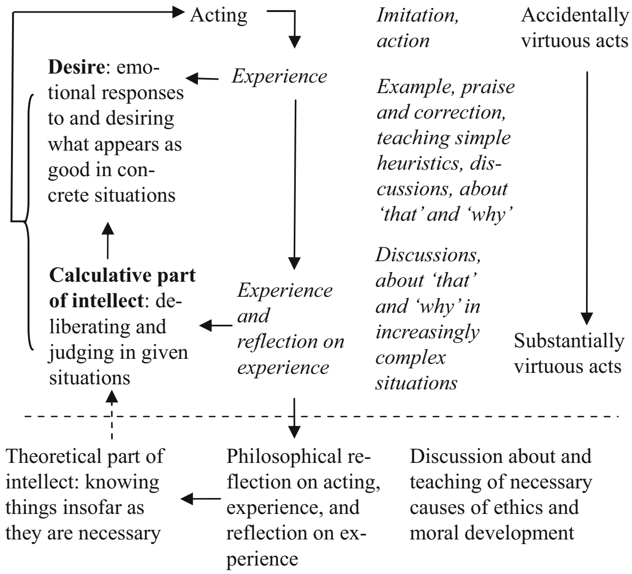


Fig. 10.2 The dynamics of habituation

The figure shows that both desire and the calculative part of the intellect develop on the basis of action and experience, i.e., given the right activity both our desires and emotions as well as our capacity for deliberation and judgment become increasingly virtuous. In addition, there are relations between desire and the calculative part of the intellect. The calculative part of the intellect can reflect on complexes of emotions, desires, and activities in particular situations. By means of this reflection, it in the first place enriches its concepts of both the fulfilled life and what contributes to it. In the second place, it cultivates beliefs shaping our emotional responses to concrete situations.

In the figure, we added the theoretical part of the intellect. Given our ethical development, this part can reflect on its unchanging and necessary causes. This part of the intellect is not directly needed for moral habituation as such. It is rather presupposed in theoretically understanding its structure and necessary preconditions. Because the major part of Aristotle's ethics is a theoretical reflection on human activity, it can be viewed as resulting from the activity of the theoretical part of the intellect.

These capacities and feedback relations allow us to focus on the process of habituation in terms of its learning devices and stages.

From early on, infants imitate behavior of, for instance, their parents, and – if the example is right – simply *act* in the right way, getting a taste for imitation and "games" of action and response. This imitative behavior might be construed as "doing the virtuous thing." However, the infant does not yet know *that* it acts virtuously. Moreover, it does not yet choose the act *because* it is virtuous. In this sense, the act is only "accidentally" and not "substantially" virtuous, for substantially virtuous acts are chosen, chosen for their own sake, and from a virtuous character. Still, the child's pleasure in imitation and games of action and response is important because they are connected with "just and noble things". In this way "evaluative responses may develop in connection with the right objects", even before the infant is able to reason or deliberate about more complex matters (Burnyeat 1999, p. 217). Although "imitation and games of action and response" may seem to be "operand conditioning" and at later stages "indoctrination," commentators such as Nussbaum (1997), Sherman (1989, 1999), Hughes (2001), and Burnyeat (1999) press the point that at every stage of our moral development intentionality, volition, and from early on, deliberation, judgment, and reflection are involved. Only in this way, the whole process is directed at developing a child's *own* discerning powers, i.e., at developing its ability to see as valuable what is of value and to feel, deliberate, judge, and act accordingly because it is seen in this way.

As they develop, the world of children "broadens" and "deepens." They gain morally relevant experience in different contexts. They experience the actions of others in their environment, possibly treating them as "examples" of "good" or "vicious" behavior. They acquire increasingly complex heuristics and roles that need to be applied in different situations. They receive praise and punishment, and they discuss about and motivate their behavior in terms of goals. At this stage of their moral development, children not simply act in accordance with virtue without being aware of this. They rather acquire the disposition to "see" *that* situations are morally relevant and *that* they are doing the right or the wrong thing. Moreover, they may learn to discuss and deliberate about their actions and to choose the right action for its own sake. Thus, as their experience, their emotional sensitivity, and their deliberative powers grow, they become disposed to desire what is right, to deliberate about what contributes to what is right, and to do this swiftly and accurately in increasingly complex situations. Aristotle calls this stage of development, knowing "that." This means that we can: (1) see that a particular situation is morally relevant, (2) see its morally relevant features, (3) see relevant value-dimensions involved in it, (4) judge that a particular action indeed contributes to acting appropriately in the given situation, and (5) act accordingly because it is the good thing to do. (cf. Book II)

Children and young adults continue to learn to articulate and discuss their emotions and motives for action in moral terms, to evaluate reasons for praise and punishment in particular situations, and to discuss the appropriateness of rules, i.e., they continue to be persuaded or persuade others to see particular situations in a particular light. In this way, they further develop the potential to deliberate and judge in increasingly complex ethical situations as well as to provide increasingly accurate and sophisticated *motivations* for the choices they make. Due to growing experience, discussion, and reflection both the concept of what it means to live a fulfilled life and the power to judge actions in terms of their contribution to it grow, thereby adding to the person's practical wisdom. Building upon the knowledge of the “that” – i.e., knowing that this action truly is in accordance with virtue – and supported by education, discussion, and reflection, the ability to understand “why” choices are in accordance with virtue further develops, contributing to the ability to motivate one’s choices. Aristotle emphasizes that the ability to motivate “why” particular actions in particular situations are virtuous depends on being disposed to see “that” situation call for particular actions and “that” particular actions are the virtuous thing to in particular situations. Insight in “why” an action is virtuous may further articulate and explain the point “that” the action is virtuous. In this sense, it does neither add nor rely on extra-moral content. It articulates, refines, and possibly adjusts one’s moral convictions, re-weaving them into the web of beliefs and concepts, enhancing one’s ability to judge that a particular action is virtuous and should be done because of this.

This ability to motivate why actions contribute to living a fulfilled life must not be mistaken for theoretical knowledge about ethical behavior. Motivating why we think an action is virtuous is always situation-dependent, referring to this particular *contingent* situation. As such, it requires experience, moral virtue, deliberation, and judgment. Theoretical knowledge about the “why” of ethical behavior – on the other hand – articulates its invariant, *necessary* structures and causes.

In both cases, knowledge of the “that,” i.e., knowing that particular acts contribute to fulfillment, is required. However, in the former case, this knowledge contributes to enriching our ability to deliberate and judge in contingent situations, allowing us to motivate why we think we should have acted in a particular way in this situation. In the latter case, knowing “that” functions as the point of departure of a theoretical/philosophical reflection on the necessary causes for and structures implied in ethical behavior. This distinction between being able to motivate why we act in a particular way in practical and contingent situations and to understand and explain why ethical behavior is structured as it is has the quite comforting consequence that one does not necessarily need to be a philosopher to be a virtuous person.

Given this explanation, we can understand that virtuous habits develop from virtuous acts and virtuous acts presuppose virtuous habits. We can also understand that moral virtue does indeed presuppose practical wisdom and practical wisdom indeed develops from moral virtue. The point is that habit and action as well as moral virtue and practical wisdom mutually condition each other involved as they are in the same spiraling process of developing, widening, deepening, in short, cultivating the humanity of actors and their actions. By each virtuous choice we make, we become better people because each of them cultivates our emotions and desires and provides our intellect with experience about what a fulfilled life should look like and about what contributes to it. As argued, this process starts with an actor performing acts that may be called “accidentally virtuous.” Dependent on the talent, upbringing, education, and continuous effort of the actor, it may culminate in the actor performing substantially virtuous acts, i.e., acts chosen for their own sake, and proceeding from a firm and unchangeable character ($1105^{a}30$–$1105^{b}1$).

10.4 Aristotle Versus the Disciplines

Broadly speaking, rich survival for organizations means: making a contribution to society in order to enable its citizens to live a fulfilled life. In this chapter, we introduced Aristotle's ethical theory to flesh out what "living a fulfilled life" means. Given Aristotle's description, we are in a position to assess whether it provides a counter-model to the disquieting features attached to the disciplines.

Against the minimalism of goals, Aristotle, right from the start, is interested in the highest good for man. Instead of pursuing the minimal goal of producing trivialized – dehumanized – minds and bodies, Aristotle focuses on the cultivation of man's humanity.

Instead of viewing human beings as objects of discipline, he provides an analysis of the functional dimensions of our existence and argues that a truly fulfilled life entails the development of the highest and most characteristic of these functions into virtuous habitual dispositions. In this way, Aristotle tries to do justice to the many dimensions of our life with others in a community, proposing an optimal instead of a minimal image of humanity and human development.

As has been argued, this development is not a matter of trivializing basically non-trivial behavior by processes of operand conditioning. It rather is a complex and subtle process of moral education that both cultivates and involves our potential to "see" what is right and to reflect on what we "see."

In the course of this process, our desire is attuned to its natural object, i.e., to that which is perceived as "good." For Aristotle, moral development entails learning to judge what is good in a given situation and to desire what is judged as good. In the end, this means that not fear, but what is perceived as contributing to the fulfilled life becomes an important motivational factor for action. Moral involvement and responsibility not only are a duty to the morally virtuous person, they become a pleasure.

By relating making and acting, Aristotle holds a position that is very different from the one implied by the disciplines. The disciplines consider the goals of the production processes as contingent. Whatever is in our power to make can be their goal.

In the disciplines, the goals of the production processes appear as the goals of making considered in terms of the isolating mode of reasoning. In this mode, cause-effect complexes are judged as more or less effective and efficient means to produce desired effects. Neither the means nor the desired effects are judged in the light of what it means to live a fulfilled life. For Aristotle, such an isolating mode of reasoning can never stand on its own. It always is related and subservient to the inclusive mode of reasoning. In this mode, actions and their goals are viewed from the perspective of their contribution to living a fulfilled life in a community. This determines their value. This means that, from an Aristotelian perspective, actions and their goals are never contingent in the sense that we are entitled to make everything that can be made.

Finally, Aristotle's philosophical reflection on the structure of the fulfilled life, on its meaning and prerequisites, can be used to invent social "mechanisms" that support human beings in their moral development. If we assume that "living a fulfilled life" is literally what makes us into human beings, the invention of such "mechanisms" appears as a highly relevant activity. Therefore, contrary to finding tools to further trivialize human behavior, we are urged to invent social mechanisms supporting human beings to develop their humanity.

Viewed in this way, Aristotle's ethics actually provides a counter-model to the disciplines that ticks all the boxes. Goals do make a difference; the development of feelings of involvement and responsibility are encouraged; not fear, but what is perceived as good is a prime motivator for action, cultivating instead of trivializing behavior is a main aim; categories to monitor behavior are developed in a subtle process of moral education, not by operand conditioning; and we are urged to invent social mechanisms supporting the cultivation of our humanity.

Based on Aristotle's ethics we can understand why the disciplines are so worrying. It makes sense of the feelings we have once we become their subject. The disciplines negate what we as human beings, given our organization, are set out to do. They negate our humanity and its fulfillment.

Given this understanding, we need to ask the next question: is it possible to find organizational structures that can support organizations and organizational members to contribute to society in a way that allows its citizens to live a fulfilled life? This is the topic of the next chapter.

References

Barnes, J. (ed). (1991). *The complete works of Aristotle: The revised Oxford translation* (The Bollinger Series LXXI 2nd ed., Vol. 1 and 2). Princeton, NJ: Princeton University.

Burnyeat, M. F. (1999). Aristotle on learning to be good. In N. Sherman (Ed.), *Aristotle's "Ethics": Critical essays* (pp. 205–230). Lanham: Rowan & Littlefield.

Hughes, G. (2001). *Aristotle on ethics*. London: Routledge.

Kenny, A. (1979). *Aristotle's theory of the will*. New Haven: Yale University Press.

Nussbaum, M. C. (1997). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*. Cambridge, U.K.: Cambridge University Press.

Ricoeur, P. (1994). *Oneself as another*. K. Bamey (Trans.). Chicago: University of Chicago Press.

Sherman, N. (1989). *The fabric of character: Aristotle's theory of virtue*. New York: Oxford University Press.

Sherman, N. (1999). The Habituation of Character. In N. Sherman (Ed.), *Aristotle's "Ethics": Critical essays* (pp. 205–230). Lanham: Rowan & Littlefield.

Vasiliou, I. (1996). The role of good upbringing in Aristotle's Ethics. *Philosophy and Phenomenological Research, 56*(4), 771–797.

Wolf, U. (2002). *Aristoteles' "Nikomachische Ethik", Werkinterpretationen*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Chapter 11
Organizational Structures Supporting Rich Survival

11.1 Introduction

Aristotle's ethics provides us with a description of what it means to live a fulfilled life. However, this description is not an end in itself. What we are really after is formulating a set of principles allowing for the design of organizational structures supporting "rich survival." In order to find both these principles and the structures that result from their application, we need to take two additional steps.

Rich survival, as described in Chap. 1, is about organizations contributing to the creation of societal conditions, enabling human beings to live a fulfilled life. Until now, we only discussed the "fulfilled life"-part of this description. To explain how organizations can provide a rich contribution to society, we first need to discuss the relation between organizations and society. This is the topic of Sect. 11.2.

Once we have a model of how organizations are related to society, we can go into the question what this relation should look like, i.e., which requirements should be met, if organizations are to make a rich contribution to society. Based on these requirements, we define principles for the design of their structure. Moreover, we can, in terms of de Sitter's parameters, discuss organizational structures fitting these principles. This is the topic of Sect. 11.3. In Sect. 11.4, we summarize and discuss our findings.

11.2 Incorporating the Organization into Society

To provide an explanation of the relation between organization and society that can function as a basis for deriving requirements and design principles for rich survival, we take our cue from Aristotle's political theory.

According to Aristotle, ethics and politics have the same end: the fulfilled life of the individual. In his ethics, Aristotle reflects on the principles of living a fulfilled life at the level of the individual. What does it mean for a human being to live a fulfilled life and what are its constituent elements and requirements?

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_11, © Springer-Verlag Berlin Heidelberg 2010

Right from the start, Aristotle stresses that individual virtue necessarily has a social dimension. Human beings are naturally disposed to live together in a community and their individual virtue necessarily has a social dimension as it both depends on and addresses problems of living together in that community. A (just) society is both a condition for and a result of the virtue of the individuals living in it.

However, in his ethical works, Aristotle does not systematically address the question how society should be ordered to allow for the fulfilled life of its citizens. He does address this question in his political writings. In his Politics, he inquires how the "polis" (the city-state) should be ordered to create the conditions for the virtuous development of its citizens, enabling them to live a fulfilled life. We do not want to go into Aristotle's answer to this question here. Parts of it seem to be bound to problems and societal institutions – such as the polis and its households – characteristic of his time. However, we do want to embrace the principle that society should be ordered to create conditions for its citizens to live a fulfilled life and that its institutions, in our case organizations, should contribute to this end.

To explore what this contribution can be, we first need to characterize both the type of society modern organizations are a part of and the role they play in it.

Of course, there is quite a difference between the Greek city-state, the polis, of Aristotle's times and modern society. One way to express this difference is to look at the differentiation of society. In Aristotle's times, segmentation (for instance, into similar households) and stratification (for instance, between nobility and artisans) were prevalent forms of differentiation. Modern society, on the contrary, can be characterized as a predominantly "functionally differentiated" society (Luhmann 1998, p. 743ff).

Luhmann (1998) defines a functionally differentiated society as a society consisting of a manifold of "societal subsystems" each devoted to handling a particular societal problem. Each societal subsystem has its own unique societal function, realizing a particular societal value. For instance, it is the function of the subsystem "law" to incorporate into society the "expectation that normative expectations will be fulfilled," i.e., that we can expect that the rule of law will be maintained. It realizes this function by enabling the production of judicial decisions stabilizing these normative expectations. Another example is the subsystem "economy" that has the function of ensuring the satisfaction of (future) social needs.

According to Luhmann, societal subsystems are autopoietic. They exist by virtue of the production of subsystem-specific communications by subsystem-specific communications. For instance, e.g., judicial decisions link up with earlier decisions and prefigure follow-up decisions.

To this purpose, each societal subsystem, in the first place, has its own "code." This code is a leading distinction specific for that subsystem. For instance, the distinction between "legal" and "illegal" is the code specific for the subsystem "law" (Luhmann calls "legal" and "illegal" the two sides of the distinction). All legal decisions constituting the legal sub-system refer to this distinction between and indicate one of its sides (e.g., a judge deciding that stealing this bread is illegal).

Another example is the economic sub-system. The leading distinction there is that between "paying" and "not paying." According to Luhmann, all communications belonging to the economic sub-system somehow refer to this distinction. By means of reference to the code specific for a sub-system, communications can link up with each other, thereby contributing to its production.

In the second place, societal subsystems develop their own "structures" guiding their autopoietic production. Luhmann calls these structures "programs." For instance, "legislation" and "jurisprudence" are programs of the subsystem "law." In the economic subsystem, "investment programs" structure communication. It is the function of these programs to furnish rules enabling the selection of one of the "sides" of the code. For instance, in the "law-system," labor law and jurisprudence are programs structuring the decision whether it was "legal" or "illegal" to sack this person. Thus, each societal subsystem produces its "own" communications referring to and linking up with other communications of that subsystem. To this purpose, each subsystem has its "own" code and produces its "own" programs structuring the selection of follow-up communications belonging to it.[1]

Because societal subsystems consist of communications referring to other communications of the same subsystem using their own code and programs, they are autonomous. In Luhmann's view, this autonomy characterizes their relation to each other and to modern society as a whole. Modern society is a unity comprising a manifold of societal subsystems. Each of these subsystems exists in the "social environment" of the others and because each subsystem is autonomous, no subsystem can determine what happens in one of the others, let alone that one of them can "preside" over all the others. For this reason, Luhmann (1998) thinks that it is impossible to guarantee the success of attempts at "overall steering" of society: integrating modern society from a central locus of control, in his view, is impossible.

The organizations we know operate in this functionally differentiated society: they are a part of it, they exist by virtue of it, and they contribute to it. To describe their relation to society, it is useful to introduce the distinction between the performance and function of organizations (Teubner 1985).

Performance means that each organization has a set of primary activities selected by that organization, constituting its raison d'être. Examples are, an organization servicing its market with energy or a hospital producing health for its patients. Each of these primary activities requires input from and delivers output to other subsystems within society. Teubner defines "performance" as the set of input/output relations of an organization to other societal subsystems, based on its primary activities (1985, p. 163).

[1] For this reason, it is impossible that communications belonging to one societal subsystem, e.g., "economy," link up with communications belonging to another, e.g., "law," thereby constituting a "mixed" subsystem. However, it is possible that, for instance, the subsystem "law," by means of its *own* communications, code, and programs deals with the question whether an event in the subsystem 'economy' was legal or illegal (e.g., was not paying this bill legal or illegal?).

The function of an organization entails that it by means of its performance contributes to the realization of the function of a societal subsystem. For instance, by producing and selling energy with a profit and by paying wages, the profit organization contributes to ensuring "the satisfaction of future needs," which is the function of the subsystem "economy." By curing its patients, the hospital contributes to realizing the function of the "health-system," and by its verdicts, the court contributes to realizing the function of the "law-system." Teubner defines the "function" of an organization as the – performance related – contribution of that organization to the realization of the function of a societal subsystem (1985, p. 163).

As a rule, organizations have one dominant function, i.e., by means of their performance they contribute to one societal subsystem (the energy supplier primarily contributes to the economic system, the hospital to the health-system, and the court to the law-system).[2] However, this does not mean that by means of their performance, organizations do not affect the functionality of other societal subsystems. For instance, by paying wages and taxes, both the hospital and the court may contribute to the economic system. Or, by providing education to its workers, the profit organization may contribute to the educational system. Although these contributions do not constitute the dominant (or even subdominant) function of these organizations, they still exist, and therefore should be and actually are taken into account in organizations.

This becomes even more important once we see that organizational performance may either be dysfunctional or cause unwanted side effects in other functional subsystems.

In the case of dysfunction, organizational performance obstructs the realization of the function of the societal subsystem to which the organization is primarily geared. For instance, a profit organization may attempt to maximize its performance by distorting competition, thereby undermining the functioning of the economic system.

In the case of unwanted side effects, organizational performance either directly or indirectly negatively affects the functionality of one or more of the other societal subsystems. For instance, to increase profit, the profit organization uses cheap, toxic materials, causing sickness among operators, causing problems for the "health-system." Or, by lacking professionalism of its heart-surgeons, (surviving) ex-patients may start lawsuits against the hospital causing problems for the law-system.

To decrease the probability of performance related dysfunction or unwanted side effects, societal subsystems – the political and law-system in particular – develop "attenuating programs." These programs are directed at decreasing the probability of dysfunction and unwanted side effects. To this purpose, they require some notion of function and wanted side effects, presupposing and involving more or less

[2] In special cases, organizations may have a subdominant function. For instance, a "private" hospital serves the economic subsystem, but also contributes to the health system. In such organizations, a struggle for dominance may occur as the different functions, e.g., contributing to health and participating in a competitive market, impose different requirements on the organization's performance. There are also "hybrid" organizations contributing to different dominant functions.

stabilized societal values and norms. Society enforces these attenuating societal programs onto organizations that have to incorporate them into the premises structuring the production of decisions.

For instance, the political system may develop legislation to regulate the use of toxic materials. Organizations, in turn, should incorporate this legislation into their safety programs, structuring decisions about buying and using toxic materials. Or, the "education-system" may develop programs to improve the professionalism of heart-surgeons. Hospitals may adopt these programs, structuring decisions about the training and development of its medical staff.

To counter dysfunction and unwanted side effects of organizational performance, societal subsystems may develop and enforce all kinds of attenuating programs. Organizations, in turn, incorporate these programs into their decision premises and decisions, thus incorporating societal values and norms into their structures and elements. As a result, organizational decisions not only take into account the goals of their primary activities, i.e., the goals specifying their performance, they also consider societal values and goals related to both their function and performance related side effects. If goals are to make a difference in organizations, they should never refer to organizational performance alone, but also to their societal function and side effects.

In spite of the incorporation of attenuating societal programs into decision premises, imbalances can still exist between the performance of organizations and their function in and effects on society. Even in the extreme case of so-called "planned economies," societal programs leave considerable degrees of freedom to organizations to produce their own decision premises and decisions. Thus, within the framework of both the law and other societal forms of regulation, organizations can still behave more or less "socially responsible." For instance, in spite of complying with the letter of anti-discrimination law, organizations can still discriminate against various groups of people.

To stimulate reflection in organizations on both their social responsibility and the societal effects of their decisions, societal subsystems may develop programs, instituting audits or duties of disclosure (see Teubner 1985, p. 164ff). For example, the requirement to publicly disclose the man/women ratio for each function type may stimulate organizations to put gender discrimination on their agenda, and to implement decision premises that increase the chance that this ratio becomes an explicit decision topic.

Therefore, in addition to attenuating programs, i.e., programs decreasing the probability of dysfunction and unwanted side effects, "amplifying programs" may be in place. These programs increase the probability of an organizational awareness of social responsibility and a reflection on the societal effects of its performance.

Figure 11.1 depicts what we have found until now. The figure shows a number of societal subsystems, each with its own function and programs. These functions and programs define societal values and goals. Some of these programs are "attenuating programs," designed to decrease the probability of dysfunction and unwanted societal side effects of organizational decisions. Other programs are "amplifying programs," developed to increase the probability of internal reflection on the

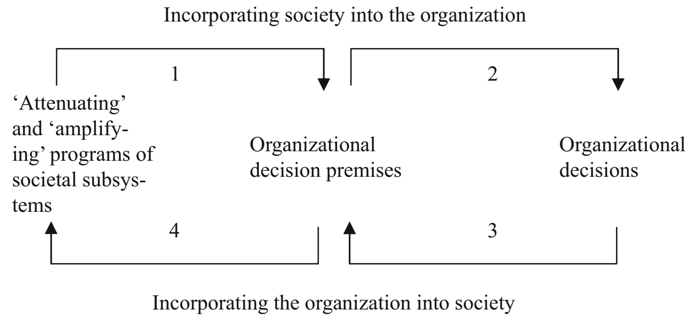


Fig. 11.1 Incorporating the organization into society

societal effects and social responsibility of organizational decisions. Organizations should incorporate both types of programs into their decision premises (Fig. 11.1, *arrow 1*).

Decision premises, in turn, structure the production of organizational decisions (Fig. 11.1, *arrow 2*). What is special here is that in one decision, multiple decision premises play a role. For instance, by means of its decision premises, a profit organization considers a decision about a new primary activity not only from the angle of its performance goals (e.g., its market potential), but also from the angle of its societal function (e.g., contributing to employment) and possible unwanted side effects (e.g., environmental risks). By incorporating societal programs into decision premises, organizational decisions are structured to weigh and integrate both performance-related goals and a variety of societal values and goals.

In this way, organizations have an integrative function. In their decisions, performance goals, attenuating programs, and amplifying programs of different societal subsystems come together. The integration of programs produced by different societal subsystems is an important organizational contribution to society. Decisions about goals in organizations always involve weighing and integrating both performance- and a multiplicity of societal values and goals.

By incorporating societal programs, values, and goals into their decision premises and by incorporating these decision premises into their decisions, organizations "incorporate society into themselves," i.e., into their structures and decisions. This is what arrows 1 and 2 of Fig. 11.1 amount to.

By actually producing decisions that balance performance and societal goals and values, organizations reinforce decision premises incorporating these goals and values. Reinforcing these premises, in the first place, can strengthen the organization's awareness of and reflection on its societal role and responsibility (Fig. 11.1, *arrow 3*). Reinforcing these premises, in the second place, can contribute to reinforcement of the societal programs they incorporate. As organizations by means of their decisions comply with decision premises incorporating societal programs, these programs themselves are reinforced becoming a common societal practice (Fig. 11.1, *arrow 4*). Therefore, by incorporating societal programs into their decision premises and decisions, organizations not only incorporate society into themselves, they also "incorporate themselves into society." That is what arrows 3 and 4 amount to.

As it appears, the relation between organization and society is one of mutual incorporation. In the next section, we explore what is required if this relation of mutual incorporation is to contribute to the organization's rich survival. Based on these requirements, we specify principles for the design of organizational structures supporting this type of survival.

11.3 Rich Survival: Specific Design Principles and Corresponding Structures

To find organizational structures that can support rich survival we proceed in two steps. First, we specify requirements to and design principles for organizational structures that can support an organization's rich survival (Sect. 11.3.1). Second, we describe–in terms of de Sitter's parameters–the structures that fit these design principles (Sect. 11.3.2).

11.3.1 Requirements to and Design Principles for "Rich" Organizational Structures

To find functional principles for the design of "rich" structures, we organize this subsection into two parts.

In Sect. 11.3.1.1, we zoom in on the relation between organization and society in order to define what is required for organizations to make "rich" contributions to society by means of their primary activities.

Section 11.3.1.2 addresses the question what this means for the design of their production and control structure. We inquire how organizations should be structured to increase the probability that they act responsibly by means of the selection and realization of their goals.

In Sect. 11.3.1.3, we "open" the organization and inquire into the principles for designing structures that allow organizational members to act responsibly. Again, we ask ourselves what is required for responsible behavior and which design principles can be derived from this.

Finally, in Sect. 11.3.1.4, we discuss how the design principles attenuate the disquieting features of the disciplines.

11.3.1.1 Responsible Organizations: Requirements

In this section, we ask what is required for organizations to incorporate themselves into society in a responsible way. To this purpose, we depart from the assumption of an organization incorporating itself into a just society, i.e., a society actually creating conditions for its citizens to live a fulfilled life by means of its societal subsystems and their programs. Later, we make amendments enabling us to drop, or at least mitigate, this assumption.

If organizations are to incorporate themselves into a just society, this in the first place means that they incorporate relevant societal programs into their decision premises. In the second place, it means that these premises are incorporated in the decisions regarding the selection and realization of their performance goals.

To further specify what this process of incorporation entails, we should take into account that we can look at it in at least two ways: in terms of the isolating or in terms of the inclusive mode of reasoning about purposeful behavior.

In the "isolating" mode, incorporation consists of nothing more than a "cost/benefit" analysis, weighing "costs" and "benefits" of compliance or non-compliance with relevant programs. For instance, criminal organizations are highly selective on this point. They incorporate societal programs into their decision premises and decisions, not because it is the right thing to do, but because it is beneficial to their particular contingent goals. In a similar vein, perfectly legitimate organizations may incorporate, for instance, laws against pollution for the sole reason of avoiding customer strikes. In such cases, incorporating the organization into society by incorporating societal programs into the organization becomes a more or less complex technical process treating these programs as means to realize the organization's contingent goals. Incorporating society into the organization (in order to incorporate the organization into society) is not regarded as something of value, but rather as a question of compliance or non-compliance to "alien" values, norms, or programs that may be enforced by society. Thus, the question of the organization's incorporation into society reduces to that of compliance or non-compliance with societal programs, which in turn reduces to weighing their respective "costs" and "benefits" relative to the organization's contingent goals.

In the "inclusive" mode, organizations incorporate societal programs into their decision premises because they decide it is the right thing to do, and they decide it is the right thing to do, because they consider these programs as enabling citizens to live a fulfilled life. Moreover, decision premises are applied in a particular case, not (only) because they incorporate these programs into organizational decisions, but because their application is judged to be the right or appropriate thing to do in the given circumstances.

The isolating and the inclusive mode have in common that incorporation of societal programs is never a "mechanical" process relying on "blind obedience," just because these programs are in place. In both cases, incorporation involves selection, interpretation, deliberation, and judgment: "Which programs or premises are relevant?", "How should one interpret and adapt societal programs to fit into the web of organizational decision premises?", "How should one interpret, weigh, and integrate program-based premises and the values implied in them to apply them in concrete decision situations?" These are all questions related to incorporation one must answer both in the isolating and in the inclusive case. The difference between the two is not that deliberation and judgment are involved, but rather the kind of deliberation and judgment

For "isolating incorporation," the organization's contingent goals are the point of departure. Although these goals may be the object of reflection and adaptation, they are not reflected in terms of their contribution to the creation of a just society, i.e., a society enabling its citizens to live a fulfilled life. In terms of the isolating mode, each contingent goal is as good as any other. These contingent goals, then, guide technical processes of incorporation, weighing performance, function, and side effects solely in terms of costs and benefits for realizing these goals. The main question here is, "Is it smart to incorporate just societal programs to efficiently realize contingent organizational goals?"

If we want to argue against contingent and minimal goals, this isolating/calculating/technical mode of incorporating societal programs into organizations should not be the one to pursue.

By contrast, "inclusive incorporation," aims at contributing to the creation of societal conditions for living a fulfilled life. Programs are incorporated into decision premises and decision premises are applied to concrete situations because this is judged the right thing to do.

The quality of this kind of incorporation depends on the practical wisdom and moral virtue of the judgments of "agents" (organizational members, organizations) about the equity of societal programs, the decision premises incorporating them, and their application in particular situations. The question no longer is, "Is incorporation a smart thing to do, i.e., does it realize our contingent goals in the most efficient way (taking into account the possible costs incurred by punishment)?", but rather, "Is incorporation the right thing to do in the given circumstances (irrespective the cost incurred by punishment)?"

If we reserve the word "integrity" for "inclusive," i.e., practically wise and morally virtuous, incorporation of societal programs into organizational decision premises and concrete decision situations, we can distinguish two instances of integrity allowing us to drop the not very realistic assumption made above of incorporation into a wholly just society. The two instances of integrity are: (1) integrity of incorporation and (2) integrity beyond incorporation.

By integrity of incorporation, we mean that societal programs are incorporated into organizational decision premises because it is considered the right thing to do, i.e., because these programs are considered to enable citizens to live a fulfilled life. In this case, an organization incorporates itself into society by incorporating these programs into its decision premises because it considers these programs as just programs, contributing to a society with just programs. Moreover, by integrity of incorporation we mean that decisions incorporate program-based decision premises because it is judged the right thing to do in the particular circumstances. In this case, decisions reinforce these decision premises, because they are considered right and appropriate in the given circumstances.

By integrity beyond incorporation, we refer to deviations from societal programs or program-based decision premises in concrete cases, because deviation is judged the right thing to do in the given circumstances. Integrity beyond incorporation provides some space to organizations to compensate for either failing or societal programs they consider unjust. For instance, organizations can do more than societal programs require, voluntarily applying stricter rules because they decide that these programs have omissions that should not be there. Organizations can also do less than prescribed, once again, because they judge that this is the right thing to do.

For instance, they may decide that it is not right to incorporate the "letter" of the law into decision premises contradicting its "spirit." Or, they may make exceptions to societal programs, applying program-based decision premises with discretion dependent on the salient features of the concrete situation. In addition, organizations may decide to lobby or protest against societal programs because they are considered unjust. In extreme cases, organizations may even decide to negate or overstep programs positively sidelining basic requirements for citizens to live a fulfilled life.

It may be clear that integrity beyond incorporation requires even more of organizations and their members than integrity of incorporation. Particularly negating or overstepping societal programs – legislation in particular – requires careful consideration. In these cases, the organizational cultivation of virtues like veracity and courage become even more important. For instance, deviating from a particular law because one considers it unjust, not only requires practical wisdom, but also endurance, veracity, and courage.

Given our analysis, we can now specify what it means that goals should make a difference in organizations. Moreover, we can specify the functional requirements related to it.

Unlike the discipline's contingency and minimalism of goals, goals should make a difference in organizations. This means that organizational decisions should contribute to society being ordered to enable its citizens to live a fulfilled life.

In the first place, this entails that organizations incorporate themselves into society by incorporating society into their decision premises and decisions. This requires that relevant societal programs are selected, interpreted, and integrated in the organization's decision premises and that these decision premises structure organizational decisions. In this way, organizational decision premises and decisions are never only performance related. Decisions about goals always involve some wider societal perspective, weighing performance, function, dysfunction, and unwanted side effects.

In the second place, this entails that the mode of incorporation should be "inclusive." This requires integrity of and integrity beyond incorporation. That is societal programs should not be incorporated because it is the technically smart thing to do relative to contingent organizational goals, they should be incorporated because it is the right thing to do, because they contribute to the creation of conditions needed for living a fulfilled life.

We are aware that these are quite high expectations that hardly ever are and perhaps never can be wholly fulfilled. To begin with, we departed from the unrealistic assumption of organizations participating in a wholly just society, ordered to create conditions for the fulfilled life of its citizens. By introducing the concept of integrity beyond incorporation, we were able to mitigate this assumption. In less than just societies, organizations can and should resort to measures such as protesting or lobbying against unjust programs to salvage integrity and contribute to society's improvement. In this way, they can contribute to the improvement of a less than just society.

However, there is a second problem, for is the assumption of "organizational integrity" itself not rather far-fetched?

In many organizations, integrity seems to be a rare thing. Their infrastructure seems to reward, and thereby, induce "isolating" rather than "inclusive" reasoning. And even if "inclusive" reasoning exists regarding some issues, it seems to be "complemented" by "isolating," merely technical compliance to societal programs regarding other issues. For instance, most organizations do not kill underperforming highly substitutable workers. In most cases, the reason for this is not only that killing them is not "smart" or cheap (for this might well be the case), but because it is not the right thing to do. The same holds in a good many instances for the use of child labor, or other issues touching on deep rooted societal values or fundamental human rights. However, the same organizations may treat less fundamental values and rights as objects of technical reasoning, calculating the "costs" and "benefits" of compliance or non-compliance. It appears that, in the best case, organizations combine integrity and calculation, sometimes posing the latter as the former.

For this reason, enforcement of societal programs and the rights, values, or norms implied in them, seems to be indispensable. In the first place, enforcement of, for instance, the law, functions as a societal safety net. It increases the probability of tipping the balance of calculative cost/benefit reasoning in favor of compliance with values and norms considered important. In the second place, societal programs and their enforcement indicate what is socially acceptable and what is not. In this way, they have an "educational" function. By punishing trespassing and trespassers, enforcement indicates that certain behavior is socially unacceptable. Moreover, programs and their enforcement institute "social routines" of compliance or non-compliance, culminating in a "culture" of, for instance, paying or evading taxes. Finally, enforcement may increase the probability of reflection on the values and norms implied in societal programs. Why is some behavior acceptable or unacceptable in a particular situation? What is the societal value implied in a program, and what is the value of this societal value?

What counts, in the end, is that we not only reflect upon and reason about means towards contingent ends, but also about ends that can be ends in themselves. As argued, this kind of reasoning is of a fundamentally different kind than calculative, technical, isolating reasoning. It requires integrity, i.e., moral virtue and practical wisdom, and even if these virtues are factually scarce in organizations, integrity of and beyond incorporation remain counter-factual "measures" or "ideals" needed to judge the contribution of organizational decisions and decision premises to the creation and improvement of a society enabling its members to live a fulfilled life. By its very nature, calculation alone cannot fulfill this function. As argued, it does not provide opportunities for reflection on the value of goals we set and on their contribution to the fulfilled life. Therefore, if we want to be able to judge the integrity of organizational decisions and decision premises, i.e., if we want to be able to judge their importance to our lives with others in a community, we cannot rely on calculation alone. Rather, we must invoke inclusive reasoning and the moral virtue and practical wisdom involved in it. Thus, although we are aware that calculative, isolating reasoning is prevalent in organizations, this is no reason to drop the appeal to integrity.

11.3.1.2 Responsible Organizations: Design Principles

If we want to take seriously our conviction, that goals should make a difference in organizations, then integrity of and beyond incorporation are prerequisites, and organizational structures should be designed to enable them. To this purpose, the organization's performance, function, decision premises, and decisions should aim at contributing to the creation of a "just" society enabling its citizens to live a fulfilled life.

Phrased in the terminology of Beer's Viable System Model, this means that throughout the organization, at all its levels of recursion, policy, intelligence, and control, should develop; and control, coordination, and primary activities should realize performance goals not only in reference to organizational viability, but also in reference to a just society and the organization's contribution to it. We say "throughout the organization" because integrity of and beyond incorporation, clearly, is not something that is realized exclusively by some corporate staff-department devoted to "ethical problems." It rather is a problem for all organizational units, each of them contributing its bit to the organization's contribution to society. To make integrity into a responsibility throughout the organization and to provide the capacity to live up to it, we formulate three principles for the structural design of organizations.

Design Principle 1

The production structure should be designed in such a way that the organization down to its smallest units reflects – as much as possible – the organization's performance and societal function.

This first principle secures that at all levels of recursion the contribution of organizational units (e.g., divisions, departments, teams) to the organization's contribution to society is both as inclusive and as clear as possible.

An example of such a production structure would be a railway company designed in units, each devoted to servicing its "own" geographical area with its transport capabilities. In this company, each unit services its own clients with the whole of the organization's performance and societal function (inclusiveness), reflecting them as clearly as possible in its own activities. This layout enables the organizational performance and function to make a difference throughout the organization, setting the stage for local processes of incorporation that are in tune with the organization's incorporation into society. Compare this to the same railway company with a different structure comprising one unit specializing in "ticket sales," one in "transport," and another in "maintaining infrastructure" (taking care of trains, rails, stations, etc.). Now, each unit performs a specialized task that is only a part of the organization's performance and is not a clear reflection of its function. As each unit is held accountable for its own performance and is geared to its own specialized societal problems and programs, it may loose sight of the performance and societal function of the organization as a whole, complicating its incorporation into society.

Design Principle 2

Given the first principle, the control structure should be designed in such a way that all levels of recursion – i.e., from the smallest units up to the organization as a whole – have the regulatory capacity needed to control, design, and regulate the performance and function as well as the possible dysfunctions and unwanted side effects of their primary activities.

This principle secures that both the organization as a whole and its lower level units have the regulatory capacity to contribute to the incorporation of the organization into society. So, in addition to the production structure setting the stage for local processes of incorporation, the control structure provides each organizational unit with the regulatory capacity to actually incorporate relevant societal programs, allowing for local discussions about the relation between the unit's and the organization's performance and function. More in particular, regulatory tasks should allow units at all levels of recursion to direct themselves to both society and the organization's contribution to it, as well as to their own contribution (to the organization's contribution) to society. Moreover, they should enable them to select societal programs relevant to their contribution, to interpret and adjust them, and fit them into their decision premises. Finally, they should enable the organization and its units to select, weigh, and integrate these decision premises to structure the decisions by means of which they actually incorporate themselves into society.

Design Principle 3

Given the first and second principle, the control structure should be designed in such a way that units at higher levels of recursion only interfere with primary activities and their regulation at lower levels if these endanger the integrity of incorporation of the unit at that level and at the higher level (Beer's principle of freedom, see Beer 1995).

This principle, in the first place, provides each lower level unit with the freedom and responsibility for its own integrity of incorporation. In the second place, it discharges the higher units from the obligation of controlling and regulating everything that is going on at the lower levels, thereby recursively distributing the total complexity of the process of incorporation over all units.

By means of these three principles, the total problem of incorporating the organization into society is "unfolded" and distributed over the organization. Moreover, this "unfolding" and "distribution" is done in such a way that incorporating the organization into society becomes a "shared" problem, reflecting itself throughout the organization.

More in particular, the first principle focuses at the grouping and coupling of primary activities. It aims at securing that lower level units, by means of their primary activities, "keep in touch" with the total problem of incorporation. The second and third principle specify how regulatory tasks should be distributed to allow, at all levels of recursion, for adapting and implementing decision premises contributing to the organization's incorporation into society (below, in subsection 11.3.2, we show which type of organizational structures can realize the requirements specified here).

11.3.1.3 Responsible Organizational Members: Inclusive Jobs

These principles mentioned above are in vain if organizational members responsible for decisions about decision premises do not get the chance to maintain or further develop their integrity in organizations. So, in addition to these principles, additional principles are needed that allow for the design of structures enabling the development of the integrity, the moral virtue and practical wisdom, of responsible organizational members. In this subsection, we explore these principles.

Before we take on this challenge, a possible counter-argument should be addressed. Moral virtue and practical wisdom, as Aristotle explains it, are virtues developing from our earliest childhood, culminating in a more or less virtuous, stable, and consistent character.

Given this developmental process, one could argue that organizational contributions to the development of the moral virtue and practical wisdom of their members can be only very modest because their characters have already been formed before they start their organizational careers. In answer to this argument, we want to propose the following considerations.

Indeed, organizational contributions to the moral development of their members should not be over-estimated. In principle, it is hard to see how organizations alone can "heal" a society that either is already corrupt or neglects the education and development of its young citizens. However, in the type of societies we live in, organizations play an important part in the education and development of children and young citizens. In this sense, they have a bearing on the development of their character long before they start to play their role in organizations. This provides a good reason indeed to reflect on structures enabling the integrity of organizational members.

Indeed, organizational structures tend to induce "isolating" rather than "inclusive" reasoning. And this may lead too the all to common and sad figure of organizational members "leaving their integrity at home," putting on their task-focused, technical, and therefore, a-moral "dustcoats" as they arrive at their work-places, shielding them off from the usual bites of consciousness. To counter this tendency of inducing "isolating" rather than "inclusive" reasoning in organizations, a reflection on organizational structures appealing to the best of organizational members, creating chances for nurturing instead of diminishing their virtues, seems to be a worthwhile challenge.

And indeed, it may true that we develop our character before becoming organizational members, we always already are – to some extent – virtuous or vicious. Yet, even virtuous persons entering organizations or organizational members changing jobs within their organization, may still have to learn to respond appropriately to the difficulties presented by their new tasks. These tasks may confront them with problems they never had to face before, requiring a refocusing of the virtues they already possess.

The question, therefore, remains, "What requirements need to be met by organizational structures to allow for the development and maintenance of the integrity of organizational members?" This also is where the discussion of the five remaining disquieting features starts, for they all relate to this question.

If we look at these five disquieting features, we can see that they all connect to the "anthropology" underpinning the disciplines. More in particular, they depict human beings as a kind of trivial machines; programmable and re-programmable by operand conditioning, "motivated" by fear, performing operations contributing to contingent goals alien and possibly unknown to them, and examined as "cases" to increase their performance.

This image contrasts sharply with Aristotle's views on man. He does not think of human beings as a kind of trivial machines. For him, human beings are intentional and intelligent beings with more or less sophisticated emotions and desires that can and should develop and blossom in the course of their lives.

For this reason, human beings should not be "motivated" by fear, but by what they consider to be right, by what in the given circumstances is judged to contribute most to living a fulfilled life. The goals they pursue should not be alien or unknown to them; they should rather be most intimate to their very existence, contributing here and now to the realization of their humanity. Because of their intelligence, it is also impossible to reduce human learning to "blind" operand conditioning. For human beings, learning also involves experience, awareness of, and reflection on themselves, their motives, and actions in their social environment.

In this way, Aristotle's conception of human beings and their development seems to offer opportunities to counter the disquieting features attached to the disciplines. Because this conception is connected to moral virtue and practical wisdom, it also allows us to define principles for the design of structures enabling organizational members to develop their emotions, desires, and intellect in accordance with the ideal of moral virtue and practical wisdom. Based on these principles, organizational structures can be designed that do not predominantly:

1. Reduce but mobilize feelings of involvement and responsibility
2. Rely on fear, but on what appears as worthwhile, as a motivator of action
3. Trivialize – principally non-trivial – behavior, but capitalize on human intelligence, reflection and learning
4. Shape categories and norms for monitoring behavior by operand conditioning, but allow for reflective learning by doing
5. Invest in tools further trivializing behavior, but search for instruments enabling human development.

To specify these design principles, we take seriously Aristotle's dictum that living a fulfilled life implies "an activity of the soul in conformity with virtue, and if there are more virtues, in conformity with the best of them." Applied to man's practical life this introduces three basic requirements to organizational structures:

1. They should contribute to the development of (technical) skills
2. They should contribute to moral virtue and practical wisdom
3. They should contribute to the guidance of the use of skill by moral virtue and practical wisdom.

To show what design principles can support the realization of these requirements, we introduce the concept of "inclusive jobs." With regard to this concept, we do two things. First, we describe what we mean by it and formulate requirements inclusive jobs impose on organizational structures. And second, we show how inclusive jobs can function as an alternative to the tasks specified by the disciplines reducing the disquieting features implied in them.

In order to describe inclusive jobs, please examine Fig. 11.2. In this figure, it can be seen that inclusive jobs are designed to combine the application and development of practical wisdom, moral virtue and skill.

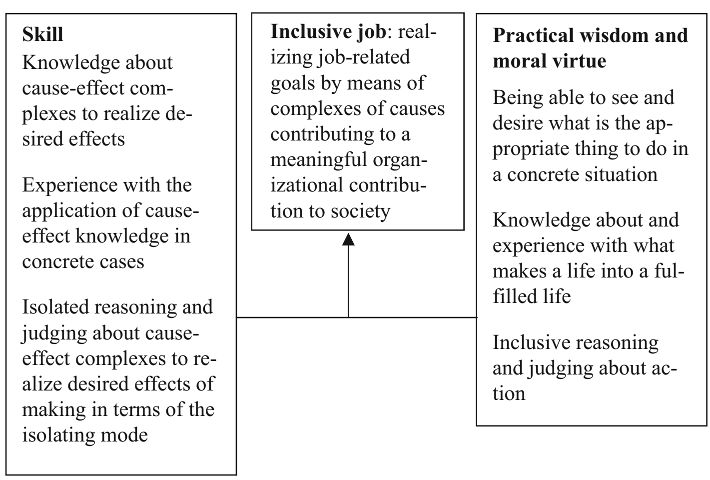


Fig. 11.2 The structure of the inclusive jobs

It might be instructive to start with some examples of inclusive jobs. The most obvious examples are the so-called "professions" situated in their related professional organizations; e.g., the "surgeon" in a hospital, the "lawyer" in a law firm, the "consultant" in a consultancy firm, the "furniture maker" in her own studio, etc. Each of these jobs is devoted to the organization's contribution to society, and is sufficiently rich to reflect this contribution as much as possible. Moreover, each of these jobs involves intelligent situational decision-making regarding means, both requiring and developing the skills of the persons performing them.

From the examples, we can see in the first place that inclusive jobs are designed to foster awareness of, reflection on, and communication about both the appropriateness of and relation between organizational and situational job-related goals. In this way, they create chances for developing moral virtue and practical wisdom. In organizations where goals make a difference, inclusive jobs reflect the performance and function of the organization as much as possible, thus enabling their meaningful local interpretation.

Without this goal-related aspect, i.e., depending on skill alone, the job would lack "direction." It could be unleashed to produce whatever societal effects, without questioning their contribution to the organization's contribution to society.

In the second place, inclusive jobs are designed to foster skills. This means that they are sufficiently complex to allow for gaining knowledge about cause-effect relations related to the goals of the job. Moreover, they allow for acquiring experience with applying this knowledge in concrete situations.

Finally, they are sufficiently rich to allow for the development of capacities for "job-related" deliberation and judgment. Without this skill-related aspect, inclusive jobs depending on practical wisdom and moral virtue alone would lack "efficacy." They would not provide chances to develop the calculating part of our intellect to find causes producing in the individual case the desired effects related to the job with high predictability and efficiency.

Thus, directed by the ends of practical wisdom and moral virtue and drawing from skill, inclusive jobs aim at realizing complexes of desired effects by means of complexes of causes in order to contribute locally to a meaningful organizational contribution to society.

Now, it could be objected that in most organizations this type of jobs is quite rare; (1) because the relation between the job and the organization's performance and function has been lost or (2) because the job's complexity is so low that it does not allow for applying and developing job-related skills.

For instance, simple jobs at production lines consisting of small, standardized tasks neither reflect the organization's performance and function nor provide opportunities for developing skills. On the contrary, it is essential to this type of jobs that they can be easily learned by in principle anyone. Soon they become a "blind" routine, demanding very little from the knowledge, experience, awareness, reflection, deliberation, and judgment of the organizational members performing them. Precisely for this reason, they offer opportunities for easy substitution and apparent, yet only superficial, efficiency, and seem to be omnipresent in the many existing functionally organized, “bureaucratic” organizations. To counter this objection, we make two remarks.

First, the factual rarity of inclusive jobs in organizations does not mean that the concept of “inclusive jobs” should not function as a counter-factual ideal to evaluate the merits of actual job design in organizations.

Second, the fact that jobs are actually designed to comprise simple, standardized tasks that do not reflect the organization’s performance or function does not mean that it is impossible to formulate principles and design organizational structures allowing for jobs that are as inclusive as possible, making room for the development of job-related skills, moral virtue, and practical wisdom. On the contrary, in addition to the three design principles derived in the previous subsection, three additional principles can be formulated.

Design Principle 4

The design of the production structure should as much as possible allow jobs to reflect the performance and function of the organization, amplifying the probability that the organization’s contribution to society is acknowledged and translated into meaningful “local” contributions.

This requirement secures that even at the level of individual jobs, goals can make a difference in organizations (as much as the division of work permits). In this sense, design principle four complements design principle one mentioned above.

Design Principle 5

The design of the production structure should include jobs to be sufficiently rich to allow for the development of job related skills and sufficiently connected to allow for job related communication.

Design Principle 6

The design of the control structure should include for the internal, external, routine and non-routine regulation needed to steer the job’s operational tasks and to secure job-related communication with other organizational members.

11.3.1.4 Inclusive Versus Discipline-Like Jobs

In the previous sections we specified six design principles for “rich meaningful survival.” In Sect. 11.3.2 we show what kind of organizational structures fit these six principles. However, before we go into these structures, we want to argue how inclusive jobs provide conditions for attenuating the disquieting features attached to discipline-like jobs, and, therefore, function as an alternative to them.

1. Against non-involvement and non-responsibility. By their reflection of the organization's performance and function, inclusive jobs can amplify local feelings of involvement with and responsibility for the organization's contribution to society. Other than the simple tasks of the disciplines that can be configured and refigured to serve whatever organizational purposes, and in this sense, are loosely coupled to them, inclusive jobs reflect these purposes as clearly as possible, allowing organizational members to identify themselves with and take pride in the organization's contribution to society.
2. Against fear as prime motivator. The coupling of jobs to the organization's contribution to society also can function as an "intrinsic" and "positive" motivator. Unlike the discipline-induced fear for punishment by others, inclusive jobs allow for intrinsic motivation because they can be perceived as worthwhile contributions to the organization's contribution to society. Similarly, unlike the discipline's "negative" motivation by fear (i.e., avoiding "pain" in whatever form), inclusive jobs allow for positive motivation by local contributions perceived as worthwhile in reference to both one's own development and the organization's contribution to society.
3. Against trivializing behavior. Inclusive jobs do not trivialize the principally non-trivial behavior of organizational members. Because they are designed to foster the development of job-related skills, moral virtue, and practical wisdom, they capitalize on learning processes involving job-related knowledge and experience as well as situational deliberation and judgment with regard to both means and ends.
4. Against operand conditioning. The types of learning related to inclusive jobs differ from the operand conditioning of the disciplines. Because jobs are designed to allow for the development of skills, moral virtue, and practical wisdom, they involve "cognitive learning" regarding cause-effect complexes related to the job's goals. Moreover, they involve "learning on the job" to acquire experience and to improve deliberation and judgment both to apply cause-effect knowledge in concrete situations and to evaluate local decisions in terms of their contribution to living a fulfilled life. Both types of learning do not treat organizational members as black boxes, but as human beings; aware of their social and operational context, of their motives and reasons, and reflecting on and learning from their operations. Organizational members performing inclusive jobs have their own learning history – ranging from novice to expert – in which they develop their job-related skills and virtues.
5. Against research for tools supporting trivialization. The abovementioned differences between inclusive jobs and jobs defined by the disciplines require the development of different types of instruments and systems to support them. In the case of the disciplines, research aims at developing tools for trivializing human behavior, contributing to its operand conditioning. In the case of inclusive jobs, research is re-directed. It is driven by the desire to find organizational structures to enable organizational members to develop the virtues related to the practical dimension of their existence as much as the division of labor admits.

As it appears, it is possible to formulate requirements to and design principles for organizational structures that can help to reduce or even overcome the disquieting features attached to the disciplines.

First, against the contingency and minimalism of goals, we formulated three principles for the design of organizational structures. Briefly, these principles allow organizational members at all levels of recursion to "keep in touch" with the organization's performance and function as well as with possible dysfunctions and unwanted side effects. They allow for the recursive unfolding of the problem of incorporating the organization into society over the organization. By applying them, integrity of and beyond incorporation can become an issue throughout the organization.

Although the first three principles set the stage for integrity in organizations, they do not yet create conditions to develop the required moral virtue and practical wisdom of their members. To this purpose, we defined a second set of principles that should guide the design of inclusive jobs. By applying these principles, jobs are designed to allow for the development of (1) job-related skills and (2) the moral virtue and practical wisdom to deliberate about and judge the appropriateness of job-related decisions and decision premises.

On closer examination, it can be argued that the two sets of principles complement each other. The first set allows organizational goals to make a difference throughout the organization, setting the stage for the integrity of the organization's incorporation into society. The second allows for the development of the job-related skills, moral virtue, and practical wisdom needed realize the potentials created by the first set of requirements.

Given these design principles, the question presents itself whether it is possible to find organizational structures that conform to them. This question is quite urgent, for if we are unable to find such structures, we are forced to share the fate of the business ethicists or "critical" students of management who are able to point into the right direction without being able to show how to get there.

11.3.2 Production and Control Structures Fitting the Design Principles

It may be nice to have a set of principles for the design of organizational structures supporting rich survival. However, it is better to show that there are structures that conform to these principles. To show what these structures look like, we relate the principles formulated above to de Sitter's design parameters.

In Chap. 7, we explained that, according to de Sitter (1994), organizational structures can be described in terms of seven parameters and their values. We also distinguished two extreme structural types: (1) "high parameter-value structures," i.e., complex structures with simple jobs and (2) "low parameter-value structures," i.e., simple structures with complex jobs. Moreover, we explained that different structural designs have different characteristics with respect to three variables considered essential to the meaningful survival of organizations: "quality of organization," "quality of work," and "quality of working relations." Given this basis, de Sitter showed how structures should be designed to increase organizational performance in terms of the three quality criteria.

In his work, de Sitter, quite convincingly argues that simple low parameter-value structures with complex jobs designed according to the controllability principle are to be preferred over complex high parameter-value structures with simple jobs. According to de Sitter, simple structures with complex jobs systematically increase the value of the three "quality variables" essential for meaningful survival.

In the present section, we have a problem similar to de Sitter's, for we are interested in organizational structures that can support rich survival, reducing the disquieting features attached to the disciplines. In the previous section, we argued that to realize this purpose, six principles should be applied to the design of organizational structures. Therefore, to answer the question what structures contribute to an organization's meaningful survival, we relate the six principles to the de Sitter's parameters and discuss which parameter values are needed to realize them. Below, we discuss the production structure first and then the control structure.

11.3.2.1 Rich Production Structures

As argued in Chap. 7, the production structure can be described in terms of the first three parameters: the level of functional concentration, the level of differentiation between making, supporting, and preparing, and the level of specialization of tasks. So, these parameters need to be taken into consideration. The design-principles that have a bearing on these parameters are principle 1, 4, and 5. Table 11.1 lists the relevant principles and parameters. Moreover, it indicates which values result from the application of the principles to the parameters. In the text below the table, we argue why these values are needed.

Design principle 1 and 4 secure that at every level of aggregation (from the organization as a whole down to individual jobs), organizational members can be – as much as possible – in touch with the organization's performance and function, allowing the organization's contribution to make a difference throughout the organization.

A high level of functional concentration has a negative impact on realizing this requirement. As explained in Chap. 7, functional concentration implies, that activities of the same type are concentrated in specialized departments and that all orders and order types are routed through these departments. As a result, organizational members concentrate on their departmental activities, losing sight of orders and order-types, and as a result, of the organization's performance and function. By means of functional de-concentration, i.e., parallel flows related to their own order type(s), the organization's performance and function remain as clear as possible to organizational members in these streams. A low level of functional concentration, therefore, has a positive influence on realizing requirements one and four.

Table 11.1 Rich production structures

Parameter : Design principles	1. Level of functional concentration	2. Level of differentiation: preparation, making, supporting	3. Level of specialization
1. The production structure should be designed in such a way that the primary activities from the organization down to its smallest units – as much as the division of labor permits – reflect the organization's performance and societal function(s).	LOW	LOW	LOW
4. The design of the production structure should as much as possible allow jobs to reflect the performance and function of the organization, amplifying the probability that the organization's contribution to society is acknowledged and translated into meaningful 'local' contributions	LOW	LOW	LOW
5. The design of the production structure should allow jobs to be sufficiently rich to allow for the development of job related skills.	LOW	LOW	LOW

Given a low level of functional concentration, low levels of differentiation and specialization have an amplifying effect on the realization of requirements one and four. By integrating preparation and support activities with parallel "making" streams, these activities are coupled to specific order-types. This means that organizational members devoted to these preparation and support can keep in touch with the performance and function of the organization. If preparation and support are differentiated and decoupled from streams, or even worse, if they are differentiated and coupled to functionally concentrated departments, the capacity to keep in touch with the organization's performance and function decreases. High levels of differentiation, especially combined with a high level of functional concentration, have a catastrophic effect on the organizational capacity for realizing requirements one and four. Something similar holds for specialization. Given a low level of functional concentration, low levels of specialization, further increase the chances for organizational members to keep in touch with "whole" orders or order types. As specialization increases, members become responsible for an increasingly small part or aspect of an order or order type, eventually losing sight of the organization's performance and function.

To summarize, realizing design principles 1 and 4 primarily depends on a low level of functional concentration. Given functional de-concentration, low levels of differentiation and specialization further amplify opportunities for organizational members to keep in touch with the organization's performance and function.

Design principle 5 applies primarily to the design of jobs. It is formulated to ensure that chances are created for the development of job-related skills and virtues. This entails that jobs are sufficiently rich to allow for acquiring job-related knowledge and experience as well as for practicing job-related deliberation and judgment. Moreover, principle 5 ensures that jobs are defined to allow for job-related communication.

To realize design principle 5, high levels of differentiation and specialization are undesirable. They diminish job-content, concentrating jobs on small aspects or parts of the entire production process, allowing only for developing "mindless" routines. Integration of preparation, making, and support activities in the job as well as de-specialization, increase the chances for developing job-related skills. In addition, low levels of differentiation and specialization require external and non-routine forms of regulation (see design principle 6 below), adding to communication and learning opportunities offered by the job.

Of course, jobs are not defined in isolation. They are part of a larger network. To allow for the development of job-related skills, moral virtue and practical wisdom, the larger network should be designed to meet design principles 1 and 4. In this way, individual jobs are embedded in order-related teams allowing for job-related communication, that are a part of order-related segments, that, in turn, are part of order-related streams, securing – as much as possible – the job's connection with the organization's performance and function.

11.3.2.2 Group 2: Rich Control Structures

Design principles 2, 3, and 6 directly relate to the control structure. In our discussion of the desired values of the parameters characterizing the control structure, we presuppose that the production structure has been designed according to the characteristics specified above. The reason for this is that control structures are defined relative to production structures and the parameter values characterizing them (see Chap. 7, Sect. 3). Table 11.2 provides an overview of the desired parameter values needed to conform to the listed design principles.

The three control-related design principles are specified to ensure that at all levels of recursion (including the level of individual jobs), organizational members have the capacity to solve job-related problems coupled to the organization's performance and function, with as little interference from higher echelons as possible. This is needed for two reasons. First, if goals are to make a difference, organizational members must have the (local) regulatory capacity to make a difference to them. This requires a well-designed production structure as well as local potentials for external and internal non-routine regulation. Second, if organizational members are to develop job-related skills, moral virtue, and practical wisdom, they must be allowed to make job-related decisions about local goals and means. To realize these requirements, the control structure should be designed according to de Sitter's principle that regulatory capacity is allocated where disturbances actually occur.

Table 11.2 Integrity requirements and the control structure

Parameter : Requirement:	4. level of separation between production and control	5. level of differentiation: monitoring, assessing, intervening	6. level of differentiation: strategic, design, operational regulation	7. level of specialization
2. Given requirement one, the control structure should be designed in such a way that all levels of recursion – i.e., from the smallest units up to the organization as a whole – have the regulatory capacity needed to control, design, and regulate the performance and function as well as the possible dysfunctions and unwanted side effects of their primary activities.	LOW	LOW	LOW	LOW
3. Given requirements one and two, the control structure should be designed in such a way that units at higher levels of recursion only interfere with primary activities and their regulation at lower levels if these endanger the integrity of incorporation of the unit at the higher level.	LOW	LOW	LOW	LOW
6. The design of the control structure should allow for the internal, external, routine and non-routine regulation needed to steer the job's operational tasks and to secure job-related communication with other organizational members.	LOW	LOW	LOW	LOW

This implies that levels of separation are relatively low. At all levels of recursion, tasks should comprise the regulatory transformations to deal with task-related problems. The same holds for the level of differentiation of control into parts. For instance, to enable the development of job-related skills, organizational members should be able to monitor, assess, and intervene in job-related problems. The level of differentiation of control into aspects should also be low. To enable goals to make a difference, organizational units at all levels of recursion should have the capacity to select local goals (strategic regulation), to design local structures (regulation by design), and to locally regulate the performance of tasks (operational regulation). Levels of specialization also should be low. For instance, to enable the development of jobs-related skills or job-related moral virtue and practical wisdom, jobs should be designed to allow for deliberations and judgments involving multiple relevant job-related aspects.

As it appears, organizations designed according to de Sitter's controllability principle, resulting in low parameter-value structures, not only distribute complexity enabling organizational viability, they also offer opportunities for reducing the disquieting features attached to the disciplines, contributing to individual and societal viability.

Unlike their high parameter-value counterparts reducing opportunities, low parameter-value structures produce opportunities for goals to make a difference throughout organizations. They allow for local responsibility and involvement with performance and function related tasks. Given their simple structures and complex jobs, they allow for the development of job-related skills, moral virtue, and practical wisdom, invoking other motivators than fear and requiring other kinds of learning than operand conditioning.

However, regarding even these organizational structures a note of caution should be made. They only offer opportunities. They cannot guarantee success. For instance, a well-designed team may be the locus of individual and team learning. Yet, it also can also become a snake pit, a place of continuous reciprocal monitoring and control: a panopticon par excellence. Asking absolute guarantees of structural design, is asking too much for at least two reasons. It is asking too much from a sound perspective on cybernetics, for asking for guarantees presupposes that virtuous behavior of organizational members can or should be determined or trivialized by structural design. This, however, is a contradiction in terms, for what will remain of virtue if it is determined by outside force? It is also asking too much from the perspective of a sound evaluation of what organizational structures can and cannot do.

11.4 Organizations and Rich Meaningful Survival

Based on the argument in this chapter, we can conclude that poor survival is not inevitable. On the contrary, we discussed a concept of organizational survival that is more "rich" than "being able to maintain a separate existence." Moreover, we can conclude that Beer and de Sitter's functional and specific design principles support the design of structures supporting rich survival. Designs based on these principles allow societal and organizational goals to make a difference in organizations. At every level of recursion chances are created for organizational members to be involved in and feel responsible for their contribution to the realization of these goals. Moreover, by means of structural designs according to these principles, organizational members get the chance to develop the job-related skills, the moral virtue, and the practical wisdom to deal with this responsibility. For several reasons, we think these conclusions may be useful.

First, they attenuate the seventh worrying feature that followed from our cybernetic and social systemic analyses of the disciplines. As may be recalled, this seventh feature referred to the cybernetic and social systemic underpinnings of the disciplines (including their disquieting features). More in particular, it highlighted that all instances of the disciplines have in common that they (1) realize only poor survival and (2) rely on basic cybernetic functions and social systemic elements and structures. Therefore, the question presented itself whether it is possible to design organizational infrastructures based on cybernetic principles that do not have the disquieting features of the disciplines and even may support rich survival. Given the argument in this chapter, we can say that this is indeed possible. The application of cybernetic principles to the design of organizations does not necessarily lead to organizational structures supporting poor survival. These principles are *pharmakon*, which means that it is possible to use them appropriately or to misuse them (either because of ignorance or intentionally). Therefore, designing organizations is not only a matter of skill, involving scientific knowledge of basic cybernetic principles and experience allowing for their application in concrete situations. It also is a matter of moral virtue and practical wisdom. Designers have the obligation to develop their skill, reducing the possible misuse of principles based on ignorance. However, primarily they have the obligation to see their design not only in terms of the isolating mode of making, but also in the inclusive mode of acting, for they can choose to build structures supporting poor survival, but they can also choose to build structures supporting rich survival.

The conclusions formulated above may be useful for a second reason. Within business ethics, broadly speaking, two different approaches can be distinguished: the compliance and the integrity approach (e.g., Trevino et al. 1999; Trevino and Nelson 1999 or Sharp-Paine 1994, 1997). The compliance approach focuses on organizational compliance with mostly legal rules. It advocates the use of codes of conduct and the monitoring of organizational decisions on sensitive issues. A basic assumption of this approach is the possibly opportunistic behavior of organizational members. The integrity approach focuses on responsible behavior. It argues that rules need to be applied, and that their application requires virtue. It views organizational members as human beings that can develop the required virtues.

If we compare the two approaches to the one developed in this chapter, two advantages of our approach can be discerned. To begin with, taken in isolation, the compliance and the integrity approach only seem to tell half the story. The focus of the compliance approach on rules seems to forget that, in concrete situations, the "blind" application of rules may result in injustice. Rules need to be applied, and their application requires virtue: compliance needs integrity. The focus of the integrity approach on responsibility seems to forget that responsible behavior and the virtues associated with it can only develop in a society or in organizations with (basically) just rules: integrity presupposes rules and compliance with them. Compliance and integrity seem to be interdependent. The approach advocated in this chapter takes this interdependency seriously. By focusing on organizations incorporating themselves into society by incorporating society into themselves, it takes into account the importance of societal and organizational rules. However, as argued, incorporation should not be "mechanical" or "isolating." What counts is integrity of and beyond incorporation. In this way, it takes into account the importance of integrity. Therefore, a first advantage may be that the approach presented in this chapter comprises the compliance and integrity approaches, overcoming their possible one-sidedness. A second advantage of the approach outlined in this chapter appears if we ask *how*, i.e., by what means, compliance or integrity should be supported in organizations. In literature on these approaches, a lot of attention is paid to what they entail, why they are important, what their view on humanity is, and what features organizational members should have to realize either of them. However, we were unable to find a more or less systematic derivation of principles supporting the design of infrastructures allowing for the realization of either compliance or integrity. The approach developed above, not only specifies what rich survival entails, it also formulates specific design principles for organizational structures supporting rich survival. In this way, it not only points into the direction of a state of affairs that should be realized. It also provides practical guidelines that can help realize it.

There is a third way in which the developed argument in this chapter can be helpful. It may be argued that rich survival is a *utopia* and nice for idealists, but that it has little to do with the hard reality of "doing business." The same may be said for the design principles supporting it. Against this "argument," we can say that the principles supporting rich meaningful survival also support survival. As argued in Chaps. 6 and 7, Beer's functional design principles and de Sitter's specific design principles, increase an organization's controllability, and thereby, the potential for realizing and adapting goals. They allow for order flexibility, control over realization, and innovation. They increase involvement, create opportunities for learning, and reduce structure-related stress. And, they increase the quality of communication in organizations. If they are anything, these features are all requirements to successfully participate in the hard reality of "doing business." Therefore, the argument in this chapter and its relation to that of the previous part of this book, may be useful, because it shows that there need not be a contradiction between high organizational performance and a "rich" contribution to society, for both have the same design principles.

References

Beer, S. (1995). *The Heart of Enterprise*. Chichester: Wiley.

Luhmann, N. (1998). *Die Gesellschaft der Gesellschaft. Zweiter Teilband*. Frankfurt am Main: Suhrkamp.

Teubner, G. (1985). Corporate fiduciary duties and their beneficiaries: A functional approach to the legal institutionalization of corporate responsibility. In K. J. Hopt & G. Teubner (Eds.), *Corporate Governance and Directors' Liability: Legal, Economic and Sociological Analyses of Corporate Social Responsibility* (pp. 149–177). Berlin/New York: De Gruyter.

Sharp-Paine, L. (1994). Managing organizational integrity. *Harvard Business Review* (March–April), 106–117.

Sharp-Paine, L. (1997). *Cases in leadership: Ethics and organizational integrity: A strategic perspective*. Chicago: Irwin.

Sitter, L.U. de (1994). *Synergetisch Produceren*. Assen: Van Gorcum.

Trevino, L., & Nelson, K. (1999). *Managing business ethics*. Chichester: Wiley.

Trevino, L., Weaver, G. R., Gibson, D. G., & Toffler, B. L. (1999). Managing ethics and legal compliance. *California Management Review, 41*(2), 131–151.

Chapter 12
Epilogue

12.1 Introduction

In this epilogue, we want to provide a brief summary of the book (Sect. 2.2), make a few closing remarks about poor and rich survival (Sect. 2.3), and propose our agenda for future research (Sect. 2.4).

12.2 Organizations, "Archai," Design Principles, and "Rich" Survival

Looking back on the path we have followed, it can be said that in Part I of the book, we explored two features of organizations: their experimental and social systemic feature. We maintained that, whatever else organizations may be, they are a particular type of social systems conducting experiments. These two features cannot be denied of organizations, without denying them altogether. They are "archai" of organizations.

To theoretically describe the experimental "arche," we introduced first- and second-order cybernetics. Based on Ashby, we argued that experiments in organizations entail a continuous process of selecting organizational goals, infrastructural arrangements, operational regulation, and transformations. These experiments are risky because there is no absolute certainty that selections regarding these "objects" contribute to what is at stake in organizational experiments: their meaningful survival. Based on von Foerster's ideas, we argued that the risky character of these selections is fundamental; it cannot be "organized away," for they are ultimately bound to our contingent pre-cognitions developing in a world of non-trivial machines.

To flesh out the social systemic "arche," we introduced Luhmann's theory. In his view, organizations are social systems consisting of decisions "producing" decisions. By "production" Luhmann understands that, given particular decision

J. Achterbergh and D. Vriens, *Organizations*,
DOI: 10.1007/978-3-642-00110-9_12, © Springer-Verlag Berlin Heidelberg 2010

premises, operations are generated that are recognized as decisions and connected to prior decisions. Based on this production process, the organization as a system of related decisions emerges that distinguishes itself from its environment.

The experimental and social "arche" of organizations are related. Goals, infrastructures, operational regulation, and transformations, are "objects" of decisions and decision premises. They are what decisions and decision premises in organizations are about. Moreover, because decisions are communications communicating selections *as* selections, they highlight the contingency and risk involved in them, making organizations into a type of social systems that explicitly reflect and deal with contingency, uncertainty, and risk.

Because the experimental character of organizations cannot be "organized away," organizations should be directed at supporting, instead of suppressing, experiment, i.e., at supporting the selection and reselection of goals, infrastructures, operational regulation, and transformations. And because an organization's potential for selection and reselection is conditioned by the design of its infrastructure, i.e., the design of the organization's structure, HR-systems, and technology, Part II of the book focused on finding design principles for infrastructures supporting organizational experiments.

In Part II, we discussed two types of design principles: functional and specific design principles. Functional design principles specify what a system must be able to do, if it is to continue its experiments with meaningful survival.

Based on Stafford Beer's Viable System Model, we argued that in order to be able to survive, an organization should – at every level of recursion – define its identity in terms of primary activities, and design relevant attenuators and amplifiers. Moreover, at every level of recursion it should realize two overlapping sets of functions. The first set supports the *realization* of the organization's identity. This set comprises primary activities and the coordination and control function. The second set supports the *adaptation* of the organization's primary activities. It comprises the control, intelligence, and policy function. According to Beer, these five functions (and their relations) are necessary and sufficient for (organizational) viability.

Beer's model specifies functional principles that may be used to design organizational infrastructures. However, it does not provide a prescriptive model explicitly relating infrastructural arrangements to the realization of these functions. To this purpose, specific design principles are required.

De Sitter formulated this type of principles for the design of organizational structures. More in particular, based on Ashby's theory, he argues that these structures should be designed in such a way that potential disturbances are attenuated as much as possible and regulatory potentials are amplified as much as needed. To this purpose, the design parameters describing the production and control structure should have low values.

Based on Beer's functional principles and de Sitter's specific principles, it is possible to design organizational structures supporting social experiments with meaningful survival.

In Part III of the book two modalities of meaningful survival were at issue: poor and rich survival. Foucault's description of the disciplines provided an extreme example of poor survival. Attached to these disciplines, we found six disquieting features. Roughly speaking, the disciplines consider organizational goals as contingent and minimal and aim at trivializing the basically non-trivial behavior of organizational members by means of a process quite similar to operand conditioning, using fear as a prime motivational factor.

To find a counter-model to the disquieting features of the disciplines, we introduced Aristotle's ideas about the fulfilled human life. This entails realizing our characteristic human capacities in the best possible way and acting according to them. However, Aristotle's theory of the fulfilled life only was a first step towards finding organizational structures that can support rich survival. In a second step, we explained what it means for an organization to contribute to a society enabling it to let its citizens live a fulfilled life. More in particular, we argued that to this purpose, organizations should incorporate themselves into society by incorporating into themselves relevant societal values and programs. This process of incorporation should not be "isolating" or "technical," but "inclusive," requiring integrity of and beyond incorporation.

In the previous chapter, we argued that this requires that (1) the organization's primary contribution to society actually makes a difference in the organization and (2) organizational members develop job-related skills, practical wisdom, and moral virtues. From these two requirements, we were able to derive six specific design principles for "rich" production and control structures. It appeared that these structures are the same as those that can be deduced from Ashby's theory and Beer and de Sitter's application of it to respectively viable systems and organizations. Organizational structures supporting rich survival, therefore, are indeed possible, and they are in line with basic insights provided by cybernetics. Apparently, there exists a, yet to be explored, connection between cybernetics and Aristotle's reasoning concerning the individual and societal meaning of and conditions for "*eudaimonia*." Perhaps this connection has to do with the point that both traditions are involved in the functional analysis and improvement of complex adaptive systems in their environment.

12.3 Rich Survival: Its Applicability to Organizations and Relation to Poor Survival

Reading Chaps. 10 and 11, the question may arise whether all that is said is "realistic." It may be fine to speak about rich survival, about organizations contributing to society, about integrity of and beyond incorporation, or about developing the integrity of organizational members. However, it may also be argued that organizations have to compete for capital, resources, and customers on global markets, forcing them to operate in "isolating" rather than "inclusive" ways. Moreover, organizational members have their own private and occupational interests and may be involved in lines of work and micro politics that do not promote the best in them. All this seems to be at odds with rich survival, making it into something that always should, but never actually is realized; making it into a "parasite of the future" (Cioran 2008).

In this epilogue, we would like to direct a few words against this and similar arguments.

Indeed, rich survival is what can be called a "parasite of the future," or to phrase it in a more positive way, a "counter-factual." Here, we understand a counterfactual as something that *should be* realized, irrespective of its actual realization. Irrespective of the fact that no, some, or all organizations provide a rich contribution to society, "rich survival," in our view, remains something to strive for. So, we are ready to admit that rich survival is counter-factual. However, this alone cannot be held against it. It is common to all basic values.

Of course, this does not mean that no objections can be raised against rich survival. Both its value and attainability for organizations operating in modern society may be contested. Without its value there is no reason to try to attain it. And without attainability, there is little point in presenting it as a value.

Arguing in favor of the value of rich survival, we would like to point out that at its heart is the concept of living a fulfilled life. As human beings, we literally cannot do more or better than to develop our humanity to the best of our ability and act accordingly. As we cannot and never actually do this in isolation, we may ask of society and its institutions, including organizations, that they are arranged to allow for this kind of development. Whatever the historically contingent lay-out of society and its institutions, we consider it of value that they are directed at providing conditions for individuals to cultivate their humanity.

The value of rich survival may also be presented by departing from its alternative: poor survival. We already argued that poor survival in the form of the disciplines is disquieting, and the whole project of developing an alternative to them was inspired by these disquieting features. This type of poor survival certainly is not valued as a goal to be pursued.

Now, it can be argued that there are other instances of poor survival than the disciplines and that some of them are more "enlightened," and therefore less worrisome. Maybe these more enlightened forms of poor survival can be of value.

For instance, suppose that an organization introduces a stakeholder approach that successfully takes into account relevant societal stakes for the purpose of goal-setting. However, the reason for introducing this approach is not because it is the right thing to do, but because it is good for the "image" of the organization. Although all kinds of positive developments may follow from this (employees, the environment, "important" client groups may benefit from it), this organization still suffers from the contingency of its goals. The stakes and the organizational goals flowing from them are not considered relevant because of their societal value. They are considered relevant because it is important for the organization's image to (be seen to) pursue them. As soon as this relation between stake and organizational image disappears, the "value" of the stake also disappears, irrespective of its actual societal value. So, even in this example of enlightened poor survival, goals remain contingent. This is disquieting because organizations like this can and will shed important societal stakes like snakes shed their skin.

In general, if organizations can pursue any goal, for any reason, by whatever means, which may be the essence of even enlightened forms of poor survival, goals remain contingent. And contingency of goals was one of the basic worrying features of poor survival, stimulating the development of an alternative to it. The fact that features attached to even enlightened forms poor survival can be or are experienced as worrying, point into the direction of the worth of underlying values that may be articulated in terms of rich survival and the fulfilled life.

Those who are prepared to grant the value of rich survival may still question its attainability. They may ask whether it is realistic or even possible to strive for rich survival in a modern global market.

Concerning attainability we want to address two issues. The first is that of the "organization of the organization": can organizational infrastructures be designed enabling rich survival? The second concerns the organization's relation to its societal environment. Is rich survival a viable option in a global market?

If we zoom in on the "organization of the organization," it can be argued that just as "habitual dispositions" are needed for virtuous human behavior, "virtuous infrastructures" (decision premises) are needed for decisions supporting rich survival. Can such infrastructures be designed?

In the previous chapter, we showed that infrastructural conditions can be created that support it.

As argued, organizational structures supporting rich survival are possible and may be realized in organizations. And what is more, such structures not only support rich survival. They also support organizational performance in terms of viability (Beer) and the quality of organization, work, and working relations (de Sitter). Their implementation not only can benefit society in terms of rich survival, it also can benefit organizations in terms of competitive advantage.

Moreover, we think that it is possible to develop specific design principles for the other two constituents of the organization's infrastructure as well. Building upon and in line with organizational structures as described in Chaps. 7 and 11, systems for human resources management and production and information technologies may be designed to allow for the development of the practical wisdom, moral virtue, and job-related skills needed to make rich organizational contributions to society. So, we think it is possible to design organizational infrastructures allowing for rich survival.

The second issue is that of the organization's relation to its societal environment. More in particular, it might be argued that organizations pursuing rich survival do not stand a chance on markets dominated by profit and shareholder value. Therefore, rich survival may be a nice idea, but organizations that need to survive in the "real" world can never attain it.

With respect to this argument, we would like to point out two things.

First, in Chap. 7 on organizational structures we argued in detail how production and control structures designed according to the controllability principle enable organizations to realize the functional requirements (quality of organization, quality of work, and quality of working relations) needed to survive in a fundamentally risky world. Particularly, realizing quality of organization, involving order flexibility, control over realization, and potential for innovation provides excellent conditions for survival, even in a global society. In Chap. 11, we showed how organizational structures designed according to the same principle also enable rich survival. Therefore, from an infrastructural point of view, there need not be a contradiction between being able to compete on global markets and rich survival. The structures enabling the one also enable the other.

Second, making a profit or realizing shareholder value need not necessarily be at odds with rich survival. In Chap. 11, we argued that organizations can contribute to different societal subsystems each devoted to handling a particular societal problem. One of these subsystems is the economic subsystem which has the function of ensuring the satisfaction of future social needs. Making profit or realizing shareholder value, may surely contribute to the realization of this function, and realizing this function may surely contribute to a society enabling its citizens to live a fulfilled life. So, there need not be a fundamental opposition between rich survival and making profit or realizing shareholder value. However, for an organization pursuing rich survival, the latter two cannot be decisive and of value *in themselves*.

12.4 Things to do

The present book is about the experimental and social arche of organizations, about principles for the design of its infrastructure, more in particular, organizational structures, and about organizational structures supporting rich survival. Given these topics, there are thee more or less natural lines for further investigation.

First, if organizations are to contribute to a society enabling its citizens to live a fulfilled life, it seems plausible that such a society needs to meet particular functional requirements and specific principles for its design can be derived. The first line of enquiry addresses these requirements and specific design principles. It consists of the application of cybernetics and social systems theory to society and aims for the development of principles for the design of societal subsystems and their institutions that can support a society enabling its members to live a fulfilled life.

Second, in the book we formulated principles for the design of organizational structures. These structures are only one aspect of the organizational infrastructure. The question remains whether specific design principles for human resources management systems and production and information technology can be formulated that are in line with the principles formulated for organizational structures. Formulating these specific design principles is the second line of enquiry.

Third, by specifying design principles for organizational infrastructures that allow for rich survival we did not yet address the problem of the transition from "poor" to "rich" infrastructures. As may be expected, such transitions may be highly problematic. The reason for this is that organizations with poor infrastructures lack the variety needed to redesign their own infrastructure. They have become "self-locking" organizations. In such cases, episodic interventions are required, involving temporary change-organizations within the larger organization. The third line of inquiry concerns the application of cybernetic and social systemic insights to the design of such episodic interventions in organizational infrastructures.

References

Cioran E (2008) Die verfehlte Schöpfung. Suhrkamp, Frankfurt am Main