DMP

A kinetics of macroscopic particles in open heterogeneous systems

by

Per-Erik E. Bergner

A naive application of the notion of stochastic process

(DMP = Dynamics of Markovian Particles)

DMP

4rd Internet edition

Linnégatan 20, SE-114 47 Stockholm, Sweden

Preface

A purpose of this monograph is to bring together some pieces of elementary knowledge and experience that are reasonably well settled though often unsystematic and scattered in the literature. To this comes that some of the matters are pretty old and almost forgotten though still being of significance.

There are of course insufficiencies in the presentation, especially as reflected by the appendices: THEY SHOULD BE LOOKED UPON AS AN INTEGRATED PART OF THE MAIN TEXT (their presence expresses my inability to organize the material in the usual way). These appendices are thus not mere technicalities aimed for the specialist. So I humbly urge the reader to consult them whenever they are referred to.

That this monograph is now published is much because of my son Sten-Erik E. Bergner, who has not only actively but also critically (very much so) helped me to put the material together. In fact, much of this book should have been presented long ago, but because of a few incidences this did not happen. To some extent this shows up in the list of references because it is old indeed, but I have not bothered to update it, simply because: 1) what was formally true years ago is still true; 2) the references are marginal for the content of this monograph; 3) they might after all have some historic significance. This is however of minor importance and what really matters is that, over the years, colleagues of mine have provided me with plenty of both criticism and encouragement, without which this little book would hardly ever have occurred.

Stockholm 2005
Per-Erik E. Bergner

PS. Theory is here looked upon as an instrument of the experimenter and, hence, it must appeal directly to the investigator's intuition. A theory should therefore appear somewhat self-evident, and *the physical significance of analysis must have priority*. But then we must keep in mind that often the languages we formulate and apply today will in a future appear insufficient or simply incorrect. Or more bluntly, presently there are "scientific" concepts of the past that we consider obsolete, and there is no reason why the future fate of the notions we create today should be much different (i.e. our concepts might appear obsolete to future investigators).

Table of Contents

Introductory footnote

The function of this book is not so much a formulation of some theory in the form of specific models (e.g. for prediction) but, rather, the function is construction of a language in terms of which models and hypotheses can be formulated. Differently phrased, the purpose of the book is to construct a language for UNCOMMITTED DISCUSSION OF COMPLEX (OR, HETEROGENEOUS) SYSTEMS. This language is in the tongue of motion of macroscopic particles: one deals with a population of particles in motion, no matter if the motion is concrete or takes place in an abstract state space or both. And that the discussion is "uncommitted" means that the discussion—of the system and of the processes within it—can be carried out without prior assumptions about the underlying physics: the system is allowed to contain an immense variety of physico-chemical processes (with or without mixing). An arbitrary complexity!

One way to observe this purpose is to note questions like: What does the word "substance" mean? What does the word "system" mean? Routinely such questions have been considered as being a matter of philosophy and then neglected by so-called active scientists. Personally I react against this: TO BE CONCERNED WITH THE WORDS WE USE AND THE CONCEPTS THEY REPRESENT must be an active matter for any active scientist.

Chapter 0
Heuristic Starter

Thermodynamics was originally called *mechanical heat theory*, because rather early (before the 19th century) it was empirically accepted that there is a definite connection between mechanical work and heat liberated (e.g. in 1797 Rumford showed that drilling a canon with a dull drill could make water boil), and those findings formed the basis of a "mechanical heat theory". It was however recognized – especially at the development of the steam engine – that much of this theory's content is quite general, and so the notion of thermodynamics was born.

The subject I shall deal with in this book has a somewhat analogous metamorphosis, though more recent and less prolonged. Thus, around the middle of last century an experimental technique was developed that gave rise to what has become known as "tracer theory". However, this theory is afflicted with such a number of insufficiencies that most of it (the theory) should be looked upon as a parenthesis of history. But the topic nevertheless induces a few problems that provide a basis for rather general a perception (Section 0.1); somewhat similar to how mechanical heat theory constitutes an initial basis for thermodynamics. However, similar to how the notion of steam engine is of small significance for today's thermodynamics, the notion of tracer is of little interest to the here subsequent theory.

0.1 Two questions

Let us proceed heuristically and consider an open system S, which might be visualized as a closed domain in the Euclidian space that is functionally open in the sense that there are particles of some specific kind moving in and out of S. We further assume that there is a constant stream of "equal" particles into S, and that each particle that enters S eventually also leaves S irreversibly. Possible examples would be a rabbit and atoms in food molecules, a lake and pollutants or a city and cars. It is intuitively reasonable to expect that under certain conditions steady state will eventually be reached. Here steady state means that the number of particles in *S* is constant (input equals output) and that, moreover, the distribution of particles within *S* is stationary (time invariant).

QUESTION (i): Are there any necessary and sufficient conditions for steady state, and if so, what are they?

This question means that we accept the notion of steady state as an existing physical entity, which we meet in nature to an extent that it is of significance for our daily life.

Yes, one might wonder about the practical relevance of this: to expect the existence of steady state here and there in nature does not seem very realistic in the light of all ongoing variations and perturbations. Perhaps the best way to meet this skepticism is to refer to classic thermodynamics, where the notion of equilibrium plays a central role: the theory has indeed proved practically quite useful in spite of the fact that equilibrium in its exact sense does not occur in nature. The most commonly applied explanation of the success of the theory is that many systems in nature are "sufficiently" close to equilibrium for that notion to be practically useful as a general reference. And, hence, a similar argument can be applied to the notion of steady state, which thus plays a role of general reference for open system. Although "reference" is something abstract we nevertheless visualize it as something physical that we frequently experience.

Now, question (i) has a somewhat delicate background, because the particles in *S* are in constant motion and the steady state we presently contemplate is a result of an, often large, family of particle motions. To visualize the meaning of this let us consider the simple *S* depicted by Fig. 0.1.1, where M is a membrane permeable to water and to the particles in question: the dots are supposed to illustrate the steady state population of particles.

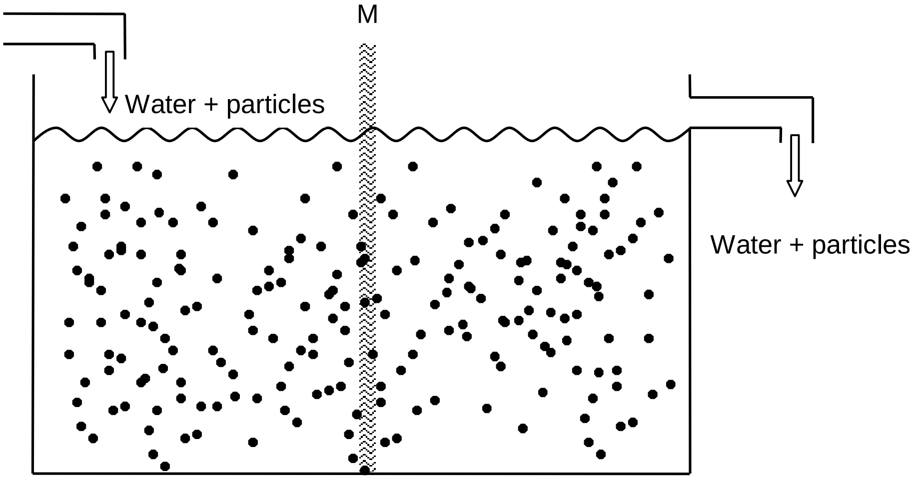

Fig. 0.1.1 An illustration of the notion of steady state: there is a constant input and equal output of water and particles (the system is "open"), and the particle distribution is constant in time; the membrane M is permeable to both water and particles. The two containers are not (!) necessarily well-stirred.

Fig. 0.1.2 depicts an imagined trajectory generated by a single particle passing through *S*. We may then try to visualize how a large population of such trajectories—of various but finite length—build up a time independent population that we refer to as a steady-state population.

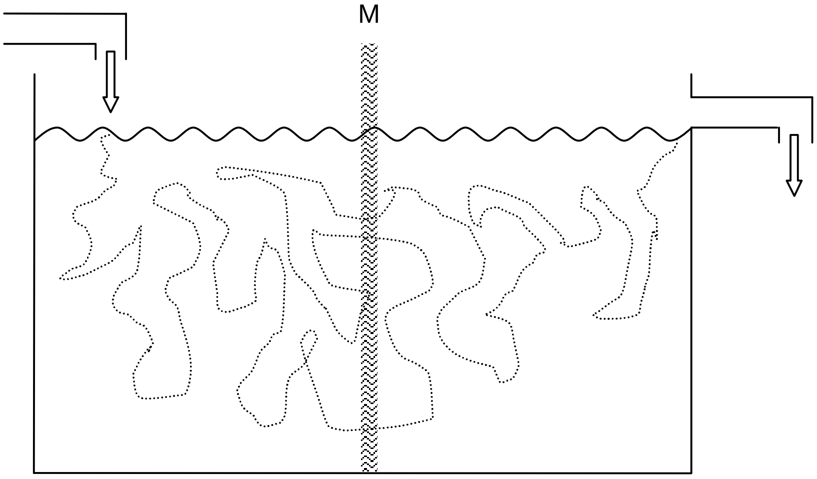

Fig. 0.1.2 The steady state in Fig. 0.1.1 is a result of the motion of individual particles passing through. This figure is supposed to show a thought trajectory of such an individual particle. The basic problem dealt with in this book concerns the relation between this kind of motion (a transient time process) and the resulting steady state. The "classic" system was closed and one particle's trajectory could be followed for as long as it pleased the investigator. The main target in this book is however open systems (all particles eventually leave for good), and the trajectory cannot be made arbitrarily long: instead one has to consider large families of trajectories (a large population of particles that is).

The situation suggested by the two figures is in no way just academic. The fact is that we can often study the time processes as well as the steady-state systems and, therefore, questions about possible connections between the two kinds of systems enter one's mind naturally. Moreover, if the time invariant steady-state, which we experience as something solid, actually is the result of a large number of transient time processes of the kind depicted by Fig. 0.1.2, we should expect that this probably affects the properties of the systems we observe. But what properties could we expect?

However, before we can at all entertain that kind of question there is of course the other and more general question that must be dealt with:

QUESTION (ii) Is there any relation between steady state and trajectories of the kind depicted by Fig. 0.1.2 and, if so, what is it?

Yes, as just discussed above, it seems natural to expect the existence of a connection, but the simple fact is that we don't know. The main task is therefore to develop means for the handling of question (ii); hopefully we shall then also obtain an answer.

The question of how to connect a time process with something time-independent is in no way new. Thus, in classic statistical mechanics a central problem is how to relate time averages (resulting from time processes) to space averages (resulting from time-invariant space distributions). Commonly this is known as the **ergodic problem**, which is relevant for what I here shall refer to as closed systems (the ergodic problem is extensively discussed in Appendix 3). In closed systems the particles stay in S for good, and a trajectory of the kind depicted by Fig. 0.1.2 can be made arbitrarily long, and this constitutes the fundamental difference between the classic case and the problem we face here: S is open and the trajectory is finite and cannot be made arbitrarily long; this means that, as I shall subsequently show, instead of following a single trajectory that can be made "sufficiently" long, we will have to deal with a population of (finite) trajectories that can be made "sufficiently" large.

This is an undertaking one accepts because, again, one should expect that the type of connection we presently consider will have effects on the properties of the systems one meets. And from a plainly practical point of view one wants to know what these properties are and how to deal with them. It is in fact an elementary and natural matter but, as almost always happens, as soon as one tries to deal with it in a somewhat formal manner it begins to look involved and academic (a deception that one seemingly can never avoid).

0.2 How to do what's needed

In the subsequent chapters I shall apply a scheme that is outlined in this section.

STRUCTURE
Fig. 0.2.1 depicts the point of view I apply on the notion of theory. Thus, one could say that *one of the purposes of this book is to construct a current theory for the handling of the two questions formulated in the preceding section*. And the main task is then to construct an appropriate language (the term "language" is soon specified below).

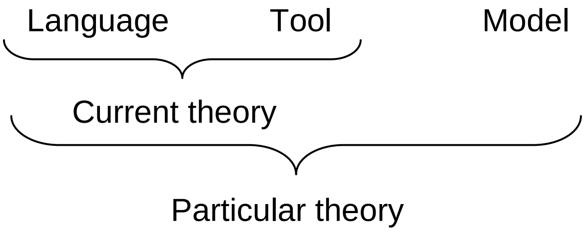

Fig. 0.2.1 The daily notion of theory is here named ***particular theory****: it is intended for particular application. It contains two components:* ***current theory*** *and* ***model****. As explained in the text below, "language" is the component of current theory that serves as material for building of theoretical model by means of "tool" (often some form of mathematics)..*

Yes, the inceptive task of this book is to formulate a language for the two questions in the preceding section. However, as so often happens in this kind of work, the very construction of such a language also generates by-products that turn the final result into something more than originally intended. Or, phrased differently, we shall end up with a language that has broader significance than the two questions in the preceding section might suggest.

Here the term *language* is used in a somewhat particular sense, namely that it represents the "material" theoretical models are made of. The structure of this particular notion is depicted by Fig. 0.2.2, which in words could be expressed as: **language** is the union of the four sets

(i) a set of **primary concepts**, which are not defined but understood intuitively;
(ii) a set of **secondary concepts** that are defined in terms of primary concepts or of other secondary concepts already defined;
(iii) a set of **postulates**, which are at will formulated "statements about concepts" (primary or secondary) and that cannot be logically verified within the language;
(iv) a set of **theorems**, which are *formally deduced* statements about concepts (often in the form of equations of a sort).

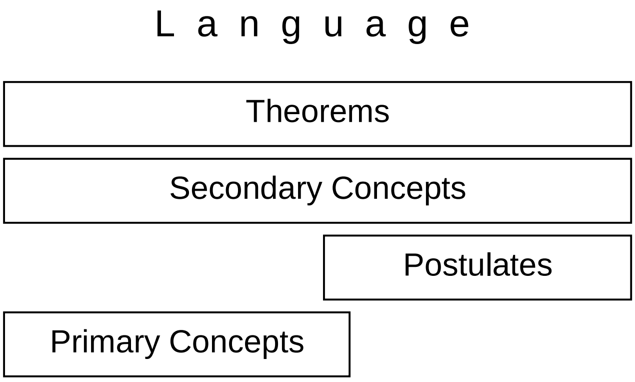

Fig. 0.2.2 One could say that the purpose of the book orbits around the very notion of language. Initially it consists of primary concepts and postulates and from there secondary concepts can be defined and theorems derived; primary concepts are not defined but understood intuitively, and postulates are statements that cannot be empirically tested within the language itself (in fact, language is the material models are made of, and as such it can be more or less convenient but never true or false). Language is not static but grows, and especially the number of theorems has a tendency to increase. Frequently this increase is a result of specific applications that easily induce increase in the number of secondary concepts. One usually tries to keep the basis "primary concepts and postulates" as small and invariant as possible.

A possible illustration – of the present notion of language – is provided by classic mechanics. For the elementary part of that theory a thinkable list of primary concepts could be: particle, mass of particle, force acting on particle, time, and particle's location as a function of time. The familiar Newton equation of motion (with time as the independent variable)

$$(\text{mass})\ (\text{acceleration}) = (\text{force})$$

then comes out as a postulate; note that acceleration is a secondary concept because it is *defined* as the second time-derivative of the particle's location. There are presentations of this language where this postulate functions as a definition of "force", which then comes out as a secondary concept; of course, there are a variety of forms of this language, for instance forms where the notion of force does not appear. In other words, application of the structure presently ascribed to the notion "theory" is to an extent a subjective topic. Note that no theorem is introduced in this example nor is any model formulated.

A sharp distinction between language and model is an absolute necessity, and an understanding of the elusive notion of model (or, rather, a feel for it) is a necessity indeed, should the experimental and theoretical work be at all positive.

MODEL AND MATERIAL

Assume I have a piece of wood and a knife good for whittling. Assume further that I have an *idea* about the notion of "boat". It is then possible for me to carve out of the wood something that makes an onlooker to exclaim: Oh dear, it resemblance a boat! What has happened is that I have in some MATERIAL and with the help of some TOOL realized an idea of mine, and I name the result a MODEL. Whether it is a good or bad model is beside the point but, in this example, because the material is wood one might speak of a wooden model. In other contexts one might have to deal with a steel model or a paper model but, no matter what, this kind of model I like to call a **physical model**. Another example of a physical model is the molecular models chemists love to make out of differently colored plastic balls (hands, with fingers, are then the usual tool).

In this book the term "model" stands for **theoretical model**, because the material is some language of the kind discussed above: the term material is replaced by the term language in Fig. 0.2.1; frequently a theoretical model is in the form of an equation of a sort; for instance, for description of the growth of a biologic cell. And instead of something like a knife the tool is some form of mathematics (at least often so). It is, and this is made particularly evident in the Section 4.1, crucial that there always is made a clear distinction between model and language.

The notions material, tool, and model seem to be such that almost everyone is empirically familiar with them. Hence, at least to me this means that the particular scheme that I have here outlined about "theoretical model" (Fig. 0.2.1) should be nothing but a mild extrapolation of something familiar (the structure I propose is in the line of common praxis, so to speak).

THEOREM AND TEST OF MODEL

To illustrate the preceding, let us once more consider Newton's famous equation: mass times acceleration equals force. Within the considered language (classic mechanics) this equation cannot be tested and, hence, it is not a model: in the present terminology it is a postulate. As it stands it is a second order differential equation that is so unspecified that it cannot be solved. However, should we insert in the force term the law of gravitation the equation can be solved, and this solution can be tested against data. That is, we can test the model (the law of gravitation), where we apply a theorem namely the *procedure* (a piece of mathematics = tool) of solving the equation. And here is a hint of why it is crucial to make a clear distinction between language and model: it is the latter and not the former that is experimentally tested against data (the only thing that matters for a language is its convenience).

0.3 Purposes: What should be done?

We face a problem that commonly is referred to as “the system’s complexity” which, in spite of the sloppy terminology, submits a kind of flavor to the particular difficulty at hand. For instance, consider a sodium ion that is in a rabbit’s hart and some time-interval later appears in the rabbit’s left lung. To describe such a process in terms of physics (classic or any other type whatever) would be quite an undertaking and involve a wealth of various assumptions; it would in fact require an amount of knowledge that hardly ever is even far away. Consequently, any effort to obtain such a description would be futile. That is, one should not be surprised to hear exclamations like “Gosh that is too complicated!”

What makes this so particular is that such a process (the ion’s motion from the hart to the lung) is in the eyes of a biologist something quite elementary – yes, a triviality. Hence, the outlined scenario demonstrates the *de facto* existence of quite some gap between what we today refer to as physics and as biology, and that gap spells complexity: what is overwhelmingly complicated to a physicist is self-evident for a biologist. And a main task is TO BRIDGE THAT GAP. And this has very much governed my choice of approach: I have selected an approach based on the notion of stochastic process.

The rational behind this choice is quite simple, namely that one can in exact terms speak of the probability for the sodium ion to move from the heart to the left lung without ever getting involved in detailed assumptions or hypotheses about the underlying physical mechanisms. In a way one could say that the laws of probability replace what we think of as physical mechanisms, and this implicates a kind of uniformity, in the sense that the mechanisms are the same no matter what point of view is applied, is it physical or is it biological.

As I hope to demonstrate in this book, this approach does generate a number of possibilities at the same time, however, as it does cause some difficulties that must be taken care of. For one thing there is the instability caused by the inherent randomness (Appendix 6). Moreover, the theory of stochastic processes has an annoying ability to soon become too complicated even at physically modest applications; therefore one has to look for concepts that are independent of those cumbersome matters. And one thing I shall do is to leave the very notion “state of particle” unspecified (it remains as a specific concept though). That nevertheless some practically useful results can be obtained I consider somewhat amazing.

Chapter 1
Description

From Section 0.2 it follows that, to discuss the questions formulated in the Section 0.1, the first step is to formulate a language so that they (the questions) can be rationally dealt with. The initial part of that step is the construction of primary concepts (of course, discussion of primary concepts always involves introduction of secondary concepts), and this is the purpose of this chapter; the formulation of postulates is postponed till Chapter 2, where the dynamics (the *motion* of particles) is the main target. As will be seen in the present chapter, the mere formulation of primary (and secondary) concepts is a somewhat involved matter, very much because of the "complexity" of the phenomena we here deal with (as discussed in the preceding section).

1.1 Fundamental concepts

The first task is thus to formulate an appropriate language. And the first thing then to do is to find a proper set of primary concepts but, as so often happens, what looks fine on paper does not function without disturbances as the practical application goes. Thus, as it turns out in this case, there are some primary concepts that look more primary than the other primary concepts, so to speak. And rather arbitrarily I shall name them **fundamental concepts**. Yes, they are indeed primary but needed at the very beginning for the formulation of the other concepts.

Here is something that might look like a contradiction: primary concepts are not defined and, hence, there cannot be a need for any other concepts for the formulation of a primary concept. And this is true except that also our intuition needs something to work with. Phrased differently, what I here call fundamental concepts are not needed for any formal definition of primary concepts—such definitions are of course excluded when it comes to primary concepts. No, the fundamental concepts are needed for our intuition to function:

as always, one must start with something. Of course this is quite a subjective matter, and should not be taken all that seriously. One could look at it as merely a matter of organizing ones thinking, in the sense that also primary concepts must be introduced in some order, and the "first" of these concepts I name fundamental.

The first fundamental concept that comes into my mind is **particle**. Now, a basic feature of primary concepts is that they cannot be defined but should be understood on intuitive grounds and about the only thing that can be done is to attempt illustrations. For instance, let us assume that a small population of rats is given a certain dose of gamma irradiation. If I am interested in how this treatment affects the calcium metabolism, calcium ions (atoms) are particles. If the concern is the effect on red cell survival, red blood cells are particles. And should my interest be the effects of the gamma dose on rat survival, rats are particles.

So this particle concept is quite broad and, as will be elucidated subsequently, this is because of what here is considered as the main property of particle is not the particle's structure but its motion (dynamics). However, I shall soon introduce some constrains on the present particle concept, but yet it will remain significantly broader than the corresponding particle concepts in classic theories like Newtonian mechanics and statistical mechanics.

The fundamental feature of "particle" is thus not its structure but its motion, and that brings us directly to the other fundamental concept, namely the **state of particle**, for which I use the symbol q; and because the particle moves it becomes a time function $q(t)$. I think that one should visualize q as a vector of large but finite dimension. In classic statistical mechanics q consists of three components that describe the particle's geometric location, and three momentum coordinates that describe the particle's mechanic energy:

$$q = \begin{pmatrix} r \\ \omega \end{pmatrix} \tag{1.1.1}$$

where r is the three dimensional geometric vector and ω the three-dimensional momentum vector; that is, in this case q is six dimensional. Following the reasoning in Section 0.2 one could say that this (classic statistical mechanics) is a kind of model, because of the detailed form given to the state vector.

Personally I have found that this subdivision of q into one **geometric vector** r and one **non-geometric vector** ω often works a long way—although it formally is a model of a sort it is rather general. However, often the dimensions of the two vectors are much higher than just three. Imagine for instance a calcium ion in a rabbit. Of course the particle's geometric location in the animal is usually of importance, but to specify the location quantitatively can be quite an undertaking, and most certainly there will be a need for more than only three coordinates (for one thing, the animal moves around and its body bends in various ways). And the same thing goes for ω; for instance, the particle can be inside or out outside a cell and there is quite large a variety of cells, which are not only of various age but inside a cell the particle can be in various locations and physical states (like different energies of various forms). In short, to specify q can be quite a formidable task that easily becomes *practically impossible to handle*.

One reason for being interested in this notion of state is that, for instance, if one has two "identical" particles the only possibility to tell them apart is with respect to q. One example would be that two physical objects cannot simultaneously have the same geometric location (geometric state).

Similarly, if we observe a particle at the time t and later at the time t' ($t < t'$), we may wonder if there is any difference between the particle at the two time points: we wonder if anything of significance has happened to the particle during the time interval $t' - t$. And we simply *define* such a change as a change in the particle's state q.

That is, q represents *all* significant properties of the particle, and therefore it does not matter how the particle reached the state q. One could say that *the particle lacks memory*. Under this condition the process of how q varies with time (i.e. how the particle moves) is in applied mathematics known as a **Markov process** (cf. Takács, 1960), after the Russian mathematician Markov in the early part of the last century. And that is why I have named the theory I am now trying to outline DMP (=Dynamics of Markovian Particles).

This lack of memory on the part of the particle logically means that q contains *all necessary information* about the particle, but clearly an investigator can hardly deem what actually is "all necessary". That is, not only do we have this problem of judgment and knowledge about "all", but there is also the practical impossibility that I have just touched upon above, namely the often immense difficulty to specify q even under modest conditions (for one thing, usually one must expect that the number of components ("dimension") of q is indeed quite large). Thus:

$$\text{The state } q \text{ must be left unspecified} \qquad (1.1.2)$$

As explained in Section 0.2, any specification of q means that a model is introduced, and that is prohibited because the immediate task here is the formulation of a current theory for handling of matters like the two questions (i) and (ii) formulated in Section 0.1 and—as follows from Section 0.2—that theory must be model free.

1.2 Some primary and secondary concepts 1

Let now the symbol *u* stand for a particle (a physical object) as discussed in the preceding section. If *q* denotes the state of *u* we shall let *Q* denote the totality of states *u* can ever attend; that is, it is always true that $q \in Q$, so *Q* denotes the **universal state set**. We shall visualize *q* as a location vector in the finite (but usually poly dimensional) **state space Q**. That is, *Q* is a point set in **Q**. Because the coordinates (components) of *q* represent physical quantities they are always numerically bounded.

It seems that to many people the distinction between **Q** and *Q* is indeed rather academic. However, there is nothing subtle about it, but the distinction is quite a natural thing. The easiest way to grasp the matter is simply to look at **Q** as a coordinate system as illustrated by Fig. 1.2.1, where **Q** is two dimensional, and *Q* is a point set that not necessarily is restricted to the first quadrant.

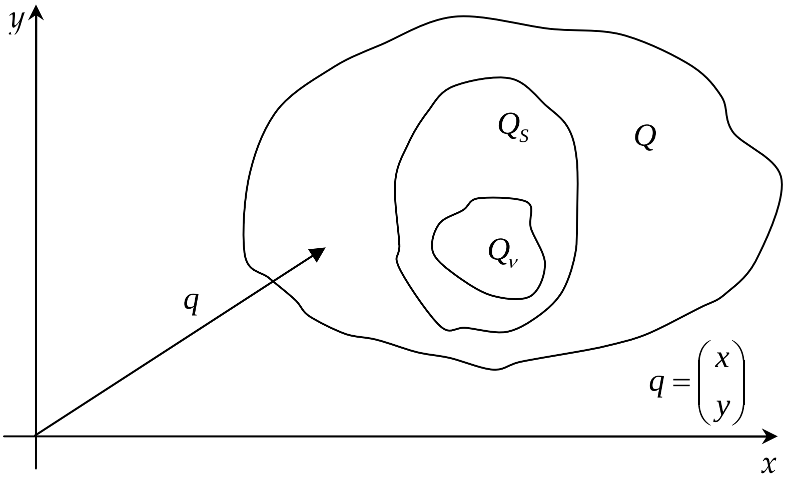

Fig. 1.2.1 A two dimensional illustration of a state space **Q***, where the state* q *is a location vector and Q is a point set that equals to the totality of states the particle can ever attend. The subset* $Q_S \subset Q$ *is the totality of states the particle can attend while being in the subject* Φ *(system* S*, next section) and* Q_v *is a similar set of states while the particle is in the subsystem* $S_v \subseteq S$ *.*

In general *u* moves which means that *q* changes with time; i.e. we are faced with a trajectory *q* = *q(t)* in **Q** (again, it is of course always true that $q(t) \in Q$). If incidentally $q(t_0) = q_0$ it is meaningful to ask for the probability that at the time $t > t_0$ the particle *u* is in a state that belongs to the subset Q_v of *Q* (that is, $Q_v \subseteq Q$), which means that we ask for the probability of the event that $q(t) \in Q_v$. In other words, we ask for the **transition**

probability (actually the conditional probability with the condition $q(t_0) = q_0$):

$$p_v(t, t_0, q_0) \equiv P\{q(t) \in Q_v | q(t_0) = q_0 \in Q\} \quad t \geq t_0 \tag{1.2.1}$$

Here are two possibilities: (1) the transition probability is there *because* of *u*'s motion or (2) the motion takes place *because* there is a transition probability acting on *u*. And it is alternative (2) that I select. This is because of the role given here to the notion of language and that transition probability enters as a primary concept (i.e. it has a physical significance of its own, and therefore it cannot exist because of some *other* physical "thing" like the particle's motion). That is, the motion of *u*, as illustrated by Fig. 1.2.2, is looked upon as being caused by a physical quantity we call transition probability (an obvious analog is the change in the motion caused by a "force" acting upon the particle in classic mechanics).

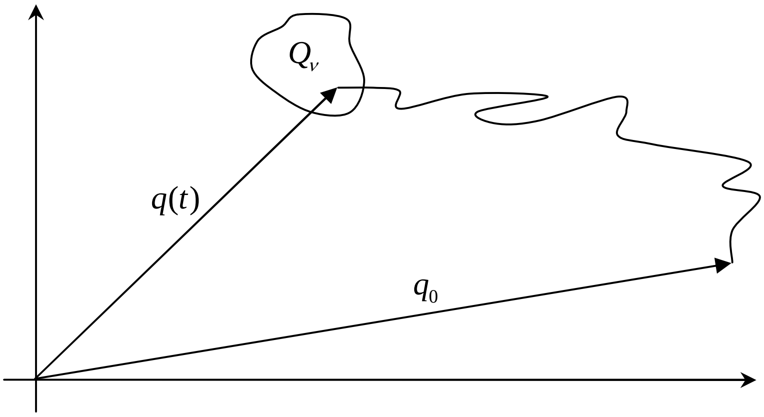

Fig. 1.2.2 If the condition is given that the particle's state is q_0 *at the time* t_0*, it is meaningful to ask for the probability that its state is somewhere in the set* Q_v *at the time* t $(t > t_0)$. *Phrased differently, we ask for the probability for the event that the particle moves from the state* q_0 *to a state in the set* Q_v *within the time* $t - t_0$. *This is a conditional probability, and "transition probability" is the name commonly given to it.*

Introducing probability as a primary concept presents a problem in that the probability theory as such could be looked upon as a "tool" (Fig. 0.2.1) – it is as if the notion of probability belongs nowhere or everywhere (is it a part of the language or...?). From this point of view it seems to me that the notion of probability should be introduced as a fundamental concept (yes, it is badly needed), but presently I leave open the question of the formal role of the notion of probability: to me the matter is, in the present context, rather esoteric.

The motion of u is thus a stochastic process with the sample function $q(t)$, which means that this time function is an individual realization of that process. One thing that must be kept in mind when dealing with time processes in general is that data usually contain less information than one would expect: often only a strikingly small number of parameters can be meaningfully estimated from kinetic data. One reason for this is that consecutive points of $q(t)$ are often not mutually independent (e.g. often there is some kind of correlation present between consecutive data points). Why I mention this is that, in the history of kinetics, attempts have frequently been made to extract more information than data allow for; an example of this is that too often attempts have been made to deduce from macroscopic data the quantitative nature of the underlying physical mechanisms - usually a totally futile undertaking.

Finally, the various concepts introduced so far make it natural to contemplate the notion of substance. Let us therefore first agree upon that a particle (a physical object) is **equivalent** to another particle if and only if its motion is a realization of the same stochastic process. So if we let *U* denote the universal population of objects equivalent to the particle *u* I shall from now on let the symbol *u* denote an arbitrary member of that population (in other words, it is always true that $u \in U$). And it does not require a lot of creative thinking to accept *U* as representing the common concept of **substance**.

Usually the notion of substance has been considered in terms of the microscopic structure of the particles $u \in U$, like atomic number or molecular structure. However, here I instead consider the dynamics of *u*; for instance, a certain kind of biologic cells could serve as equivalent particles in spite of the fact that there are no two cells that are "physically" equal; and there are occasions when *U* stands for a population of a certain kind of fish or, why not, of cars (with driver).

1.3 Some primary and secondary concepts 2

The preceding section might appear somewhat abstract so let us try to be a little more physical so to speak. And to that end I have found it fruitful to introduce the concept of **subject** with the symbol Φ, which should be understood as a natural or artificial construct like a rabbit or a chemical reactor (e.g. for industrial synthesis). That is, Φ should be looked upon not only as a geometric domain like a rabbit that contains some *U* (i.e. there is a subpopulation $U_S \subset U$ of particles in Φ, or simply the "substance *U* in the subject Φ ") but a subject is also something that in some way processes (does something to) the particles $u \in U_S$; in other words Φ is exactly what it is called—a *subject* that does something to the particles. (The substance serves as object and the symbol "*S*" in the subscript of U_S stands for "system", which is a concept that will replace Φ in the next section.)

To an onlooker it will appear as if there is a number of $u \in U$ that pass through Φ all the time, and it would be tempting to use a specific symbol like, say, U_S^0 for the steady state population of particles. But I have found that the symbol induces unnecessary problems; it seems as if the difficulty is caused by the fact that "steady state" is something *collective*, and which therefore cannot refer to the *individual* $u \in U_S$. This becomes especially evident in RULE (1.3.1):

> Necessary for a *u* being a particle with respect to a subject Φ is that it retains its identity while being in Φ and never returns once it has left Φ. (1.3.1)

This rule will be modified slightly when the concept of System *S* replaces Subject Φ in the next section.

In other words, individuality (identity) is here an essential property of particle: the primary concept "identity" is ascribed to another primary concept, namely "particle". Yes, it could be argued that (1.3.1) is a postulate; the reason why I refer to it as a "rule" is merely an old habit of mine, and a discussion of this could be justified in some other context. At any rate, (1.3.1) is the "restriction" on the notion of particle I indicated in Section 1.1.

The concept of particle is here indeed primary and, therefore, one has actually freedom to ascribe to it, at will, various properties; the necessary properties specified by (1.3.1) reflect simply my personal experiences. Please note that we here contemplate *necessary* properties: because as "particle" is a primary (even fundamental) concept it is not logically possible to formulate *sufficient* properties.

The two notions Φ and *U* make it natural to entertain a concept represented by $U_S \subset U$, namely the population of particles that are in subject Φ. But obviously a $u \in U_S$ cannot enter all states represented by *Q*, and it becomes instinctive that to given Φ and *U* there corresponds a subset $Q_S \subset Q$; this subset stands for the totality of states a *u* in Φ (a $u \in U_S$) can ever attend.

By selecting a subject Φ and a substance *U* one obviously also selects an **environment** *E*, namely the rest of total population *U* (note that there is nothing abstract about this, but something quite "physical"). Thus, to specify *E* it is not enough to specify a subject Φ but also a substance *U* must be specified as well. Phrased differently, we define the environment *E* as the part of the world where one can find $u \in U$ (particles of the substance *U*) that are not in Φ . And this means that to E there corresponds a population $U_E = U - U_S$, namely the population of those $u \in U$ that are not in Φ .

E exemplifies a truly secondary concept because it is *defined* as the physical counterpart to the secondary concept $Q_E = Q - Q_S$. Personally I look at Q and Q_S as primary concepts but I do admit that it might not be all that clear cut (for instance, Q_S could be regarded as a secondary concept defined in terms of the primary concepts Φ and Q). That is, the subdivision into primary and secondary concepts is in a way a subjective matter. And as I have already pointed out, it appears to be rational to consider a subdivision of primary concepts into fundamental concepts and other primary concepts, where the fundamental concepts are more primary, so to speak.

1.4 System and subsystem

If one has a subject Φ it is intuitively natural to contemplate its environment. But as is evident from the preceding section, the notion of environment E appears to require specification of some substance U. This idea of E is closely related to the notions $U_E = U - U_S$ and $Q_E = Q - Q_S$. For instance, if Φ is a rabbit, the situation is different if U is selected to be iron (we consider iron metabolism) compared to the situation we would have if U is selected to be iodine (iodine metabolism); that is, even if Φ is one and the same rabbit the two situations are entirely different: *we have two different systems*—and also two different environments.

Thus, the notion **system** S is forced upon us:

$$S \equiv \{\Phi, U, p_v(t, t_0, q_0)\} \tag{1.4.1}$$

So to repeat, in order to specify the environment E not only the subject Φ but also the substance U must be specified.

It may seem redundant to include the transition probability in the symbol S: if U is given, also $\mathbf{Q}$ and Q are given, which means that, from a practical point of view, also the transition probability is given as well. However, Φ and U are qualitative quantities, and the transition probability is necessary for the quantitative aspect on the notion of system. For instance, assume again that Φ is a rabbit and that U is selected to be iron. Assume further that the rabbit is exposed to some infection that affects the iron metabolism. We then have the same subject and the same substance before as after the infection; the only difference between before and after is reflected by the transition probability. That is, also quantitative aspects must be included into the symbol S, and the transition probability seems to be the most natural quantity to select for that purpose.

Let us now consider a part of Φ and denote it Φ_v, to which there corresponds $Q_v \subseteq Q_S$; should $Q_v = Q_S$ there is, from the viewpoint of *U*, no difference between Φ_v and Φ. It is then natural to introduce the notion of **subsystem**:

$$S_v \equiv \{S, \Phi_v\} \tag{1.4.2}$$

Assume as usual that Φ is a rabbit, and assume further that part of its food consists of AB molecules that in the animal are irreversibly dissociated: $\mathrm{AB} \to \mathrm{A} + \mathrm{B}$. If we identify $u \in U$ with A, the particle can be in two states "bound" (bound to B) and "free", where the latter is an absorbing state (the reaction is truly irreversible). And it is quite natural to look at this absorbing state as belonging to Q_E, and this means that the A particles are in *E*. But physically the reaction takes place in Φ, and it is not only possible but even likely that immediately when free A appears as a result of the chemical reaction it is in Φ. In other words that *u* enters *E* does not necessarily mean that it leaves Φ. On the other hand, should *u* leave Φ it definitely enters *E!*

This is a kind of uncomfortable asymmetry that is due the physical character of Φ (the notion of subject). For that reason I shall only make limited use of this notion—by and large it has done its job—and I shall mainly apply the concepts represented by the symbols *E, S, U, Q* and S_v. The expressions $Q_v \subseteq Q_S$ and $U_v \subseteq U_S$ are obviously equivalent, at the same time as we can observe that if a $u \in U_v$ it means that its state $q \in Q_v$, and the converse. But "physically" this is the same thing as saying that *u* is in the subsystem S_v and, therefore, I shall apply the notation $S_v \subseteq S$ in spite of the fact that, strictly speaking, the symbols S_v and *S* do not represent sets in the usual sense. Hence, we have three rather equivalent expressions $(U_v \subseteq U_S, Q_v \subseteq Q_S, S_v \subseteq S)$, and which one to use is matter of taste and what emphasis one for the moment wants to make.

Definition (1.4.1) should perhaps have been given the form

$$S \equiv \{U, Q_S, p_v(t, t_0, q_0)\}$$

because formally the notions of system S and subsystem S_v are more rationally defined in terms of the state sets $Q_v \subseteq Q_S$ $(\subset Q)$. That I here introduce and apply the notion Φ is merely an expression of an attempt of mine to make the approach less abstract so to speak.

As indicated at the introduction of Rule (1.3.1) the notion of System calls for a modification as we substitute Φ with *S*:

> Necessary for a *u* being a particle with respect to a system *S* is that it retains its identity while being in *S* and never returns once it has left *S*. (1.4.3)

1.5 Content

There are quantitative aspects other than the transition probability on the notions system, subsystem, $U_v \subseteq U_S \subset U$ and $Q_v \subseteq Q_S \subset Q$. The aspect I have in mind is simply the number of particles the corresponding *U*-population contains at the time *t*; phrased differently, we ask for the number of particles for which $q(t) \in Q_v \subseteq Q_S$.

Please note my writing $U_v \subseteq U_S$ and not $U_v \subseteq U$ $(= U_S \cup U_E)$ which would otherwise be seemingly a rational thing to do. The reason for this my writing is that the number of $u \in U_S$ is a *physically possible quantity*, whereas the number of $u \in U_E$ is physically impossible to specify; at most one can often say that it is immense. This is the reason why a $u \in U_S$ cannot enter *S* once it has entered *E*. This is actually according to Rule (1.4.3), namely that the fundamental *identity* of a $u \in U_S$ is drowned as it were as soon as the particle enters the huge population U_E; a parallel with the proverbial raindrop in the ocean.

Personally I view Q_v and Q_S as canisters that contain some particles $u \in U$ but as far canisters go they may be quite abstract matters. Therefore I introduce the notion of **content** $\mu_v(t)$ and $\mu_S(t)$ that basically stands for the expected number of $u \in U$ that at the time *t* have states $q(t) \in Q_v$ and $q(t) \in Q_S$ respectively. In other words, "content" stands for the number of particles $u \in U_v \subseteq U_S$, and often this number is contemplated with respect to the time *t*.

The numerical value of the content is in a way arbitrary as long as it is *directly proportional to the number of particles*. For instance, frequently a $u \in U$ is a molecule so Avogadro's number is involved, and that could result in uncomfortably large values of the content, which then is more conveniently measured in something like "gram-molecules".

Let us now return to the notion of system (1.4.1). Assume that we specify a subject Φ and a substance *U*. The whole situation would definitely be ridiculous if the subject cannot possibly contain any particles of that substance (e.g. we might be dealing with a live rabbit that we want to keep alive while the substance is a strong poison for that animal). Therefore, if we specify a subsystem $S_v \subseteq S$ it is necessary that there is a definite value of the content μ_v; yes, it is of course possible that incidentally the value is zero, but the content is nevertheless there as a meaningful physical quantity.

We note that it is necessary that to any $S_v \subseteq S$ $(U_v \subseteq U_S)$ there corresponds a $Q_v \subseteq Q_S$ such that at any time *t* there is a unique content $\mu_v(t)$, namely that "number of *u* for which the state $q(t) \in Q_v$" is a well defined physical quantity.

The notion of content thus comes out as something quite basic. The notion means that there is quite a demand on the concepts of system and subsystem. That is, not only must there be corresponding subsets, but these sets must have a well defined content at nay time *t*. In fact, it is occasionally convenient that a subsystem S_v is *operationally defined* by the experimental procedure that determines the content $\mu_v(t)$. Phrased differently, the procedure of the estimation of content is used as an **operational definition** of the notion $S_v \subseteq S$, and the concepts that go with it. For instance, a subsystem could be defined by a procedure like the following: one takes, with a syringe, a sample of blood from a specified place of a specified blood vessel of some creature, centrifuges that sample and, eventually, estimates the content (the amount of *U*) in the plasma by some physio-chemical method. *To that content there corresponds then a subsystem, together with a particle population and a set of states.* Naturally, such a subsystem may be (clinically, say) a rather abstract thing, and the estimated numerical value of the content, together with the procedure-of-estimation, is then what makes it concrete (real, as it were); Section 4.2 illustrates how this kind of definition may be applied.

Again, the notion of content comes out not only as something quite basic but also as something so concrete ("physical") that it substantiates otherwise abstract concepts.

1.6 Summary of Chapter 1

Let us start with **particle** *u*, which simply is a physical object signified by its motion. Any given particle belongs to a universal population of equivalent particles which we call **substance** *U*; two particles are **equivalent** if and only if their motions are realizations of the same stochastic process. To represent all aspects of particle at any particular point in time we assign it a **state** *q*, which usually is a time function $q(t)$. Much of the potential of the language comes from the fact that *q* is left unspecified. The state *q* is considered to be a location vector in a **state space Q** (a kind of co-ordinate system), where all states *u* can ever attend is the **universal state set** *Q* (i.e. it is always true that $q \in Q$).

The probability that the state of a particle *u* will belong to a specific state set at a specific point in time, *given* that it is in a given state at a given earlier time, is called the **transition probability** $p_v(t, t_0, q_0)$ and is a function of assigned parameters.

A **subject** Φ affects a particle throughout a consecutive period of time during which it is a property of the particle to be "in" the subject.

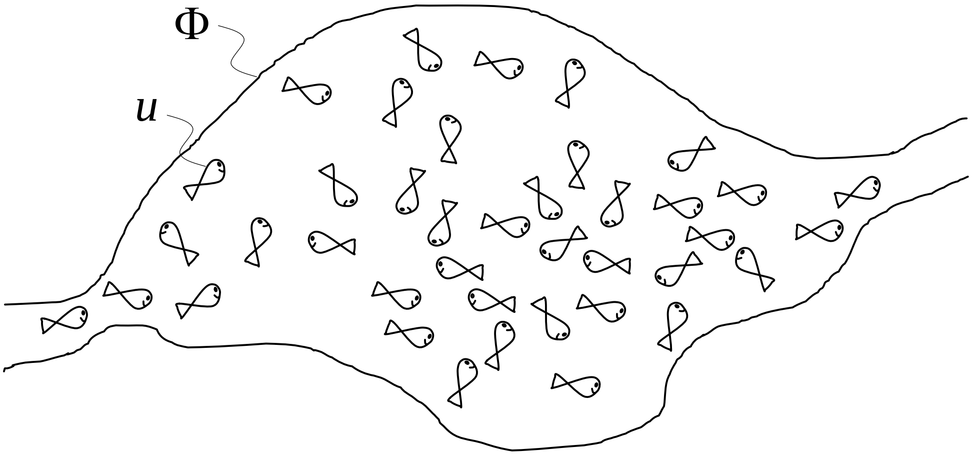

Fig. 1.6.1 An example of a system S *where* Φ *is a lake and* U *is fish of some kind; the influx is from one river and another river takes care of the efflux. The individual* $u \in U_S$ *is thus a fish and a significant state variable could be swimming velocity (a vector denoting direction and speed), which on the other hand hardly is an appropriate state variable should, say,* u *be a molecule. This illustrates one of the reasons why the* state of particle *should be left unspecified: the notion of state must be applicable to a variety of particles.*

The combination of a particular subject, substance, and transition probability is a **system** *S*. As whether or not being in the system *S* is a property of the particle and therefore covered by the particle's state *q*, we can distinguish between the states Q_S where a particle is in the system as opposed to in the remainder of states $Q_E = Q - Q_S$ where the particle is in the **environment** *E*.

Given a system we can discuss a **subsystem** S_v by regarding only the portion where a particle's state is in the subset $Q_v \subseteq Q_S$. This allows us to define the **content** $\mu_v(t)$ as the number of particles that, at the time *t*, is in a portion of the system determined by the state set $Q_v \subseteq Q_S$. Here, one can equally well say that the content $\mu_\nu(t)$ is the number of particles in the population $U_\text{v} \subseteq U_S$ at the time *t*.

Chapter 2
Motion

As pointed out in Chapter 0, a fundamental property of the kind of system we here contemplate is that the particles are in motion, and it is now time to consider this dynamic aspect more explicitly. Well, motion has already been touched upon in the preceding chapter, through the introduction of the transition probability $p_v(t,t_0,q_0)$ and the state q as a time function; perhaps somewhat particular is that, as explained in Section 1.2, this transition probability is here looked upon as the CAUSE of the particle's motion: the physical quantity *transition probability* acts on the particle.

Now, as stated by (1.1.2) the state q is left unspecified, but this does not mean that that quantity can be left totally free as it were. On the contrary some kind of formal structure (properties) must be imposed not only on q and the sample function $q(t)$ but also on notions like $\mathbf{Q}$ and Q; yes this involves also a quantity like $p_v(t,t_0,q_0)$. The reason for the requirement of formal structure is that the language must be compatible with the kind of mathematics (tool in Fig. 0.2.1) we shall apply at the construction of models. Without such a structure the language would be totally amorphous, incapable of serving as MATERIAL at the model construction that after all is a final goal. Or phrased differently, although the present chapter might occasionally appear abstract, it is, practically(!) speaking, an absolute necessity. At the same time it is important to note that the choice and wording of formal structure (like the postulates) is partly a subjective matter, and what essentially is one and the same theory can often be given quite different forms.

2.1 On the notion of stationarity

Before we proceed the crucial notion of stationary must be specified. Thus, let us consider the transition probability $p_v(t, t_0, q_0)$ introduced by (1.2.1), and let us further define the quantity

$$\Delta t \equiv t - t_0 \tag{2.1.1}$$

which leads to the substitution $p_v(t, t_0, q_0) = p_v(\Delta t + t_0, t_0, q_0)$. The Markov process (Section 1.1) is said to be **stationary** if and only if the transition probability is independent of t_0. That is, the transition probability is independent of when in time the initial time point t_0 represents. One could say that the t-axis is homogeneous in the sense that all points are equivalent as starting point and, therefore, instead of "stationary" one occasionally meets the term **time homogeneous**.

We can now simplify the rather clumsy notation of transition probability and consistently write $t_0 = 0$:

$$\pi_v(t, q) \equiv p_v(t, 0, q_0)\ ,\ q_0 = q(0) = q\ ,\ \ t \geq 0 \tag{2.1.2}$$

because, from (2.1.1) it follows that $\Delta t = t$ for $t_0 = 0$; as explained in the following IT WILL BE CONSISTENTLY ASSUMED THAT S IS STATIONARY, and it is therefore of no importance where on the time axis the point $t_0 = 0$ is chosen. The symbol $\pi_v(t, q)$ thus stands for the probability that the particle *u* will be somewhere in S_v at the time $t > 0$ *given* that its state is $q \in Q$ at the time zero. Please note also that here the time *t* represents the time of an individual particle; as a rule *t=0* is the time when the particle enters *S*.

The notion of stationary I have introduced here is in terms of the transition probability and it's independency on the starting point t_0 However, in the literature the term stationary has not an all that unambiguous meaning and, above all, the definition is often in other terms than the transition probability. For instance, one might find a term like *strong stationarity*, which means that the moments of the process are time independent (e.g. the mean value $\langle X(t) \rangle$ should be a time-independent constant), and usually this kind of stationarity does most certainly not apply to the processes we contemplate in DMP. A common term is *weak stationarity* (or, Khinchin stationarity), which refers to the covariance of the process and its time independency (the covariance depends only on the time difference between the considered two points of the process); this concept has been of some importance in time-series analysis (cf. Yaglom, 1962). In spite of their fundamental difference there is one salient feature that these two kinds of stationarity have in common, namely that the definitions are in macroscopic terms, whereas the DMP stationarity is microscopic in the sense that it is defined in terms of transition probability.

In the following I shall repeatedly state the importance of physical aspects, namely matters that we must take into account from the viewpoint of physics and which do not follow from formal (mathematical) considerations. And the notion of stationary is an example of this: the transition probability depends also on other factors than those explicitly listed, and in the next section the notion of *systemic factors* will be introduced. An example of such a factor is the particle distribution: if that changes with time the type of stationarity we presently contemplate is usually not feasible. This is because the simple truth is that the vast majority of systems one meets in nature are nonlinear, and one common reason for this is that the transition probability depends on the distribution of particles (e.g. there is a mutually interaction between particles). For that reason I shall in the next chapter introduce some time-independent distributions of particles, which are of major importance for the application of DMP.

2.2 Postulate I

From the preceding we have that to any subsystem $S_v \subseteq S$ there corresponds not only a unique set of states $Q_v \subseteq Q_S$, but each subset also has a definite content $\mu_v(t)$. That this content might be a function of the time *t* is an expression of the fact that the particles in *S* are in motion, a fact that is recognized by the very notion of the transition probability (1.2.1). In a way this means that I resume and continue the discussion in Section 1.1: the fundamental concepts *particle* and *state of particle* are indeed central, and this propagates to the role of the notion $Q_v \subseteq Q_S \subset Q$,

The next step is to note that this transition probability depends on other quantities than those explicitly stated $\{v, t, q\}$. Obvious examples are what in older literature were called intensive parameters like pressure and temperature; of course, it is often gradient fields of those quantities one has to deal with—especially in biology. Typical for the effect of such quantities is their collective character: they act on the physical mechanism in *S* as a whole. For that reason I refer to them as **systemic factors** (as mentioned in the preceding section, one such factor is the particle distribution). Several of these factors might be considered as belonging to *E* rather than to *S*; the diet of laboratory animals is an example.

> POSTULATE I: There is a finite set of systemic factors such that, when kept constant in time, $p_v(t, t_0, q_0)$ does not depend on how $t_0 = 0$ is chosen. That is, the motion of particle becomes a realization of a definite and STATIONARY (time homogeneous) Markovian process.

That the process is definite means that the transition probability is well-defined.

Nothing here is said about what the set of systemic factors actually is: it is up to the experimenter to identify the appropriate set. In other words, this postulate incorporates in the theory the simple fact that one always deals with the *art of experimentation*; in the present vocabulary this art consists of 1) to recognize the appropriate set of systemic factors, and 2) to keep this set constant in time. In other words, this postulate acknowledges the fact that experimentation is a piece of art.

> In the following we assume that the appropriate set of systemic factors is kept constant, as required by the postulate. That is, the motion of particle $u \in U_S$ is a realization of a STATIONARY MARKOVIAN PROCESS. (2.2.1)

Among other things the stationarity implies that, again, it is not important which point in time one selects as the time $t = 0$; i.e. $p_v(t, t_0, q_0)$ depends on the difference $t - t_0$ and not on the two quantities separately. Hence, because of (2.2.1)—the art of experimentation—we can now use the simplified notation from (2.1.2).

One implication of this Markovian property and stationarity is that a $u \in U_S$ does not age, nor does *S* age. And of course this is seemingly a severe limitation of the current theory presently under construction. However, many processes in nature, including biology, are "finished" long before age begins to have any significant influence; should age nevertheless be of significance a different form of language is required (cf. Bergner, 1965).

2.3 Postulates II and III

Let us consider the situation where there is no *U* in *S* at the time $t < 0$; that is, $\mu_S(t) \equiv 0$ for $t < 0$. Let us further assume that at $t = 0$ there is the amount $\mu_S(0)$ of *U* (number of *u* that is) injected into *S*. And let us for the moment be "abstract" and neglect all mathematico-physical details.

Well, there is one detail that must be taken care of and that is the simple fact that the content actually is a stochastic quantity, so the actual content of Q_v at the time *t* is $X_v(t)$ and, if we use the notation $\langle \xi \rangle$ for the mean value of the stochastic variable ξ, we have

$$\langle X_{\nu}(t) \rangle \equiv \mu_{\nu}(t) \tag{2.3.1}$$

Each *u* injected at $t = 0$ may be looked upon as a trial in the usual statistical sense (a Bernoulli trial) and, hence, the ratio $X_v(t)/\mu_S(0)$ is an estimator of $P_v(t)$, namely the *(absolute) probability* of a *u* to have a state $q \in Q_v$ at the time $t \geq 0$, given that the particle is in a state $q \in Q_S$ according to the initial distribution γ at the time $t = 0$ (the "initial condition" γ is introduced in Section 2.4 below). That is,

$$\langle X_v(t)/\mu_S(0) \rangle = P_v(t) \tag{2.3.2}$$

Here it is important to keep in mind that the probability refers to a particle that entered the system at the time *t=0* and, thus, has spent the time *t>0* in the system. Often there are a number of particles present in *S* and if their distribution is varying with time the function $P_v(t)$ can be quite messy (partly because of the nonlinear particle interaction mentioned in Section 2.1.

This kind of probability is necessarily a primary concept, so in DMP it consequently has a status similar to that of the transition probability. However, the two probabilities are fundamentally different but, as we shall soon see, two simple postulates implicate a deterministic connection between them.

The difference between transition portability $\pi_v(t,q)$ and the probability $P_v(t)$ is indeed fundamental. Thus, the former is conditional (Section 1.2) whereas the latter is in a way unconditional, and therefore often referred to as *absolute* probability (cf. Takács, 1960); as we shall soon see this probability is however not *totally* unconditional: as discussed in Section 2.4 below, there is a kind of condition also for $P_v(t)$, namely the initial condition γ.

> POSTULATE II: At any time *t* the distribution of particles $u \in U_S$ over the states Q_S is a realization of a definite probability distribution. This distribution has a density $\rho(t,q)$ such that the Lebesgue-Stieltjes integral

$$P_v(t) = \int_{Q_v} \rho(t, dq) \tag{2.3.3}$$

exists for $t \geq 0$ and all $Q_v \subseteq Q_S$, and has the ascribed significance.

The expression "all $Q_v \subseteq Q_S$" has its usual meaning, except for the restriction discussed previously, namely that we only accept *physically meaningful* sets. This means, we only accept sets that have content. In the light of the present postulate this is a kind of tautology—but an important one!

Let $f(q)$ be a finite and single valued state function for $q \in Q_S$. One experiences such functions more often than is commonly recognized. For instance, if U is a substance that is pharmacologically active, and the effect of an individual molecule u depends on its "state" (e.g. if and how it is located at a certain kind of nerve synapse), $f(q)$ could be the expected (macroscopic) effect of molecule. More generally speaking, we shall say that $f(q)$ is a **semantic function** if it, apart from being finite and single valued, is meaningful to the investigator (it could for instance be some hypothesis of a sort).

POSTULATE III: If $f(q)$ is a semantic state function on Q_S, the integral

$$\int_{Q_v} f(q)\rho(t, dq) \tag{2.3.4}$$

exists and is finite for $t \geq 0$ and all $Q_v \subseteq Q_S$.

A postulate of this kind should be easy to accept, which means that it should be close of being trivial. That is, in the present case, if the integral should not exist the function $f(q)$ could hardly be meaningful (i.e. the function would not be semantic in contradiction to the assumed condition). Phrased differently, for $f(q)$ to be semantic on Q_S it must at least be "well behaved" on that set of states.

2.4 The motion and initial condition

We observe that $\pi_v(t,q)$ in (2.1.2) is semantic and, therefore, one should be able to apply Postulate III. But what would that integral mean, if anything? Yes, that is one of the questions that naturally appear. And another such "natural" question concerns the work we have done so far: why this work, what does it lead to?

Now, in this section a theorem will be deduced that is more-than-half-the-way a result of the present approach, and which apart from answering the first question, also gives to the language the kind of stability we need according to Appendix 6. But it will become clear that this forces us to accept an additional notion, namely the notion of *initial condition* that is not only of fundamental importance theoretically but also of crucial significance for the design and analysis of experiment.

Or phrased differently, we shall obtain one of the very fundamental equations for stochastic processes that not only gives meaning to the integral but also makes it possible to proceed towards deterministic relations (the next chapter), which allow for a formal analysis of data (Chapter 4). The significance of this is that, as must be kept in mind, I have applied the theory of stochastic processes mainly for conceptual reasons and the stability problems this induces are often (not always) an annoyance and they must be taken care of (cf. Appendix 6 where is emphasized the need for determinism in the stochastic process).

To simplify matters let us assume that Q_S can be subdivided into a finite number of mutually exclusive sets δq_i ($i = 1,\ldots,N$), where δq_i is a set such that it contains the point (the state) q_i and which, in addition, is so "small" that

$$\pi_v(t,q) \approx \pi_v(t,q_i), q \in \delta q_i \tag{2.4.1}$$

Consider then the sum

$$A_v(t,Q_S) \equiv \sum_{i=1}^{N} \pi_v(t,q_i)\rho(0,\delta q_i), q_i \in Q_S \tag{2.4.2}$$

Here $\pi_v(t,q_i)$ refers to an individual *u* and is a conditional probability for the event that $u \in U_v$ at the time *t*, given (!) that $q(0) = q_i$; that is, the condition is that the particle *u* is in the state q_i at the time $t = 0$. The term $\rho(0,\delta q_i)$ is the absolute probability for *u* to have a state somewhere in δq_i at the time zero and, hence, the products in the right hand side of (2.4.2) are nothing but an application of the multiplication axiom of probability algebra. In other words, each product in (2.4.2) is the probability for the event that u is somewhere in $S_v \subseteq S$ at the time $t \geq 0$ and its state is somewhere in δq_i at the time $t = 0$.

But the sets δq_i ($i = 1, \ldots, N$) are mutually exclusive and, thus, the terms in the sum of (2.4.2) are probabilities for mutually exclusive events. This means that $A_v(t, Q_S)$ is, according to the addition axiom of probability algebra, the probability for the event that $u \in U_v$ at the time $t \geq 0$. We should thus have

$$A_v(t, Q_S) \approx P_v(t) \tag{2.4.3}$$

where the approximation is due to (2.4.1): $\pi_v(t,q)$ is not *exactly* constant for $q \in \delta q_i$.

Obviously the sum in (2.4.2) is an analog to the Riemann sum for ordinary integrals, but the "length of interval" is replaced by the "weight of set". Such sums are occasionally called Darboux sums, and a today classic description of that mathematics has been given by Cramér (1954). Here I simply make direct use of what has been done by mathematicians in the past, and replace (2.4.3) by:

$$P_v(t) = \int_{Q_S} \pi_v(t,q)\rho(0,dq) \tag{2.4.4}$$

This equation is thus a combination of the multiplication axiom and the addition axiom of probability algebra. It is consequently a rather particular thing, and it is sometimes referred to as the **Chapman-Kolmogorof equation**. And it is indeed quite fundamental for the theory of stochastic processes, and to see why this is so let us introduce the symbol γ for the concept of **initial condition**:

$$\gamma \equiv \{\rho(0,q); q \in Q_S\} \tag{2.4.5}$$

namely the distribution that applies to particles at the instant they have entered *S*; because of Postulate I and (2.2.1) it is immaterial where on the *t*-axis 0 is chosen. And from (2.4.4) it then follows immediately

> THEOREM 2.4.1: There is a probability distribution $\{P_v(t); Q_v \subseteq Q_S, t \geq 0\}$ that is uniquely determined by *S* and γ.

There are two things to observe here. The first one is that according to the definition of the symbol *S*, if the system is given also the transition probability $\pi_v(t,q)$ is given, so that part of (2.4.4) is determined; and as the initial condition γ determines the other part of the equation (the density or weight quantity) the theorem comes out loud and clear. The other thing to note is that one could have written $\{\rho(t,q); q \in Q_S, t \geq 0\}$ instead of the writing in the theorem. There are simply two possibilities to formally write the probability distribution over the state set Q_S at the time $t \geq 0$, and which one to use is a matter of personal taste that can change from one instant to the next.

Equation (2.4.4) gives the deterministic connection between the conditional (transition) probability and the absolute probability that I promised in Section 2.3 above. To this comes that the equation also gives an answer to the question asked at the very beginning of the present section, namely about the meaning of the integral on the right hand side of the equation: it is an absolute probability for the event $\{q(t)_{t>0} \in Q_v | \gamma\}$.

As an "existence theorem" the Theorem 2.4.1 has some significance. Thus, if the system *S* and the initial condition γ are given then the probability, for the event that a $u \in U_v$, is given for all acceptable $U_v \subseteq U_S$ and all non-negative values of the time *t* (please do recall that U_v is acceptable only if it has a well-defined content). In other words, we do know(!) that there is actually a function $P_v(t)$ that has a definite meaning and with which we can work; for instance, it is *meaningful* to attempt to numerically estimate that function from data.

Another implication of the theorem is the central role it ascribes to the initial condition that, for instance, can be practically difficult indeed.

However, the main importance of the theorem is probably the determinism it introduces. That is, as discussed in Appendix 6 it is necessary with some determinism in order to obtain the STABILITY that can be observed in nature and is needed for practical application of the theory as dealt with in Chapter 4.

In summary, we see how the postulates and primary concepts yield the basic equation (2.4.4), which then gives Theorem 2.4.1 that demonstrates explicitly the importance of the notion of initial condition γ and, in addition, offers a kind of needed determinism.

Of course, the choice of primary concepts and postulates is to some extent a subjective matter, and the present approach should be looked upon as a suggestion. But it is a good suggestion I guess because, as will be seen in the next chapter, not only does the resulting language contain a kind of necessary determinism but it also provides means for definition of such fundamental notions as open and closed systems; to this comes that it (the language) provides some answers to the two basic questions formulated in Section 0.1 and they have, as shown in Chapter 4, direct experimental application. Hopefully this is an adequate response to the second question asked at the beginning of the present section.

2.5 Postulate IV

Please recall the meaning of $\pi_v(t,q)$, namely that it is a conditional probability for *u* to have a state somewhere in Q_v at the time $t \geq 0$ (i.e. the considered event is $q(t)\big|_{t\geq 0} \in Q_v$ *under the condition* that $q(0) = q$). Clearly then:

$$\pi_v(0,q) = \begin{cases} 1 & \text{when } q = q(0) \in Q_v \\ 0 & \text{when } q = q(0) \notin Q_v \end{cases} \tag{2.5.1}$$

This can be looked upon as: *u* cannot change state during the time interval zero, which is nothing but an elementary physical fact. And from this fact follows from (2.4.4):

$$P_v(0) = \int_{Q_S} \pi_v(0,q)\rho(0,dq) = \int_{Q_v} \rho(0,dq) \tag{2.5.2}$$

Yes, this is exactly what one should get according to Postulate II, and here the point is merely that it is nice to see that ones reasoning is consistent (a kind of confirmation, so to speak).

But what happens if the time interval is not zero but $\Delta t > 0$? That is, instead of $\pi_v(0,q)$ we contemplate $\pi_v(\Delta t,q)$, or more generally $\pi_v(t+\Delta t,q)$. This is in fact a rather subtle question with physical—and then also mathematical ramification. It is thus easy to get involved in lengthy deliberation on this topic, and there is no room for this in present context. So as usual I shall try to take an easy way out and, to that end, introduce the **indicator function**:

$$I_v(t) = \begin{cases} 1 & \text{when } q(t) \in Q_v \\ 0 & \text{when } q(t) \notin Q_v \end{cases} \tag{2.5.3}$$

That is, in the derivation of (2.5.2) the π-function was used as an indicator function of a sort. Please note that $I_v(t)$ refers to the individual $u \in U_S$ or, phrased differently, each particle *u* in the system has its own indicator function.

It is important to recognize this indicator function, for fixed *t*, as a stochastic variable. It is actually a binomial variable with the mean value:

$$\langle I_\nu(t)\rangle = P_\nu(t) \quad t \geq 0 \tag{2.5.4}$$

Though being elementary the equation is quite important and will soon be applied.

Let us now write $B_\nu(t,\Delta t) \equiv |I_\nu(t+\Delta t) - I_\nu(t)|$, where $t,\Delta t > 0$. If $u \in U_\nu$ at the time *t* and at the time $t+\Delta t$, then we know that $B_\nu(t,\Delta t) = 0$, and the same thing happens if *u* stays out of S_ν all the time. Actually, the only instant when *B* is non-zero (equal to one, that is) is when the particle change state between S_ν and its complement $S - S_\nu$ during the time Δt; yes, it is assumed that Δt is so small that there is no time for the particle to both leave (enter) and re-enter (leave) during that time interval.

POSTULATE IV:

$$\lim_{\Delta t \to 0} P\left\{|I_\nu(t+\Delta t) - I_\nu(t)| > 0\right\} = 0$$

for all acceptable $S_\nu \subseteq S$ and $\Delta t > 0$.

Of course, one could have written $|\;| = 1$ instead of $|\;| > 0$; to me it is simply a question of personal taste—and for the moment (Sic!) I like the present writing of the postulate.

More formally, the postulate imposes a kind of the mildest possible form of continuity on the stochastic process, namely the process that motion of the particles realize. A similar result could have been obtained by imposing properties directly on the transition probability, but that would have involved quite some subtle mathematics and would be out of the frame of present purposes; to this comes that the physical significance of such work would have been substantially more difficult to grasp.

I have met different opinions among mathematicians about the need for this postulate; some say it is needed and others say it is not needed, at the same time as no one seems to like it. However, there appears to be no dispute about the mathematical "difficulty" that the probability determined by (2.4.4)—and then also by Theorem 2.4.1—is not well defined, analytically speaking. Thus, does $P_\nu(t)$ have derivatives, or is it integrable? One reason for this uncertainty should be that although the transition probability $\pi_\nu(t,q)$ has a well defined and understood condition (q(0)=q), the *probability for that condition is often zero.* Attempts have been made to handle this problem by the introduction of various models, but as the goal here is to obtain a current theory any approach in that direction is out of question—the current theory must be model-free. However, I shall nevertheless later discuss a particular model (a discrete model, Appendix 5) that actually induces some definite analytical properties on $P_\nu(t)$.

It is motivated to ask whether the Postulate IV actually is a "postulate" in the sense of the section 0.2. It could be argued that the postulate is a statement about the underlying physics (the probability for *u* to change state during Δt goes to zero with Δt), and therefore it is a model of a sort. However, the postulate appeals to my understanding of physics and probability in such a way that it is impossible for me to regard it as a model. Perhaps one should have used the term "law" instead of postulate (analogous to the "laws of thermodynamics"), but that would have made necessary a discussion of this notion "law" and just now I feel that this would be outside the frame of present purposes.

At any rate, the postulate is a kind of answer to the question about what will happen if we in (2.5.1) write Δ t instead of 0, or, more generally, if we write $\pi_v(t+\Delta t, q)$ instead of $\pi_v(t, q)$. Yes, of course we know that the two quantities (with and without Δt) approach each other when $\Delta t \to 0$, but *a priori* we don't know how. Something like Postulate IV is needed.

2.6 A matter of notation

In order to minimize the number of symbols (for me personally this is a matter of highest priority) I have not made systematic difference between symbols that refer to individual particles and those that refer to particle populations. This means that I put some burden on the reader: it is up to the reader to keep track of a quantity's significance in this regard.

Thus, the indicator function $I_v(t)$ defined by (2.5.3) signifies a property of an individual $u \in U_s$ which has spent the time $t \geq 0$ in *S* (each particle has its own indicator function). And as follows from (2.5.4), the same holds for $P_v(t)$.

But let us now turn to the notion of content introduced in Section 1.5. Thus, the content $\mu_v(t)$ is simply the expected number of particles $u \in U_v$ at the time *t*. Of some significance to note is that I have *not* written $t \geq 0$. The reason is quite simple, namely that there is no condition on the time *t* except that the investigator works according to a definite *t*-scale. To repeat more explicitly, the symbol $\mu_v(t)$ stands for the expected number of particles in the population U_v at some time *t* (thus, here the symbol *t* does not stand for the "age" of particle). One might say that the difference in the meanings of *t* is a matter of how the population of particles has been generated.

Usually this shall not cause any problem, though carefulness is called for, as exemplified by (2.3.1) and (2.3.2). In (2.3.1) *t* denotes the time as such whereas in (2.3.2) *t necessarily* stands for the length of time the particle has been in *S*, given that it entered at $t = 0$ (according to γ).

Chapter 3
The Open System

The general notion of system is taken care of in the preceding two chapters, and what remains now is to clarify the specific notions of *open* and *closed* systems. Thus, a formally precise definition of openness and closeness will be introduced in this chapter. However, perhaps the main result to be presented – in this chapter – is the mass-time equivalence, which has indeed both theoretical and practical implications. Chapter 4 discusses some applications of that equivalence principle.

3.1 Mean sojourn time

Fig. 3.1.1 depicts an imagined trajectory of a particle *u* passing through an open system *S*, and in particular its behaviour with respect to a proper subsystem S_v. Thus, in this thought-example *u* enters and leaves the subsystem three times, and with each trip is associated a **transit-time**, namely the time spent by *u* in S_v between entrance and exit.

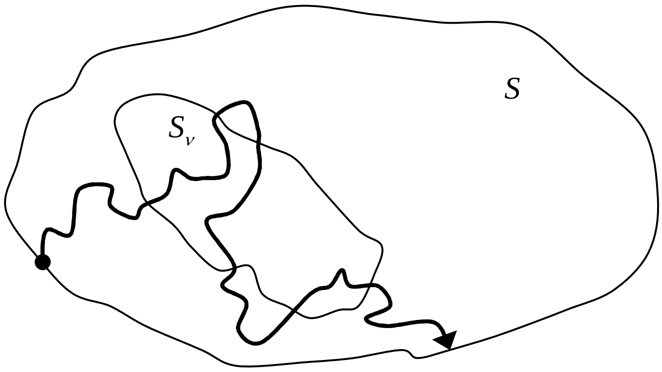

Fig. 3.1.1 While passing through the open system S the thought particle visits the subsystem $S_v \subset S$ *three times. To each visit there corresponds a transit time, and the sum of these time quantities equals the sojourn*

> *time (for this particular particle that is). It may perhaps be possible for a particle to move through S without ever visiting* S_v*, and then the sojourn time is, incidentally, zero. The main concept here is "mean sojourn-time" (for which the symbol* θ_v *is used), and it could be visualized as the average value taken over a large population of particles.*

In this example the TOTAL time *u* spends in S_v before leaving *S* for good is the sum of three transit-times, and that sum of times is named **sojourn-time**. Hence, we have here two different time quantities, namely transit-time and sojourn-time.

Let us now make use of the indicator function $I_v(t)$ introduced in the preceding section. Thus, for an individual *u* we can contemplate the stochastic integral

$$\eta_v(T) \equiv \int_0^T I_v(t)dt \tag{3.1.1}$$

that, because of Postulate IV, exists in quadratic mean for all $S_v \subseteq S$. If one considers the corresponding Riemann sum it is not difficult to grasp the meaning of $\eta_v(T)$, namely the total time the particle has spent in S_v when it has spent at most the time *T* in the system—it might have left *S* before $t = T$ (it is for simplicity assumed that *u* enters the system at the time zero).

Now, one easily grasps from the corresponding Riemann sum that $\eta_v(T)$ is a stochastic variable with the mean value according to (2.5.4):

$$\langle \eta_v(T) \rangle = \int_0^T \langle I_v(t) \rangle dt = \int_0^T P_v(t)dt \tag{3.1.2}$$

That is, as was questioned in the preceding section, the present function $P_v(t)$ should be integrable (see the next section).

If the limit

$$\theta_v \equiv \lim_{T \to \infty} \langle \eta_v(T) \rangle \tag{3.1.3}$$

exists, then it is called the **mean sojourn-time** of S_v with respect of *U*:

$$\theta_v = \int_0^\infty P_v(t)dt \tag{3.1.4}$$

The symbol θ_S is used when $S_v = S$.

From the point of view of practical application it is important to note that this equation together with (2.3.2) shows how θ_v can be estimated experimentally, namely

$$\theta_v \approx \frac{1}{\mu_S(0)}\int_0^\infty X_v(t)dt \tag{3.1.5}$$

This will be applied in the next chapter.

Of course, similar to mean sojourn-time one can also speak of **mean transit-time**: the expected time a $u \in U_v$ spends in S_v between entrance and exit of that subsystem. That is, also transit-time is a stochastic variable and there is quite a likely-hood that it has a definite expected-value; among several things, the value of an individual transit-time depends on at which state of Q_v the particle *happens* to enter S_v.

The mean sojourn-time is quite a central quantity and it is important to thoroughly grasp its meaning: the *total* time a $u \in U_S$ is *expected* to spend in S_v before leaving *S* for good. This significance is very much due to that "leaving *S*" is an irreversible process—as explicitly stated by (1.4.3). And from a practical point of view it is important to have a good idea of what factors influence the value of θ_v, and personally I have experienced that the following factors are important:

- The expected length of individual transit-times.
- The probability for a $u \in U_S$ to ever enter S_v.
- The number of times a $u \in U_S$ can be expected to reenter S_v.

(3.1.6)

For instance, the transit-time might be very long whereas θ_v is small, because the probability for a particle to ever enter S_v is quite small. Yes, there are several pitfalls, and simultaneous thinking in terms of transit-time and sojourn-time may occasionally be somewhat tricky.

There is a complicating factor here, namely that *mean* sojourn-time is not just the sum of mean transit-times but, as already indicated, there is a probability involved, namely the probability of a particle $u \in U_S$ to (ever) visit S_v before leaving the system for good. How that probability enters the calculations is a complicated matter, which fortunately is automatically taken care of in an expression like (3.1.4). For instance, a particle can as a principle pass through *S* without ever visiting $S_v \subset S$, in the case of which the sojourn-time for that particular particle is zero (of course, this is incidental and it is indeed not only possible but necessary that the mean value $\theta_v > 0$).

One way to keep hold of the difference between mean transit-time and mean sojourn-time is that the former is conditional but not the latter. That is, mean transit-time refers to a particle that *has* entered S_v whereas this is not so for sojourn-time. Well, as with $P_v(t)$, mean sojourn-time is not *totally* unconditional: the condition is that the particle *has* entered the system *S* according to γ.

> THEOREM 3.1.1: Only when $S_v = S$ (i.e. $\theta_v = \theta_S$) can mean sojourn-time be *generally* interpreted as a mean transit-time.

This theorem is essentially due to the fact stated by (1.4.3) that should a $u \in U_S$ leave *S*, it does so irreversibly. Of course, the theorem is self-evident from the preceding.

Yes, there is a reason for me being somewhat fussy about the distinction between the two terms transit-time and sojourn-time, and that reason is the history of the last sixty years of kinetics: in the past the notion of transit-time has dominated the thinking to the extent that it has hidden the "mass-time equivalence", which is a basic relation that will be derived in Section 3.3 below. The matter is further discussed in Appendix 1.

3.2 The open and closed system

In Section 0.1 it is emphasized that in an open system the trajectories generated by the particle motions cannot be made arbitrarily long. By and large this means that the expected time a particle spends in the system must be finite. And from this simple observation follows naturally the definition:

> *S* is **open** if and only if $\theta_S < \infty$ exists for all initial conditions γ. If in addition $\theta_v > 0$ for all γ and all $S_v \subseteq S$, the system is both open and **fully connected**.

The haunting question about $P_v(t)$ being integrable now has an answer of a sort: according to the definition, if *S* is open the function $P_v(t)$ is integrable.

That the system is fully connected means that that a particle $u \in U_S$ can reach any state in Q_S from any other state in Q_S, within finite time (this holds even when one deals with continuous state space and the probability for such a transition is zero). In the past it has been frequently assumed that the investigated system is fully connected, very much because it has occasionally simplified the math and the understanding with respect to the theoretical instruments that have been at hand. It seems that the fully connected system is something one should expect to meet occasionally, and it will subsequently be applied in a couple of instances.

One matter that has been of concern among investigators is the notion of **absorbing states**: a state is said to be absorbing if a $u \in U_S$ can enter the state but not leave it. Or, phrased differently, a particle enters an absorbing state irreversibly. Automatically then, if the subject Φ contains absorbing states they belong to the environment *E* (it was this feature I had in mind at the discussion of the reaction $\mathrm{AB} \rightarrow \mathrm{A} + \mathrm{B}$ at the end of Section 1.4).

Now, a reasonably good understanding of "open *S*" cannot be obtained without the notion of "closed *S*". And to this end we need the concept of **relative mean sojourn-time**:

$$\sigma_\nu \equiv \lim_{T\to\infty} \frac{\langle \eta_\nu(T) \rangle}{T} \tag{3.2.1}$$

where I use the notion introduced by (3.1.1).

It is not difficult to see that, if the limit exists,

$$0 \le \sigma_\nu \le 1 \tag{3.2.2}$$

And this gives us the necessary tool for the definition: *S* is said to be **closed** if and only if for every $S_\nu \subseteq S$ there is a γ_ν such that there exists a $\sigma_\nu > 0$ for $\gamma = \gamma_\nu$. If in addition $\sigma_\nu > 0$ exists for all γ and all $S_\nu \subseteq S$, the system is said to be both closed and **fully connected**.

The present definitions of open and closed system have one important thing in common – when *S* is fully connected – namely that they exclude absorbing states. However, whether or not there are any absorbing states present is very much a practical matter. For instance, ions like Na^+ and Ca^{2+} are commonly called "bone seekers" because of their tendency to have quite long transit-time in bone material: once such an ion is picked up by some bone crystal it appears to stay there. In fact, practically speaking (e.g. clinically speaking), it occasionally looks as if such particles were permanently trapped in parts of the skeleton. The reason for this "look" is that some of the bone processes are so slow that the particles appear to be *permanently* trapped (absorbing states!) when considered in time scale of the organism's life span. Yes, it is very much a practical matter.

It is easily seen that

$$\theta_S < \infty \Rightarrow \sigma_S = 0 \tag{3.2.3}$$

but the converse is not true. There are for instance data that suggest the model $P_S(t) = a(1+t)^{-1}$, and if that function is inserted in the appropriate integrals the result becomes "$\theta_S = \infty$ and $\sigma_S = 0$". In a somewhat subtle way such a system is neither open nor closed according to present definitions and it is therefore called a **hermaphroditic system**. Yes it has been argued that such systems cannot occur in nature, but from a formal point of view they exist—at least as a logical possibility. And, as with "absorbing states" discussed above, also this is a practical matter: within the time scale that the experimenter (e.g. clinician) works, *S* might behave *as if* it contained absorbing states or is hermaphroditic or both.

Apart from what is obvious from above, there are features in terms of which open and closed system differ in quite a fundamental manner. One such feature has to do with "limit distributions" to which I shall return later in this chapter.

3.3 Preliminary derivation of the mass-time equivalence

Let *S* be a system that is open for a substance *U*. From the definition in the preceding section we know that, for the mean sojourn-time of a subsystem $S_v \subseteq S$,

$$\theta_v = \int_0^\infty P_v(t)dt \tag{3.3.1}$$

exists and is finite. Note that nothing is said about the system being fully connected.

We shall consider the situation where open *S* does not contain any substance *U* for $t < 0$; i.e. $\mu_v(t) = X_v(t)\big|_{t<0} \equiv 0$. Assume now that at the time $t = 0$ is started a deterministic flow of particles that we label ζ^0_{SE} (subscript *SE* stands for "to *S* from *E*", and throughout the superscript *0* will be used to denote that the corresponding quantity is constant in time and, often, that the system is in steady state); this means that $\zeta^0_{SE}\Delta t$ = number of particles entering the open *S* during the time interval Δt. *We also assume that all the particles $u \in U_S$ have entered the system S according to a predetermined fixed γ.*

This assumption about the initial condition γ is not as alien as it might sound. For instance, in normal man there is a reasonably constant uptake of sodium ions (and a corresponding efflux), mainly via the food. But in food Na^+ appear in different forms, like free ions (table salt) and inside *a great variety* of cells (animal, vegetable) where the ions can be in a great variety of forms. In short, the sodium ions enter *S* at various states in Q_S , and γ is simply the probability distribution—realized by the "new" particles—over those "states of entrance". To keep γ constant might then simply mean that the amount and quality of food intake is kept constant. At occasions it might however be quite difficult a task to keep γ satisfactorily constant, and then we touch upon the *art* of experimentation discussed in connection with Postulate I (Section 2.2).

Let us now consider the time interval $[0,T)_{T<\infty}$ subdivided into the sequence $\Delta t_0, \Delta t_1, \Delta t_2, \ldots, \Delta t_n$ of consecutive and mutually exclusive time intervals of equal length Δt. During each such interval, $\zeta_{SE}^0 \Delta t$ is the number of particles that enter *S* and, to simplify the notation, we assume that all those particles enter at the very beginning of the time interval. Thus, at the beginning of the interval Δt_0 the time is $t = 0$, and at the beginning of the interval Δt_k ($0 \le k \le n$) the time is $t = k\Delta t$; at the end of Δt_n the time is $T < \infty$.

The following model will be modified in the next section:

> MODEL (assumption): All $u \in U_S$ are mutually independent, in the sense that all indicator functions are stochastically independent binomial variables with the same distribution, except for $k\Delta t$-units time shift.

To see the meaning of this let us note that $I_\mathrm{v}(k\Delta t)$ is the indicator function of a $u \in U$ that entered *S* $k\Delta t$ time units earlier and, moreover, all those *u* that enter during (rather, at the beginning of) that interval have equal indicator functions. From (2.5.4) we have for the mean value

$$\langle I_\mathrm{v}(k\Delta t)\rangle = P_\mathrm{v}(k\Delta t) \tag{3.3.2}$$

where $P_\mathrm{v}(k\Delta t)$ is the probability for this *u* to belong to the population U_v at the time $k\Delta t$. But there are $\zeta_{SE}^0 \Delta t$ such particles (i.e. the particles that entered the system $k\Delta t$ time units earlier) and from the probability algebra we then know that $\zeta_{SE}^0 \Delta t P_\mathrm{v}(k\Delta t)$ is the expected number of that particular kind of particles in U_v at the time $t = k\Delta t$. Obviously, the content of U_v at the time $n\Delta t$ is

$$\mu_\nu(n\Delta t) = \zeta^0_{SE}\Delta t\sum_{k=0}^{n} P_\nu(k\Delta t) \tag{3.3.3}$$

If we move Δt inside the summation sign the result is a classic Riemann sum, which means that

$$\lim_{\Delta t\to 0}\mu_\nu(n\Delta t)\Big|_{n\Delta t = T<\infty} \equiv \mu_\nu(T) = \zeta^0_{SE}\int_0^T P_\nu(t)\,dt \tag{3.3.4}$$

Here enters now *physics* in the picture. Thus, classic thermodynamics is based on the empirical experience that if an isolated system is left alone for a "sufficiently" long time it reaches equilibrium and, similarly, if an open system as defined here is left alone with a constant input of a substance it eventually reaches steady state. We translate this as

$$\lim_{T\to\infty}\mu(T) = \mu^0_\nu \tag{3.3.5}$$

because from (3.3.1) we know that the limit exists. That is,

$$\mu^0_\nu = \zeta^0_{SE}\theta_\nu \tag{3.3.6}$$

This is the sought MASS-TIME EQUIVALENCE, but the derivation is based on a specific and in fact unrealistic model, and this matter will be dealt with in the next section.

3.4 On the derivation of the mass-time equivalence

Mathematics is often great, but one must then recognize that it is entirely based on practical experiences (the term "practical" can or even should often be replaced by "physical"). In a way this is exemplified by Postulate III: for a function to be semantic it must be well behaved. And of course there are no formal reasons for this demand—the reasons are practical, based on experience.

Let us consider the derivation of the mass-time equivalence in the preceding section. The procedure of derivation used there is often a physical impossibility: a build up of steady state is usually not possible along the lines used, and one reason for this is simply that the particles are commonly not mutually independent as has been assumed, and instead one must *start* with a system already in steady state—how it originally has reached that state is not a problem of ours. *This is similar to the situation in* *classic thermodynamics: the given system is already in equilibrium, and the question of how this has happened is not a concern of that science.* That is, as will be shown below we have the following theorem from (3.3.6) for the **mass-time equivalence**

> THEOREM 3.4.1: If an open system is in steady state then, for any meaningful subsystem $S_v \subseteq S$,
>
> $$\mu_v^0 = \zeta_{SE}^0 \theta_v \qquad (3.4.1)$$
>
> where ζ_{SE}^0 is associated to a definite γ.

It is crucial to note that *the linear look of (3.4.1) is misleading*. For instance, if there is a change in ζ_{SE}^0 one should not be surprised if there is some change in θ_v as well; that is, μ_v^0 is not necessarily (in fact, usually not) a linear function of ζ_{SE}^0.

The Theorem 3.4.1 is somewhat remarkable in the sense that presently it seems to be the only generally established relation between the steady-state population and the time process that generates it. Another particularity to observe about (3.4.1) is that it is the flow into *S* that is crucial, not the flow into the proper subsystem $S_v \subset S$ as one, I think, would intuitively assume.

The simple fact is that most systems one meets in nature are non-linear or, phrased differently, usually there are mutual interactions between the particles $u \in U_S$, and this results in various non-linear relations. In chemistry one often has been speaking of "ideal solutions", which essentially has meant water solutions that are so diluted that there are no significant interactions between the particles $u \in U_S$, with various forms of linearity as a result; similarly, the term "ideal gas" is frequently used for a gas that is of a sufficiently low pressure so that the common "linear" equation $PV = nRT$ applies.

ON THE PROOF OF THEOREM 3.4.1

To get an understanding of the importance of the steady state condition, let us consider the population δU_S of particles that enter the steady state system *S* during the short time interval δt. The content of that population is $\zeta_{SE}^0 \delta t$, which is so small that it does not cause any perturbation of *S*. As soon as these particles enter *S* they become mutually independent because of the large number of steady state particles $u \in U_S$ surrounding each of the $u \in \delta U_S$. At the time when this happens each one of the particles in δU_S will experience itself being surrounded by a *stationary* phase of particles (the steady state population) WITH WHICH IT INTERACTS but, and this is important, that interaction is stationary; and this experience of the individual particle goes on, as long as the particle stays in the system. The same holds for particles entering the system during subsequent time intervals δt and, eventually, all the particles originally present have left *S*, and U_S consists by and large of particles that satisfy the model used for the derivation of (3.3.6).

But here comes the practical reality into the picture. Hardly any flow of particles into a system is truly *constant*; for instance, frequently the particle uptake is quite irregular because it is associated with the food intake. But practical experiences teach us that the "dynamic inertia" of a steady state population is often so large that those variations merely show up as ripples on the surface as it were. But, naturally, this does not mean that one should disregard the matter totally. That is, an expression like (3.4.1) should occasionally be regarded as a model of a sort, and the notion "test of model" might become a piece of reality that must be taken into account.

Yes, there is a matter of notation. That is, μ_v^0 and ζ_{SE}^0 refer to *S* as a whole, they are a kind of collective quantities so to speak, whereas θ_v refers to the individual $u \in U_S$ (entered *S* according to γ) and for that reason there is no superscript ^{0} on that time symbol. Thus, θ_v denotes the expected behavior of the *individual* particle entering *S*: it is meaningless to speak of steady state for a single particle.

Finally, by now we can see an answer to question (i) in Section 0.1. Thus, necessary for steady state is, according to the definition of open system in Section 3.2: that θ_S exists and is finite. For the condition also to be sufficient it is necessary that the θ-quantity is constant, which means that "all essential" parameters of *S* must be constant (including the presence of a constant ζ_{SE}^0, of course) but that merely brings us back to Postulate I in Chapter 2. These necessary and sufficient conditions could have been formulated in terms of the transition probability $\pi_v(t,q)$, but they are not—one could perhaps say that the present form of DMP is in this regard macroscopic.

3.5 On asymptotic distributions

We have just seen how physics enters into the analysis; for instance, the stead-state-condition in Theorem 3.4.1 is not there for mathematical reasons but for physical reasons. Thus, when an open *S* is left for itself, and all relevant factors are kept constant according to Postulate I, steady-state is eventually reached. This is first of all a physical experience, and it is for this reason one writes:

IF

$$\lim_{T\to\infty}\langle\eta_\nu(T)\rangle \text{ exists} \qquad (3.5.1)$$

THEN

it is called mean sojourn-time.

The phrase "if... then..." is thus crucial. That we actually expect the limit to exist is because of our physical experience—there are no formal ("mathematical") reasons for our expectation. And this goes for several of the limit properties that will appear in the section below.

One thing about Postulate I is that it does not say whether *S* is open or closed. So the matter about stationarity should hold for both open and closed *S*. And this is exactly what we require: classic thermodynamics is based on the "physical law" that says that if an isolated *S* is left for itself under the condition of (2.2.1) it eventually reaches equilibrium. And the language must be consistent with this for the current theory to be applicable to situations (systems) where the classic theory is applicable (the correspondence principle, Appendix 8 and Section 4.2). A kind of reflection of this is what I have tried to stress, namely that the concept of steady-state for open system corresponds to the concept of equilibrium for closed system. And both concepts are heavily based on the notion of stationarity of particle motion—together with physical experiences; that is, even if the two system conditions are fundamentally different they do have a few non-trivial features in common, one of which is the demand for stationarity.

Let us now continue the consideration of "physical experience", and begin with a closed *S* that is stationary according to Postulate I. Thus, when *S* is *closed* one expects that

$$\lim_{t\to\infty} P_\nu(t) \equiv P_\nu^e; S_\nu \subseteq S \qquad (3.5.2)$$

exists, where P_ν^e is a positive constant. As shown by Khintshine (1960), recalling that σ_ν is the relative mean sojourn-time,

$$\sigma_\nu = P_\nu^e \qquad (3.5.3)$$

where P_ν^e is referred to as equilibrium probability. Or, more generally, $D^e \equiv \{P_\nu^e; S_\nu \subseteq S\}$ is referred to as the **equilibrium distribution**. And it is a complete distribution in the sense that $P_S^e = 1$.

Nothing similar holds for open system. Thus, when *S* is *open*

$$\lim_{t\to\infty} P_S(t) = 0 \qquad (3.5.4)$$

Such a system can naturally not be in (non-trivial) equilibrium.

However, we have already noticed that it is steady state that matters when S is open. So let us simply define the steady-state probability as

$$P_\nu^0 \equiv \frac{\mu_\nu^0}{\mu_S^0}; S_\nu \subseteq S \tag{3.5.5}$$

This is nothing but the original understanding of probability (its success by thinking in terms of random selection of white and black balls in an urn). And, consequently, we have the **steady-state distribution** $D^0 \equiv \{P_\nu^0; S_\nu \subseteq S\}$.

One common characteristic of the two distribution D^e and D^0 is of course that they are constant in time. If we look for more such distributions, one possibility would be

$$P_\nu^l \equiv \lim_{t\to\infty} \frac{P_\nu(t)}{P_S(t)} \tag{3.5.6}$$

If(!) this limit exists it would be reasonable to call $\{P_\nu^l; S_\nu \subseteq S\} \equiv D^l$ a **limit distribution**. Like the other two limiting distributions also this one is complete, because $P_S^l = 1$. For closed systems the limit and the equilibrium distributions are identical, but for open system this is hardly ever so. To see this feature of open S one can use the mass-time equivalence and get from (3.5.5)

$$P_\nu^0 = \frac{\theta_\nu}{\theta_S} \tag{3.5.7}$$

That is, we should indeed
expect that

$$P_\nu^0 \neq P_\nu^l \tag{3.5.8a}$$

Or more generally,

$$D^0 \neq D^l \tag{3.5.8b}$$

In spite of its simplicity, which borders to triviality, the expression (3.5.7) has some significance of its own: it shows a kind of relation between a space distribution (P_ν^0) and the underlying transient time process (θ_ν); of course, it is a direct result of the mass-time equivalence, which does just that. This means that we touch on the notion of ergodicity, which is a matter treated in Appendix 3 (where also is considered the answer to question (ii) in Section 0.1)

One could thus say that there are two time independent probability distributions for open *S*, whereas closed *S* has only one such distribution. So from that point of view the open system is richer, so to speak, than is the closed system (almost drivel, but why not). At any rate, questions concerning existence and properties of D^l seemingly require specific models.

A particular problem in present context has to do with the definitions of openness and closeness, because they are in terms of a specific substance: it is indeed possible that if another substance, than the one originally chosen, is considered the corresponding system will be classified differently. This may be of importance when it comes to classic thermodynamics: in particular, the present concept of equilibrium might differ from the similarly named concept of classic theory. This matter is discussed in Appendix 2.

Chapter 4
Two Applications

As goes for all PARTICULAR THEORIES (applications), more or less specific models are involved and, hence, heuristic elements are included in the proceeding *ad hoc* analysis. This analysis will be deterministic, because I shall disregard the underlying stochasticity; by and large this means that it is silently understood that the number of particles is assumed to be "large" everywhere and all the time.

4.1 A particular bioaccumulation

Organisms of all kind commonly demonstrate ability to pick up and retain various non-desirable substances (pollutants), which thereby become part of the "food chain" and may cause harm to the ecosystem including man. The following is a specific case of that matter—qualitatively discussed in terms of DMP.

Thus, let an open *S* be a fish (e.g. shark or pike) or a mollusk (e.g. oyster). The considered substance *U* is the polluting substance and W_S denotes the total (body) weight of the organism *S*. If $\mu_S(t)$ stands for the amount of U in *S* at the time *t* (the content of U_S that is) the **weight concentration** is defined as

$$w(t) \equiv \frac{\mu_S(t)}{W_S(t)} \tag{4.1.1}$$

which means that $w(t)$ refers the total system: there will be no need for the notion of subsystems – what is measured is the amount of pollutant per unit body weight. The practical procedure behind (4.1.1) often consists of the making of a homogeneous blend of the organism by placing it in a mixer and the recording of $\mu_S(t)$ becomes then an easy task.

The *basic assumption* throughout is that S is in (or sufficiently close to) steady state, which means that

$$w(t) \equiv w^0 \tag{4.1.2}$$

Implicit in this assumption is that whatever the changes are in $W_S(t)$ and $\mu_S(t)$ they are slow compared to the rates of the physico-chemical processes that determine the metabolism of U and, hence, of the distribution of particles throughout S. And to the best of my knowledge there are no empirical findings whatsoever that even hint that this assumption should be unrealistic.

EMPIRICAL DATA

When W_S increases w^0 tends to increase for fish (shark) and decrease when S is mollusk. (4.1.3)

This is thus an empirical finding, and occasionally it has had commercial significance (e.g. fish can be toxic if it is large, and then impossible to sell.) Although there are quantitative data in the literature my general impression is that by and large those data only allow for qualitative analysis (Appendix 7). Therefore, only qualitative aspects on the empirical findings are here considered.

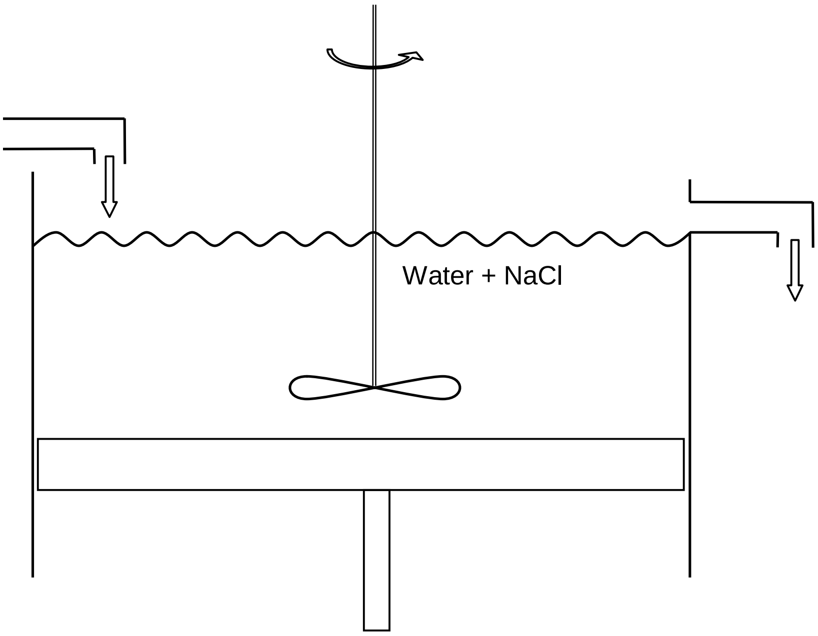

Fig. 4.1.1 This could be called a physically classic system. There is a constant input of a water solution of NaCl (table salt) and, to fix ideas, let the volume concentration of salt be 0.2 mg/l. Provided that the stirring is efficient, this is the concentration throughout the volume (weight) W_S *(in general we shall consider the weight of the system and for the present water solution of NaCl it is about the same as volume). There is a balancing efflux of salt solution with the same concentration (0.2 mg/l). If the piston is moved a few centimeters downwards the weight (volume) increases but* THE CONCENTRATION STAYS THE SAME. *In this section we shall however deal with a situation where the system* S *is a living organism, and where one has empirically observed a systematic change in the substance's concentration when* W_S *increases. Earlier this kind of finding has been used as evidence for biologic systems violating physics. However, it turns out that at least this particular phenomenon has a simple explanation in DMP terms, without any violation of physical principles.*

These empirical results are somewhat difficult to grasp in terms of classic physics, as Fig. 4.1.1 tries to illustrate. Moving the piston down increases the size (weight) of the system, but the concentration in terms of w^0 does not change. In fact, results of the type (4.1.3) have occasionally been used as a kind of "proof" for biologic systems violating physical principles (especially the second law of thermodynamics has been a popular target). However, as will be shown here the results (4.1.3) have actually a rather simple explanation that in no way violates any physics.

PROBLEM
How to comprehend the properties of w^0 described by the findings (4.1.3), *given* that S is in steady state according to (4.1.2)?

Note that this means that (4.1.2) is implicit in the problem-formulation and is thus not a specific assumption as the text immediately above the equation might suggest. That is, whatever model may subsequently be called for, the steady-state condition (4.1.2) is not a part of it.

The first step of the approach to this problem is to apply the mass-time equivalence:

$$w^0 = \frac{\zeta_{SE}^0 \, \theta_S}{W_S} \tag{4.1.4}$$

PROBLEM REFORMULATED
Get the nominator in (4.1.4) in the form of a function of the organism's weight, so that the dependency of w^0 on the weight W_S can be made explicit, which then can be compared with the empirical data (4.1.3). Or phrased on a slightly different form

> How do ζ_{SE}^0 and θ_S depend on W_S when S is a fish or a mollusk respectively? (4.1.5)

Hitherto we have not formulated a model. That is, so far we have only translated the situation and associated experimental findings into DMP. But now comes the first part of the model to be used in this study:

> For most organisms the uptake of most substances increases with increasing weight of organism. (4.1.6a)

and from this is concluded that

$$W_S \uparrow \Rightarrow \zeta_{SE}^0 \uparrow \tag{4.1.6b}$$

which is the model about ζ_{SE}^0 we shall use (the other part of the model concerns the mean sojourn-time θ_S, and is formulated below).

One could say that this is a model of the common experience that the demand for oxygen and various nutritious material increases with increasing biomass of S, and this demand is met by, among other things, increase in absorption surface (e.g. of gels and intestine). The model (4.1.6) expresses merely that it would be queer indeed if this increase should not result in increased uptake also of non-requested but available substances like *U*. Among other things, we know from the irreversible thermodynamics that there is often a positive coupling between various flows of matter and energy; for instance, an increase in uptake of one or a few substances often induce increase in the uptakes of other substances as well.

The next step is to formulate a model for how θ_S depends on W_S, and for that purpose let us consider *S* as composed of *n* distinct cells, so that

$$\theta_S = \sum_{i=1}^{n} \theta_i \tag{4.1.7}$$

where the mean sojourn-time θ_i refers to cell number i. This can be given the approximate form

$$\theta_i \approx \tilde{P}_i (1 + r_i) \tau_i \tag{4.1.8}$$

with $\tilde{P}_i =$ the probability for a *u* that enters *S* to ever enter the cell number *i*; $r_i =$ expected number of times *u* re-enters the i^{th} cell; $\tau_i =$ expected length of time a *u* stays in cell number *i* after entering (transit time).

Obviously τ is the mean transit-time. So here we have a clear illustration of why it is necessary to discriminate between transit-time and sojourn-time. This "discrimination topic" is treated in Chapter 3 and Appendix 1. Yes, there are a number of reasons of why a clear distinction between the two time concepts is crucial.

This brings us to the notion of cell population as a group of equivalent cells (like liver cells or heart cells). They are not equivalent because they have the same $\tau -$ value (which they don't) but because the cells' individual transit times form *together* a realization of a definite probability distribution. This means that as a principle it is possible to tell for each individual cell (as member of a particular population) what the probability is for its $\tau -$ value to be located within some numerical interval. The narrower the interval is the "better" is the population specified. Of course other quantities than $\tau -$ values might be considered; the important thing here is that we leave the classic morphologic point of view for a stochastic aspect. And it is just this that is done in Section 1.1 at the definition of equivalent particles, where it is the particle's motion that is considered – considered in general terms, namely in terms of the notion of stochastic process.

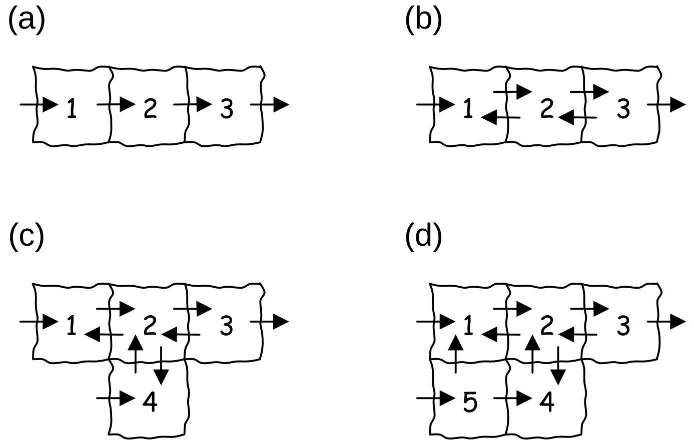

Fig. 4.1.2 The various versions of cell cluster discussed in the text. The basic principle is that the mean sojourn-time θ_S usually increases when the number of cells increases. However, this size of this increase in θ_S depends on how the cells are located geometrically relative each other and, in addition, how they are functionally related (e.g. reversible or irreversible connections). The principle that θ_S increases when the weight W_S of the organism increases remains by and large valid, though as illustrated by Fig. 4.1.3, it is a matter of degree.

The purpose of Fig. 4.1.2 is to improve the intuitive understanding of (4.1.7) and (4.1.8): a well-disciplined intuition is about the best theoretical tool there is. Yes, a more formal treatment would be possible, but it would also be rather messy and intuition would be given less than priority.

With some effort it is seen from Fig. 4.1.2 that in (a) one has $\tilde{P}_1 = \tilde{P}_2 = \tilde{P}_3 = 1$ and $r_1 = r_2 = r_3 = 0$, whereas for (b) one gets the same $\tilde{P}$ but $r_1, r_2, r_3 > 0$. For (c) we have $\tilde{P}_2 = \tilde{P}_3 = 1$ and $\tilde{P}_1 + \tilde{P}_4 = 1$ together with the addition $r_4 > 0$. And finally for (d): $\tilde{P}_2 = \tilde{P}_3 = 1$, $\tilde{P}_1, \tilde{P}_4, \tilde{P}_5 < 1$ and $r_{1,}, r_2, r_3, r_4 > 0$ but $r_5 = 0$.

No, it is not as messy as it looks at a first glance, and I do urge the reader to work through the four cases so that the message (4.1.8) comes through loud and clear. A part of that message is that θ_S depends on how the cells are geometrically located relative each other, and whether or not the connections are reversible.

Different cells have of course different mean transit-time. But cells are organized in specific organs and tissues, within each of which the cells are pretty equivalent with similar τ-values. However, the analysis here is qualitative with no intentions to be dependent of "true" values, like the actual τ-values.

At any rate, even if one feels sure of made conclusions, the general rule is to play it on the safe side. Therefore, from the preceding one can formulate the second part of the model:

> Unless environmental factors and the composition of *S* change drastically, one should expect θ_S to increase with W_S. (4.1.9)

The term "composition" refers to proportions of cell groups like those of muscle, liver, connective tissue, lipids etc.

Let us now try to translate the preceding reasoning into a form that better serves our immediate needs. Thus, for the first part of the model—(4.1.6) that is:

$$\zeta^0_{SE} = aW_S^\alpha \ , \quad a, \alpha > 0 \tag{4.1.10a}$$

And then we do the same thing to the second part of the model—(4.1.9) that is:

$$\theta_S = bW_S^\beta \ , \quad b, \beta > 0 \tag{4.1.10b}$$

The choice of possible functions for these purposes is of course unlimited, and the present functions have been selected for their simplicity and should be considered qualitatively only. That is, no attempt will be made to estimate parameter values; any such attempt would not only be totally pointless but, even worse, would demonstrate utterly bad taste.

Let us begin with (4.1.10a), which is the model for how the uptake of pollutant depends on the organism's weight. As already pointed out one support of this model comes from irreversible thermodynamics – when (4.1.6a) is accepted that is, and it does not look as if we had much choice but to accept. But this does not solve the problem about the value of α, and to that end one can notice that there are some uptake processes that have been rather extensively studied (e.g. the uptake of oxygen as a function of body weight). And a general result seems to be that although the uptake increases with W_S the rate of increase decreases. In other words, if we venture a generalization (and it seems that we are forced to do exactly that) then $\alpha < 1$.

A kind of teleological argument for this conclusion about the value of α is that cells seemingly support each other's life function (like manufacturing "essential" substances for each other). Therefore, new cells being added to *S* – the organism grows – cause increased demand of about almost everything, but this *increase* is proportionally less when *S* is grown up than when the organism is small, because each new cell ("added" to *S* as a result of the growth process) has proportionally lesser cells to support the larger *S* is.

The next concern is the second part of the model expressed by (4.1.10b), namely the properties of β. And to that end let us turn to Fig. 4.1.3, which depicts two extremes regarding simplified aspects (equal cells) on the geometric growth of *S*. Hardly any comments are needed, except perhaps that it seems difficult to find (physical, biologic) arguments for $\beta < 0$, whereas there is a lot that speaks for $\beta > 0$, which is already acknowledged by the way the model (4.1.10) is written. Now, from Fig. 4.1.2 one would intuitively conclude that it is hardly possible to visualize a $\beta > 1$, and there is in fact a rather independent support for this feeling of intuition, and that comes from Fig. 4.1.3. That is, addition of new cells – again, the organism grows – might result in a decrease in some of the $\tilde{P}$ for the older cells (i.e. those cells already present as a part of the organism) and this might lead to a value of θ_S that is less than one would otherwise expect. In short, it is not difficult to accept $\beta < 1$.

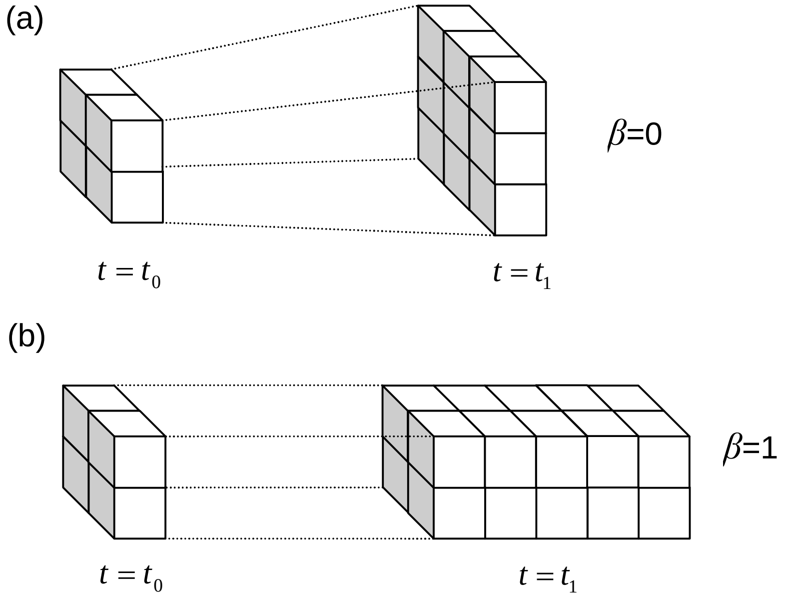

Fig. 4.1.3 Simplified principles of growth of organism: (a) a mollusk, (b) a shark. It is evident from the mere look of the two figures that θ_S *increases faster for shark than for mollusk. In fact, for (a) we have* $\beta = 0$, *which means that, according to the model (4.4.10b),* θ_S *stays constant when* W_S *increases; well, the figure is indeed* simplified, *so some increase in* θ_S *is to be expected also when* S *is a mollusk. And something similar holds for (b), but in the other direction as it were: although* $\beta = 1$ *we could hardly expect such "efficiency" in reality so, as discussed in the text, we should expect* $\beta < 1$.

Now, from (4.1.4) and (4.1.10) one gets

$$w^0 = abW_S^{(\alpha+\beta-1)} \tag{4.1.11}$$

and from Fig. 4.1.1 it seems reasonable to expect that for fish one has $\alpha+\beta>1$. That is, it is indeed quite possible for w^0 to increase with W_S, and that is exactly what empirical data for fish show.

But $\alpha,\beta<1$ and it is thus possible that the "margin" is small for the result we have just obtained. That is, although the exponent in the right hand side of (4.1.11) is positive for fish it could be rather small. In other words, it is possible that just a seemingly minor decrease in either α or β, or both, could change the picture totally: suddenly the weight concentration would decrease with increase in body weight, as it does for mollusk. And if we consult Fig. 4.1.3 we can see that this is actually not only possible but quite possible. Thus, a primitive organism like mollusk has no circulatory system and the cells must be in more or less direct contact with the environment. That is, as Fig. 4.1.3 illustrates, there is a difference in the principle of growth of fish (b) and of mollusk (a): one should expect that the β-value for mollusk is smaller than for fish. And this makes quite understandable indeed the empirically recorded difference between fish and mollusk: as stated by (4.1.3), w^0 for mollusk decreases with increasing body weight whereas the opposite holds for shark.

In terms of DMP we have thus obtained an understanding of why the departure from classic physics discussed in Fig. 4.1.1. It is in fact almost inviting to characterize classic physics as something rather particular, namely the universe where the equality $\alpha+\beta=1$ holds (the weight concentration does not change with body weight). Actually, from this point of view the empirically observed departures come out as something more "natural" than the property provided by the classic perspective. Perhaps it would be sensible to speak of three different kinds of physics, namely: $\alpha+\beta<1 \ \ (molusc), or \ \alpha+\beta=1 \ \ (classic), or \ \alpha+\beta>1 \ \ (shark)$?

Of course it is not a matter of different physics because, as we have just seen, there are no violations of any basic physical principle but, on the other hand, it demonstrates how easily biologic processes can *look* as if there were such violations. No wonder that in the past investigators have occasionally believed that sometimes biology violates basic physics.

At any rate, it is crucial to note that the analysis in this section has been qualitative, in the sense that no attempt has been made to estimate the various parameters . This kind of analysis has proved useful also in other contexts, and the topic is further discussed in Appendix 7.

4.2 Turnover factors and the correspondence principle

In this section I shall introduce a generalized rate constant, where "rate constant" stands for the well-known quantity of classic chemistry (e.g. the quantity where the temperature dependency is given by the Arrhenius equation); a result of the analysis will be a concept applicable to systems where the classic mixing condition is not necessary (heterogeneity on all levels). My intention is to demonstrate the use of the correspondence principle (Appendix 8); hopefully this will also show how probability theory can serve as a substitute for "physical mechanisms" (Section 0.3)

In fact, the whole approach I apply in this book is based on the idea that this kind of substitute for "physical mechanism" is valid. As discussed in Section 0.3, the very idea is that a particular gap between physics and biology can be bridged simply by replacing physical mechanisms with the laws of probability.

Fig. 4.2.1 is supposed to depict an open S—in steady state!— which is subdivided into two mutually exclusive subsystems S_1 and S_2. Between these subsystems, and between them and the environment *E,* there are flows of particles of some substance *U*, as indicated by the figure (where each symbol ζ^0 denotes "sum" of sub flows). Introduce a good mixing device in the system and one obtains a counterpart in the form of closed system dealt with in classic chemistry.

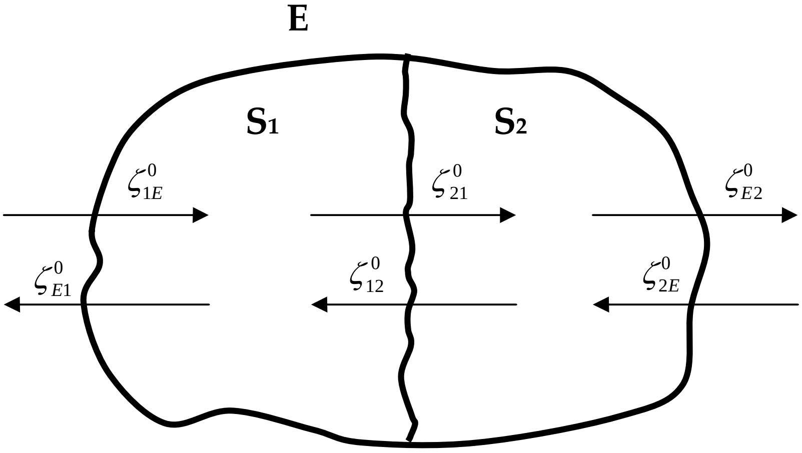

Fig. 4.2.1 An open system in steady state and subdivided into two subsystems: the flow symbols represent "sums" of sub flows. There is a corresponding closed system (see the text) but then $\zeta^0_{iE} = \zeta^0_{Ei} = 0$*, and the steady state condition is replaced by equilibrium. The subsystems can of course be "non geometrical" with respect to* S *(e.g. in the sense of representing 'free' and 'bound'* u*).*

We thus have here an open system that is in steady state and of arbitrary complexity (heterogeneity). The subdivision means that we have subsets Q_i ($i = 1$ or 2 , or S), so if q_i denotes a state (point) in Q_i, the symbol $\pi_j(\Delta t, q_i)$ stands for the probability of a u in the state $q_i \in Q_i$ ($i \neq j$) to move into S_j during the time $\Delta t > 0$.

The next step is then to introduce the *ad hoc postulate*

$$\pi_j(\Delta t, q_i) = \lambda_j(q_i)\Delta t + o(\Delta t) \ , \ i \neq j \tag{4.2.1}$$

where we shall accept the possibility of $j = E$; that is, $\pi_E(\Delta t, q_i)$ is the probability of a u in the state q_i (the particle is in S_i) to move into E (to leave S that is) within the time Δt. It is important here to keep in mind that the system is stationary (steady state), because otherwise we might have had to consider $\lambda_j(q_i)$ as a function of the time t.

The term **ad hoc postulate** that I here introduce should be understood in the following way: The language (DMP) is given with, among other things, the postulates listed in Chapter 2. And a language often grows in the connection with applications. But perhaps it has not been made sufficiently clear that each application frequently induces the need for new concepts *specific* for that particular application, which occasionally may require additional postulates, and I have chosen to call them ad hoc postulates, because they are specific for a specific application. That such a need sometimes occurs is not more remarkable than, say, that the scales on coordinate axes often have to be adjusted in regard to the specific application at hand – the coordinate system as such remains the same.

This postulate defines the **turnover intensity** $\lambda_j(q_i)$, which we shall apply for the definition of the **turnover function**

$$\overline{\lambda}_{ji}(t) \equiv \int_{Q_i} \lambda_j(q_i) \frac{\rho(t, dq_i)}{P_i(t)} \ , \ i \neq j \tag{4.2.2}$$

i= 1 or 2 or S, and $j = 1$ or 2 or E. That is, the turnover function is the local mean value of the turnover intensity (Appendix 4). Perhaps it should be emphasized that the system is stationary (time invariant, that is) and the time symbol t here stands for the time elapsed since the particle $u \in U_S$ entered the system S (cf. Section 2.6).

If we select S_1 as our "home", standard probability algebra gives

$$P_1(t + \Delta t) = \left[1 - \overline{\lambda}_{01}(t)\Delta t\right]P_1(t) + \overline{\lambda}_{12}(t)\Delta t \, P_2(t) + o(\Delta t) \ , \ i \neq j \tag{4.2.3a}$$

where

$$\vec{\lambda}_{01}(t) \equiv \overline{\lambda}_{E1}(t) + \overline{\lambda}_{21}(t) \tag{4.2.3b}$$

So the inside of the brackets is the expected probability of a particle in S_1 for not leaving that subsystem within the time Δt. If we let $\Delta t \to 0$ the result is the first order differential equation

$$\dot{P}_1(t) = -\bar{\lambda}_{01}(t)P_1(t) + \bar{\lambda}_{12}(t)P_2(t)\ ,\ i \neq j \tag{4.2.4}$$

One might say that this equation is a consequence of the ad hoc postulate (4.2.1), which thus "warrants" the existence of the first time-derivative of $P_i(t)$. In other words, the postulate is not all that trivial.

A particular matter to observe here is that the symbol *t* denotes the time the particle has been in the system (it enters at the time *t*=0). Yes, the system is assumed to be in steady state from the very beginning so to speak; this is an important condition for the below presented analysis of the equation.

As it stands (4.2.4) looks indeed impossible to handle: not only are the coefficients non-constant but, even worse, they are *unknown* functions of the time *t*; again it should be emphasized that *t* here denotes the time elapsed since the particle $u \in U_S$ (particle to which (4.2.4) applies) has entered the system. But, however ugly the equation (4.2.4) is, here we actually have a situation where some of the results of the preceding work (the trouble) will come to explicit use, because it will in fact be possible to handle (4.2.4) in an effective way (even if the equation is not "solved" in classic sense). Thus, should all terms in (4.2.4) be integrated from zero to infinity, the Theorem A4.1.2 gives:

$$-P_1(0) = -\bar{\lambda}^0_{01}\theta_1 + \bar{\lambda}^0_{12}\theta_2 \tag{4.2.5}$$

because, as *S* is open, we have $\lim_{t\to\infty} P_1(t) = 0$. The λ values are referred to as **turnover factors** (the turnover functions with respect to the steady state distribution).

The quantity $P_1(0)$ can here be interpreted as the proportion of the total flux ζ^0_{SE} that goes directly to S_1 from *E*:

$$\zeta^0_{1E} = P_1(0)\zeta^0_{SE} \tag{4.2.6}$$

Hence, if we multiply all members of (4.2.5) by ζ^0_{SE} we obtain from the mass-time equivalence (Theorem 3.4.1):

$$\zeta^0_{1E} = -\bar{\lambda}^0_{01}\,\mu^0_1 + \bar{\lambda}^0_{12}\,\mu^0_2 \tag{4.2.7a}$$

where it follows from (4.5.3b) that

$$\overline{\lambda}_{01}^{0} = \overline{\lambda}_{E1}^{0} + \overline{\lambda}_{21}^{0} \tag{4.2.7b}$$

Similar to what has been stated about Theorem 3.4.1, the linear look of this equation can indeed be misleading. That is, there are no reasons for us to expect any linearity what-so-ever: a deliberate change in any of the quantities of the equation is likely to induce changes in the other quantities.

A rather common as well as generally known form of departure from linearity might be exemplified by the situation where the process $u \in U_1 \leftrightarrow u \in U_2$ is the reversible biochemical reaction: let first

$$U_1 \equiv \textit{population of free } u \in U_S$$

and

$$U_2 \equiv \textit{population of } u \in U_S \textit{ bound to an enzyme molecule}$$

which could be expressed as: in *S*, *u* moves between the two states "free" and "bound to an enzyme molecule". If the number of enzyme molecules is small compared to the number of free *u* molecules a "competition" will take place: if μ_1 increases λ_{21}^{0} will decrease (the probability for a *u* in S_1 to collide with a *free* enzyme molecule decreases); hence, this turnover factor will depend on the content of the corresponding set Q_1, which means that in (4.2.7) we would have $\lambda_{21}^{0} = f\left(\mu_1^{0}\right)$—a decreasing function of μ_1^{0}—and certainly that is a departure from linearity of (4.2.7a).

One might view the two equations (4.2.4) and (4.2.7) as different aspects on the same thing, namely on a steady-state system according to Fig. 4.2.1. That is, the equation (4.2.7) describes the population aspect (the collective view so to speak), whereas the equation (4.2.4) gives the expected behaviour of the individual particle.

Yes, the two equations are physically speaking quite trivial indeed (conservation of mass in classic physics). What however is important is that they have been *formally derived* in plain probabilistic terms, which demonstrates equivalence between the laws of probability and the laws of nature (physics).This equivalence is further demonstrated by application of (4.2.7) to a closed system. That is, if the system is closed the flow ζ_{1E}^{0} is zero and, consequently,

$$\frac{\mu_1^{e}}{\mu_2^{e}} = \frac{\overline{\lambda}_{12}^{e}}{\overline{\lambda}_{21}^{e}} \tag{4.2.8}$$

Of course, the super script "*o*" is replaced by "*e*", because steady state is for closed system as meaningful as is equilibrium for open system. The formal transition from open to closed system is however a rather involved thing, which would require some space that I do not think is warranted in present context and, therefore, I abstain from going deeper into the matter.

Every one familiar with classic chemical kinetics recognizes the right hand member (the ratio between two turnover factors) as an expression for the old familiar “equilibrium constant”. That is, we have applied DMP on a classic type of system and obtained a direct correspondence with classic theory.

What the present derivation adds to the classic picture is that the chemical rate constant comes *automatically* out as a local mean value (Appendix 4) when the vector r in (1.1.1) is removed (because of the classic system being well stirred the geometric state vector in (1.1.1) is of no significance); if the non-geometric state vector ω is identified with the particle’s (molecule’s) various energies, one has the old classic theory. That is, apart from the correspondence between the classic theory and DMP, we get a general interpretation of the notion of rate constant as a turnover factor, namely a local mean value of turnover intensity—when steady state (alternatively equilibrium) is at hand.

Even if the turnover factor appears as a generalized rate constant (meaningful for a large number of physically quite different processes, under a variety of conditions—occasionally far away from classic chemistry) the generality is of course not total. That is, there is a definite restriction, and that restriction spells stationarity (steady-state or equilibrium). Presently there seems to be no escape from that constraint. Thus, although there are no conceptual problems with the notion of turnover functions as introduced above, the fact that there presently does not seem to exist any general means to specify the forms of these functions certainly hamper their practical usefulness.

By direct use of Theorem (A4.1.2) and of the mass-time equivalency we have thus been able to transform the impossible looking (4.2.4) into the substantially more reasonable looking (4.2.7). However, it cannot be overemphasized that in physical terms this (4.2.7) is rather trivial; what is important is that the equation has been given a purely probabilistic derivation. And a notable feature is then that the equation is not expressed in terms of volume concentrations but in total amounts: the total amount of substance (“content”, section 1.5) is more generally a meaningful quantity (e.g. observable) than is the classic volume-concentration. But the equation does something more, and in my opinion something rather important, namely that it demonstrates the experimental difficulties that are involved even in seemingly quite simple situations. Thus, the equation contains five quantities and, of course, four of them must be estimated at practical applications. Perhaps the number of quantities could be lessened by the use of various indices, but presently the field is open.

It is possible to apply a slightly different aspect on the above analysis, because one can namely apply the notion of theorem introduced in section 0.2. Thus, it is obvious that (4.2.4) is actually a theorem (no model has been introduced: nothing to test) and, as (4.2.7) is a result of some other theorems being applied on (4.2.4), also (4.2.7) is a theorem. Hence, we have been dealing with a collection of theorems: the above introduction of turnover factor and the analysis in general have been a matter of language—again, no particular model has been introduced. From the above established correspondence between DMP and classic chemical kinetics it follows that this classic thing is actually a collection of theorems—a matter of language. This differs deeply from the original impression I received when I as young student first learned of equilibrium constant and rate constant in chemistry: to me it was clear that it was a matter of fundamental principles of nature, and the fact that data were such that the constants could be estimated from them verified this understanding of mine—I thought. That it by and large was a matter of language never crossed my mind. Yes I was young and inexperienced at that time.

Appendix 1
On the difference between Sojourn-Time and Transit-Time

I think that the term "mass-time equivalence" is pretty self explaining, namely that it concerns a direct and simple relation between mass quantity μ_v^0 and a time quantity θ_v; the symbol ζ_{SE}^0 could be looked upon as a scale factor, somewhat like c^2 in the famous $E = mc^2$ (well, in the mass-time equivalence the scale factor is not absolute though variable, but why not). Of course, other derivations and other terms for the mass-time equivalence have appeared in the literature.

Yes, in the literature there are a number of various terminologies and derivations of the mass-time equivalence. An example is the derivation that I gave in 1964, where method and terminology differ from the one used here. What to me appears to be good with the reasoning in Chapter 3 is that the physics is rather explicit as it were; e.g. the reasoning makes clear the possibility of mutual interaction between particles, with non-linearity of the equation as a consequence.

One term that I personally like very much is *occupancy principle*, where then θ_v is called *occupancy* instead of mean sojourn-time. This terminology is due to Orr and Gillespie who in 1971 published a derivation of the equation that, though perhaps superficial, is illustrative and provides a kind of intuitive understanding of the matter. However, in general I do feel that the terminology should be descriptive, as far as this is possible, and to me the term "occupancy" is certainly not descriptive; in particular I have in mind the symmetry in the *distinction between mean transit-time* and *mean sojourn-time*.

Yes, there is indeed a good reason for making that distinction, and that reason spells mass-time equivalence (what thus Orr and Gillespie call the occupancy principle); that is, the equation (3.3.6) or (3.4.1) in Theorem 3.4.1:

$$\mu_{\nu}^{0}=\zeta_{SE}^{0}\theta_{\nu}$$

What should be noticed here is that the equation holds as a principle for an arbitrary subsystem $S_{\nu}\subseteq S$. A reason for why this is noteworthy is that Stephenson (1960) published a similar equation but only for (μ_{S}^{0},θ_{S})—for the whole *S* that is—and instead of sojourn-time he uses a terminology corresponding to "transit" time. And the same holds for Takács who published, the same year (Takács, 1960, probl. 8, page 39), a similar equation for the number of electrons between the cathode and the anode in an electronic vacuum tube (the particles are emitted by the cathode and absorbed by the anode), and there he uses the term "time of flight" of the electrons, which has the same meaning as "transit-time" (the process *electron caught by anode* is namely irreversible). What makes this particular derivation especially interesting is that he (Takács) looks at the input ζ_{SE}^{0} as a Poison process and obtains not a deterministic value like the mean value μ_{S}^{0} but a Poison distribution for number of particles (electrons) between the cathode and the anode when steady state is achieved. Thus, also here the analysis is restricted to the whole system *S*, namely the "space and force field" between the cathode and the anode. Thus, also in Takács' analysis a notion corresponding to transit-time is used and only the whole system can be considered (not an arbitrary subsystem as in the mass-time equivalence).

In present formalism the Stephenson (Takács) equation assumes the form

$$\mu_{S}^{o}=\zeta_{SE}^{o}\theta_{S}$$

without the notion of steady state however. It is of some interest to note that in the theory of queuing this equation has occasionally been known as Little's theorem (1961), and earlier it was in the literature on physiology known as one of the Stewart-Hamilton equations. In other words, the kind of science we presently deal with is rather compartmentalized with hardly any communications. Yes, that's sad.

The matter I thus want to stress is that the concept of transit-time must be replaced by the concept of sojourn-time in order to get the equation in the Theorem 3.4.1 for an *arbitrary* subsystem $S_{\nu}\subseteq S$ (again, not *only* the whole system *S*). In other words, the difference between the two time concepts transit-time and sojourn-time is not only formal, as the discussion in Chapter 3 might suggest, but also necessary when one wants to consider an arbitrary subsystem.

But I must admit that to me the meaning of the word "sojourn" is not as evident and precise as is that of "transit". For instance, instead of sojourn-time could one not have used spending-time? Or visiting-time? Or...?

There is certainly no objective answer to these questions. The simple fact is that I have taken the term “sojourn” directly from Khinchin (1960, p.69), or perhaps rather the translator, who uses the word sojourn with the same meaning as I do. It was not until I found that term that I realized that there are these two kinds of time quantities, and that there is a formal as well as a necessary distinction between them.

Appendix 2
Equilibrium

The basic principle behind the notion of classic equilibrium is **microscopic reversibility**. To get a glimpse of the meaning of this principle, let us consider a system (*S*) that is isolated in thermodynamic sense: there is no exchange of either matter or energy between (*S*) and *E*. Let us further assume that the system (*S*) consists of a population of particles $u \in U_S$ that are moving between the three states A, B and C. If (*S*) is left alone for a sufficiently long time the basic idea is that it reaches the systemic state of equilibrium, and two different situations appear to be natural possibilities as shown by Fig. A2.1 and Fig. A2.2.

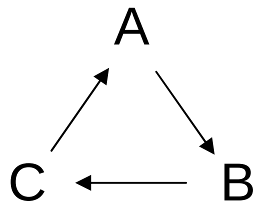

*Fig. A2.1 Let (*S*) be a thermodynamically isolated system where u can be in three different states A, B and C. At equilibrium a cyclic situation is conceptually possible but not so physically speaking: a cyclic situation means an ongoing entropy production, and that is impossible because the system is isolated (no exchange of either matter or energy between (*S*) and *E*). However, when the system is not isolated the situation changes fundamentally, and the very notion of equilibrium changes meaning in quite a basic fashion.*

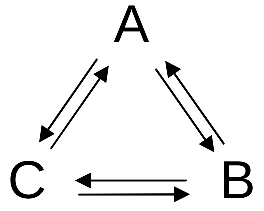

Fig. A2.2 The same situation as in Fig.A2.1, but here we have microscopic reversibility, *which means that at equilibrium (the system is isolated) there are no net-flows between the states. This is accepted as the only physical possibility of an isolated system in equilibrium and, in fact, this condition has been of fundamental importance for the development of thermodynamics in classic chemistry. Also at the development of basic parts of irreversible thermodynamics the notion of microscopic reversibility has been essential. However, as hinted in Fig.A2.1, the condition of microscopic reversibility loses its fundamental status as soon as the system is no longer isolated (e.g. when the system is merely closed in DMP sense; that is, the system could be closed to one substance but open to other substances; and indeed there might be exchange of energy between the system and its environment.*

The fundamental principle then is that Fig. A2.1 is physically impossible: it is Fig. A2.2 that holds in the sense that there are no net-flows between the states. That is, all processes are reversible—microscopically reversible that is. Or formulated differently, Fig. A2.1 implicates an entropy production, and that is not possible because (*S*) is isolated: there is no exchange of any energy between (*S*) and *E*. This principle of microscopic reversibility has had an enormous influence in both older and younger chemistry and, more recently, in the development of irreversible thermodynamics—especially at the derivation of the Onsager reciprocity relation.

It is probably correct to say that the development of irreversible thermodynamics began to catch momentum around 1930, and a major forward step was the work of Onsager. What to me seems to be particularly important is how this work paved for an understanding of the coupling between different transport mechanisms; e.g. it became gradually clear why some processes go in the converse direction to the one expected on physical grounds, when the processes were considered alone; for instance, it became understandable why one can in nature find diffusion against the concentration gradient. And much of all this was based on the principle of microscopic reversibility.

And here is the problem: In DMP a system *S* is classified (open or closed) with respect to one (Sic!) substance *U*. This means that a closed system is closed with respect to the selected substance *U* , and quite likely there are other substances involved and some of them may be just passing through (i.e. with respect to those substances the system is open). And above all, such a closed system is usually by no means isolated: commonly *there are fluxes of energy through the system,* and the principle of microscopic reversibility can by no means be taken for granted. In fact, Morowitz showed in 1967 that a flow of energy through a multi-component processing system S is bound to induce cyclic processes according to the principle depicted by Fig.A2.1 (e.g. the flow through Earth of energy from the Sun should induce some basic organization as it were). Thus, the equilibrium distribution D^e defined in Section 3.5 should usually have little to do with the thermodynamic notion of equilibrium.

Well, there is at least one commonness as it seems between the two kinds of equilibrium, and that should be the first ergodic principle discussed in Appendix 3. That is, the thermodynamic equilibrium distribution and D^e are both independent of initial condition γ, provided the two systems are fully connected.

Appendix 3
On Ergodicity

The purpose of this appendix is to demonstrate the fundamental difference between open and closed systems in terms of the so-called ergodic theory. Thus, as stated in Section 0.1, whereas in classic theory of "closed" systems a particle's trajectory can be made arbitrarily long, this is not possible for open systems where the trajectories as a principle are finite (according to the definition of openness in Section 3.2). When dealing with open systems one must replace the trajectory's length with the size of a family of (finite) trajectories; each trajectory corresponding to an individual particle (rather, the trajectory is the thought result of the individual particle's motion).

A3.1 On the ergodicity concept

The term ergodic originates from the work of Bolzmann on statistical mechanics, in the last part of nineteen's centaury: an isolated system is said to be ergodic if it eventually passes through all microscopic states allowed for by the macroscopic conditions. To illustrate what that means—in terms of classic statistical mechanics—we select a system consisting of a closed container that harbors N molecules of a gas at constant pressure and temperature. We assume that the system is isolated, so there is exchange of neither matter nor energy with the environment.

In classic statistical mechanics such a system is represented as a point v in a $6N$-dimensionsal **phase space** (a kind of state space) that we denote V, where each molecule is represented by three Euclidian-space coordinates and three momentum coordinates. Each point in that space corresponds to the system's microscopic state, and we let the symbol v denote the corresponding location vector. The function $v = v(t)$, which reflects the change of the system's microscopic state with time, is determined by the Hamiltonian principles, and it generates a trajectory on the "surface" V_A corresponding to the system's constant energy A. Bolzmann's ergodicity concept states that, when given sufficient time (for a molecular system this is an "instant" to a human), $v(t)$ will eventually pass through all points on V_A that are possible for the given pressure and temperature. Here v is the microscopic state of the system, and pressure and temperature illustrate the meaning of the macroscopic conditions referred to above.

Let us now assume the existence of a probability density $\vartheta(t,v)$ on V_A; we know that, by physical necessity, $v(t) \in V_A$ for all values of t (the system's energy A is constant). By thermal (or thermodynamic) equilibrium is understood the situation where

$$\vartheta(t,v) \equiv \vartheta^e(v) \tag{A3.1.1}$$

for all $v \in V_A$. If we define, with respect to the equilibrium distribution, the **space** (or **phase**) **average** of a semantic state function $f(v)$ as (cf. Postulate III)

$$\bar{f}_A^e \equiv \int_{V_A} f(v)\vartheta^e(dv) \tag{A3.1.2}$$

it can be demonstrated that Bolzmann's ergodic hypothesis implies that

$$\bar{f}_A^e = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(v(t))dt \tag{A3.1.3a}$$

The right hand member is called the **time average** (along the single trajectory $v = v(t)$) and the equation can be expressed in words:

Space average equals time average (A3.1.3b)

The reason why this is important is that the theory (statistical mechanics) deals with space averages whereas macroscopic measurements are interpreted as time averages (at least this appears to be the natural thing to do). Hence, relations like (A3.1.3) connect theory with experimental data.

In more general terms $v(t)$ may be referred to as a time process, and if (A3.1.3) holds *this process is said to be ergodic,* without reference to Bolzmann's concept. This habit reflects a common phenomenon, namely how the meaning of words change, by accident or for rational reasons or as a result of the natural linguistic "progress". As for this specific case one can indeed speak of accident and a tragic one. Thus, because he felt unappreciated, and even laughed at, Bolzmann committed suicide just a few years before his work became commonly recognized, and it was therefore up to others to describe and interpret his work. A major contribution in this regard was made by the two brothers Ehrenfest who, as it seems, made the slight error to put the label ergodicity on (A3.1.3). It is since then in solid tradition of physics (and also in, say, statistic time-series analysis) to use that terminology: the process is said to be ergodic if the space average equals the time average.

Other, mathematically equivalent expressions of (A3.1.3) are possible, and for present purposes an important such an expression is:

> If S is closed and fully connected, the ergodic principle—in the form of (A3.1.3a) and (A3.1.3b)—holds if and only if the equilibrium distribution $\left\{\vartheta^e(v);\, v \in V_A\right\}$ is independent of the initial condition $v(0)$. (A3.1.3c)

Hence, the three expressions (A3.1.3) are formally equivalent, and are formulations of one and the same ergodic principle.

That seemingly different formal expressions or aspects can be applied on one and the same idea is no problem—on the contrary it shows that one deals with a logically rich concept. And although an expression like (A3.1.3) is deterministic, it is quite understandable that the principle those equations express has been adopted in the theory of stochastic processes, but then the convergence must be expressed in probabilistic terms: the integral in (A3.1.3) is a stochastic integral like (3.1.1), and instead of (A3.1.3a) one has to deal with convergence in probability:

$$\lim_{T\to\infty} P\left\{\left|\frac{\int_0^T \tilde{f}(t)dt}{T} - \bar{f}_A^e\right| > \varepsilon\right\} = 0 \qquad \text{(A3.1.4)}$$

where $\tilde{f}(t) \equiv f(v(t))$.

It is of some importance to realize that $\tilde{f}(t)$ is a stochastic function or process (i.e. for fixed value of t it is an ordinary stochastic variable, like the indicator function in (3.1.1)); underlying all this is that $v(t)$ is looked upon as a sample function of a stochastic process not governed by Hamiltonian principles but by, say, transition probabilities. For comparison purposes we shall use another form of the above equation, by making the substitution

$$\beta(T) \equiv \int_0^T \tilde{f}(t)dt \tag{A3.1.5}$$

which means that (A3.1.4) can be written on the form

$$\lim_{T\to\infty} P\left\{\left|\frac{\beta(T)}{T} - \bar{f}_A^e\right| > \varepsilon\right\} = 0 \tag{A3.1.6}$$

where it is important to keep in mind that $\beta(T)$ is a function of a stochastic time process $v(t)$ or, rather, it is an integral along the trajectory $v = v(t)$ over the time interval $[0,T)$, and this trajectory in turn is a realization of a stochastic process.

In this form the idea of ergodicity becomes easier to handle. Well it is not just "easier to" but in fact "possible to". The reason for this is that Bolzmann's original idea turns out to be an impossibility (physically, logically), very much because his analysis is deterministic; yes, even if terms like "probability" and "average" appear here and there in the reasoning, the Bolzmann analysis is deterministic!

A3.2 Ergodicity of open *S*

Let us return to the situation initially considered in Section 3.3, and for simplicity we assume that the model is valid (from Section 3.4 we know that some of the model's problems can be lessened by a steady-state condition). But instead of building up a steady state by a regular input regimen (ζ_{SE}^{0}, as done in Section 3.3) we now consider the input of the total number of *N* particles over the finite time period $[0, M]$ (that is, $M < \infty$). This means that we consider a supply of a finite number of particles over a finite period of time; a pedestrian formulation could be "a single dose of *N* particles prolonged over a finite time interval". For the content of S_v at the time *t* we get

$$X_v(t) = \sum_{j=1}^{N} I_{vj}(t) \tag{A3.2.1}$$

where $I_{vj}(t)$ is the indicator function for particle number j in the dose of N particles. Obviously (3.3.2) is not applicable because we don't know where in the time interval $[0, M]$ the individual particle in the dose of *N* particles enters *S*. But we however do know that the stochastic integral

$$\int_0^\infty I_{\nu j}(t)dt$$

exists and is finite for $j = 1,2,\ldots,N$.

Perhaps it should be explained that because the integral is from zero to INFINITY its value is independent of when the particle entered S—as long as the particle's entrance took place within a FINITE time interval. Of course, essential to this context is the notion of stationarity initially introduced in Section 2.1 and accepted by (2.2.1).

Hence, if we make the definition

$$\tilde{\eta}_\nu(N) \equiv \int_0^\infty X_\nu(t)dt \tag{A3.2.2}$$

We get from (A3.2.1)

$$\tilde{\eta}_\nu(N) = \sum_{j=1}^{N} \int_0^\infty I_{\nu j}(t)dt \tag{A3.2.3}$$

This shows clearly how $\tilde{\eta}_\nu(N)$ is a sum of N mutually independent stochastic variables with the same mean value θ_ν. Therefore,

$$\frac{\langle \tilde{\eta}_\nu(N) \rangle}{N} = \theta_\nu \tag{A3.2.4}$$

for all values of $N > 0$. Hence, from what occasionally is called the "law of large numbers"

$$\lim_{N \to \infty} P\left\{ \left| \frac{\tilde{\eta}_\nu(N)}{N} - \theta_\nu \right| > \varepsilon \right\} = 0 \tag{A3.2.5}$$

And from (3.5.7) we thus get the following

THEOREM A3.2.1:

If *S* is in steady state

$$\lim_{N\to\infty} P\left\{\left|\frac{\tilde{\eta}_\nu(N)}{\theta_S N} - P_\nu^0\right| > \varepsilon\right\} = 0$$

where the somewhat annoying θ_S should be looked upon as a scale factor.

A comparison of this theorem with (A3.1.6) shows clearly 1) the similarity between closed and open system; 2) the fundamental difference between the two systems, which I indicated already in Section 0.1: for closed *S* the concern is how large the time *T* is, which corresponds to how long the particle-trajectory is, and when it comes to open system the concern is about how large *N* is, namely how large the population of (finite) trajectories is.

In summary then, when it comes to understand Theorem A3.2.1 as such, there are three different aspects that appear possible. Thus, one of those aspects stresses the space-time perspective. That is, $\tilde{\eta}_\nu(N)/N$ stands for time component whereas P_ν^0 represents space (cf. (A3.1.3b) above). And it is this particular point of view that gives to the theorem an ergodic character.

Another aspect is that the theorem establishes a relation between a stationary and a transient population of particles. Of course, although being stationary the steady state population is on constant move so to speak: particles are all the time leaving and being replaced by new particles.

A third aspect is that the theorem gives a connection between two different families of realizations of the same stochastic process. The difference between the families is in the time schedules according to which the families are generated. One schedule is quite strict and ongoing and results in steady state, whereas the other schedule is arbitrary and transient – it takes place within a finite time interval – with the result of a transient particle population.

Finally one can observe that question (ii) in Section 0.1 is given an answer of a sort by Theorem A3.2.1, though the answer is not for a single trajectory (Fig. 0.1.2) but for a *population* of transient trajectories.

A3.3 Another ergodic principle

From (A3.1.3c) it follows that if the system is closed and well-connected, ergodicity means that

$$\bar{f}_A^e \text{ is independent of the initial condition } v(0) \qquad \text{(A3.3.1)}$$

It is of some importance that this principle has never been proved under general conditions: one can occasionally find it referred to as "the ergodic hypothesis". On the other hand, there are quite a number of mathematical and physical indications (circumstantial evidences) that this ergodic principle holds under pretty general conditions. It seems to me rather safe to accept it, although it means a departure from formal cautiousness—but so what.

Yes, I do admit that I have problems here: Theorem 3.4.1 should perhaps not be considered as a theorem. This is in fact indicated in Section 3.4 but then from an entirely different point of view (the main concern there is the notion of steady-state). However, I do feel that the theorem is so well founded that it would be over-kill to doubt it on formal grounds. On the whole there is a wealth of formal subtleties involved in the ergodic matter as such, and I do intend to avoid them and, therefore, apply a slightly heuristic attitude.

From (2.3.1) and (2.3.2) we get

$$P_v(t) = \frac{\mu_v(t)}{\mu_S(0)} \qquad \text{(A3.3.2)}$$

Of course, this is nothing but the classic definition of probability (again, white and black balls in an urn).

At any rate, the limit distribution D^l is given by

$$P_v^l \equiv \lim_{t\to\infty} \frac{\mu_v(t)}{\mu_S(t)} \qquad \text{(A3.3.3)}$$

and if we apply the discrete model of Appendix 5 it is clear that, when S is fully connected,

$$D^l \text{ is independent of the initial condition } \gamma. \qquad \text{(A3.3.4)}$$

This indeed confirms (3.5.8) because D^0 is, as we know, quite dependent on the initial condition γ. So again we are shown why the dilution method is based on a totally erroneous idea, namely that there should generally be equality between the limit distribution and the steady-state distribution.

Should it now be that S is not only fully connected but closed as well, then

$$P_\nu^e = \lim_{t\to\infty} P_\nu(t) \qquad \text{(A3.3.5)}$$

This follows directly from (A3.3.3) because $\mu_S(t)$ is positive constant when the system is closed. Hence, from (A3.3.4) we have that, for the equilibrium distribution (again, the system is assumed to be fully connected)

$$D^e \text{ is independent of } \gamma \qquad \text{(A3.3.6)}$$

From an historic point of view (A3.3.6) expresses, according to (A3.3.1), an ergodic principle and it seems natural to refer to it as the **first ergodic principle** that, in the form of (A3.3.4), also has an application to open systems. And from the same point of view, it seems reasonable to refer to Theorem A3.2.1 as the **second ergodic principle** that is applicable to open systems only. So if one continues the drivel from the end of Section 3.5, there is only one ergodic principle for closed systems whereas there are two such principles for open systems and, hence, the open system is also in this regard richer than is the closed system.

Appendix 4
The Mean-Value Theorems

Let S be in steady state and let $f(q)$ be a semantic state function on $Q_\nu \subseteq Q_S$. According to Postulate III we can then define the **systemic mean-values** over Q_ν:

$$\hat{f}_\nu(t) \equiv \int_{Q_\nu} f(q)\rho(t,dq), \qquad \hat{f}_\nu^0 \equiv \int_{Q_\nu} f(q)\rho^0(dq) \tag{A4.1.1}$$

THEOREM A4.1.1:

$$\hat{f}_\nu^0 = \frac{1}{\theta_S}\int_0^\infty \hat{f}_\nu(t)dt \tag{A4.1.2}$$

Proof:

$$\int_0^\infty \hat{f}_\nu(t)dt = \int_0^\infty dt \int_{Q_\nu} f(q)\rho(t,dq) = \int_{Q_\nu} f(q)\int_0^\infty \rho(t,dq)dt \tag{A4.1.3a}$$

From Postulate II and (3.1.4) we have

$$\theta_\nu = \int_0^\infty dt \int_{Q_\nu} \rho(t,dq) = \int_{Q_\nu}\int_0^\infty \rho(t,dq)dt \tag{A4.1.3b}$$

To simplify the notation let us introduce the measure

$$r(q) \equiv \int_0^\infty \rho(t,q)dt \tag{A4.1.3c}$$

that is,

$$\theta_i = \int_{Q_i} r(dq), \quad Q_i \subseteq Q_\nu \tag{A4.1.3d}$$

Consequently, from (A4.1.3a)

$$\int_0^\infty \hat{f}_\nu(t)dt = \int_{Q_\nu} f(q)r(dq) \tag{A4.1.3e}$$

But the equality in (A4.1.3d) holds for all Q_i in Q_ν and therefore we get from (3.5.7):

$$r(q) = \theta_S\, \rho^0(q) \tag{A4.1.3f}$$

That is, according to (A4.1.3e) and definition (A4.1.1):

$$\int_0^\infty \hat{f}_\nu(t)dt = \theta_S \int_{Q_\nu} f(q)\rho^0(dq) = \theta_S\, \hat{f}_\nu^{\,0} \tag{A4.1.3g}$$

QED

And now, let us define the **local mean-values**:

$$\bar{f}_\nu(t) \equiv \int_{Q_\nu} f(q)\frac{\rho(t,dq)}{P_\nu(t)}, \qquad \bar{f}_\nu^{\,0} \equiv \int_{Q_\nu} f(q)\frac{\rho^0(dq)}{P_\nu^0} \tag{A4.1.4}$$

THEOREM A4.1.2:

$$\bar{f}_\nu^{\,0} = \frac{1}{\theta_\nu}\int_0^\infty \bar{f}_\nu(t)P_\nu(t)dt \tag{A4.1.5}$$

Proof:

From Theorem A4.1 above and the definitions of systemic mean-value it follows that

$$\frac{1}{\theta_\nu}\int_0^\infty \bar{f}_\nu(t)P_\nu(t)dt = \frac{1}{\theta_\nu}\int_0^\infty \hat{f}_\nu(t)dt = \frac{\theta_S}{\theta_\nu}\hat{f}_\nu^0 = \frac{1}{P_\nu^0}\hat{f}_\nu^0 \tag{A4.1.6}$$

where the last equality follows from (3.5.7). Finally it only remains to observe the triviality $\hat{f}_\nu^o = P_\nu^0 \bar{f}_\nu^0$.

QED

Theorem A4.1.2 is actually a rather trivial corollary of Theorem A4.1.1, but does have conceptual significance. In fact, the notion of local mean-value is applied in the next appendix (Appendix 5). And the same kind of mean value is used in the Section 4.2 at the definition of turnover factor, which is a kind of generalization of the rate constant in classic chemistry.

Appendix 5
The Discrete Model

A5.1 The description

We begin by considering a finite subdivision of S into a finite number of mutually exclusive subsystems $S_1, S_2, \ldots\ldots, S_n, S_\varepsilon$ such that

$$Q_i \bigcap Q_j = \varnothing,\ i \neq j \tag{A5.1.1}$$

and

$$Q_S = \bigcup_{i=1}^{i=n} Q_i \bigcup Q_\varepsilon \tag{A5.1.2}$$

where Q_ε refers to what perhaps could be named the **residual subsystem**, which is formally given significance in the model formulated a few lines below.

The **mean transition probability** is defined as

$$\pi_{ij}(t-\tau,\tau) \equiv \int_{Q_j} \pi_i(t-\tau,q) \frac{\rho(\tau,dq)}{P_j(\tau)} \tag{A5.1.3}$$

which is the probability of a u in S_j at the time τ to be in S_i at the time $t \geq \tau$. The expression (A5.1.3) is in fact a "local mean value" of the function $\pi_i(t-\tau, q)$ as discussed in Appendix 4.

If we add Δt to t and set $\tau = t$, it is possible to write (2.4.4) on the form

$$P_i(t+\Delta t) = \sum_{j=1}^{n} P_j(t)\pi_{ij}(\Delta t, t) + P_\varepsilon(t)\pi_{i\varepsilon}(\Delta t, t) \tag{A5.1.4}$$

This is merely the old (2.4.4) formulated in the new symbols and so far we have not introduced any specific model. But that is the next step.

MODEL A5.1.1:
We assume that

I For any arbitrarily small $\varepsilon > 0,$ a finite subdivision according to (A5.1.1) and (A5.1.2) can be found such that
  (i) $P_\varepsilon(t)\pi_{i\varepsilon}(\Delta t, t) < \varepsilon,\ i = 1,2,\ldots,n$
  (ii) $\pi_{ij}(\Delta t, t)$ is *t*-independent for $i, j = 1,2,\ldots,n$

II In the light of (ii):

$$\pi_{ij}(\Delta t, t) = \delta_{ij} + \lambda_{ij}\Delta t + o(\Delta t) \tag{A5.1.5}$$

where λ_{ij} is a constant and δ_{ij} is the Kroenecker symbol

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{A5.1.6}$$

From (A5.1.3) it is seen that condition (ii) is satisfied if $\rho(\tau, q)/P_j(\tau)$ is independent of τ and/or if $\pi_i(t - \tau, q)$ is constant for $q \in Q_j$; the former is often associated with the idea rapid mixing and the latter with some form of homogeneity (as pointed out in Appendix 9 the two properties rapid mixing and homogeneity go together like hen and egg). The two are combined in the term **efficient stirring**. It should also be noted that (A5.1.5) is indeed so common an assumption in various fields that it could be called habitual (e.g. one finds it at the derivation of the Poisson process), but it is stronger a statement than is Postulate IV.

A5.2 The model of motion

On accepting Model A5.1.1 we obtain from (A5.1.4)

$$\frac{P_i(t + \Delta t) - P_i(t)}{\Delta t} = \sum_{j=1}^{n} P_j(t)\left[\lambda_{ij} + o(\Delta t)\right] \tag{A5.2.1}$$

We thus assume that *n* is selected large enough so that we can neglect the residual subsystem S_ε. This means that the model implies that the derivative

$$\dot{P}_i(t) \equiv \frac{dP_i(t)}{dt} \qquad \text{(A5.2.2a)}$$

exists for $i = 1,2,\ldots,n$. We thus have here an example of how an even somewhat general model determines analytic properties of the function $P_v(t)$; i.e. the problem touched upon at the end of Section 2.5.

A sum up of the preceding can be done in the form of a vector differential equation of first order with constant coefficients:

$$\dot{\mathbf{P}}(\mathrm{t}) = \left\{\lambda_{\mathrm{ij}}\right\}\mathbf{P}(\mathrm{t}) \qquad \text{(A5.2.2b)}$$

where

$$\mathbf{P}(t) \equiv \begin{pmatrix} P_1(t) \\ P_2(t) \\ \vdots \\ P_n(t) \end{pmatrix} \qquad \text{(A5.2.2c)}$$

The solution of this differential equation depends on certain properties of the matrix $\left\{\lambda_{\mathrm{ij}}\right\}$, and because of the probability condition ($0 \le prob. \le 1$) some of the constraints on the matrix are:

$$\sum_{i=1}^{n} \pi_{ij}(\Delta t) \le 1, \quad j = 1,2,....,n;\ \Delta t \ge 0 \qquad \text{(A5.2.3a)}$$

and

$$0 \le \pi_{ij} \le 1; \quad i \ne j; \quad i, j = 1,2,\ldots,n; \quad \Delta t \ge 0 \qquad \text{(A5.2.3b)}$$

Should equality hold in (A5.2.3a) for all values of *j* it means that Q_S contains all possible states that a $u \in U_S$ can ever attend and, hence, *S* is closed for the substance *U*. Now, to avoid uninteresting technicalities we make the otherwise non-necessary assumption that the matrix $\left\{\lambda_{ij}\right\}$ is what in matrix algebra is called *irreducible*, which means that *S* is fully connected (again, this assumption is not necessary and made for algebraic convenience only). And from that assumption we can conclude that should there be inequality in (A5.2.3a) for at least one value of *j*, *S* is open: there are accessible states outside Q_S, which means that a $u \in U_S$ can leave *S* for the environment *E*.

Insertion of (A5.1.5) into (A5.2.3a) gives

$$\lambda_{jj} \leq -\sum_{\substack{i=1 \\ i \neq j}}^{n} \lambda_{ij}, \; j = 1,2,\ldots,n \tag{A5.2.4}$$

From (A5.2.3b) it then follows that the diagonal elements of $\{\lambda_{ij}\}$ are negative and the other elements are non-negative. According to general matrix algebra this means that the eigenvalues of $\{\lambda_{ij}\}$ have non-positive real parts and that the eigenvalue with the largest real part is single and real.

Let us then introduce an additional assumption

$$\{\lambda_{ij}\} \text{ has no multiple eigenvalues} \tag{A5.2.5}$$

This assumption is not an assumption in the sense of model (cf. section 0.2) but it is in fact rather well supported. Thus, it seems to be commonly agreed upon that, for physical reasons, the probability for multiple eigenvalues is negligible if not even zero.

The assumption about the absence of multiple eigenvalues is certainly not a postulate, and perhaps we are back to the idea about "law" as considered in Section 2.5 in the discussion of Postulate IV. It could be that the concept of law should after all be seriously considered. Thus, we have seen the need for taking into account various physical experiences and perhaps the notion of "law" could provide a tool for doing this formally. That is, a *law of nature*, as I understand it, should be the kind of experience that technically serves as a postulate but is not a postulate in the sense applied in section 0.2: its character of unusually well established empirical experience is what is crucial. Perhaps this kind of thinking would apply to Postulate IV, but concerning the present assumption I am not all that convinced: is this question about multiple eigenvalues actually a matter of a "well established physical *experience*"?

At any rate, if we accept the assumption then, from what we know about the eigenvalues because of (A5.2.4), the general solution to (A5.2.2b) can be written on the form

$$\mathbf{P}(t) = \sum_{l=1}^{n} k_l \boldsymbol{Z}_l e^{z_l t} \tag{A5.2.6a}$$

where $\{z_l\}$ are the eigenvalues (all have non-positive real parts) and $\{\boldsymbol{Z}_l\}$ the corresponding eigenvectors of the matrix $\{\lambda_{ij}\}$; the elements of the set $\{k_l\}$ are constants uniquely determined by the initial condition

$$\gamma \equiv \begin{pmatrix} P_1(0) \\ P_2(0) \\ \vdots \\ P_n(0) \end{pmatrix} \tag{A5.2.6b}$$

so again we see the central role played by the initial condition γ and the necessity of it being well specified.

This model has an obvious appeal to intuition, namely that the system S can be subdivided into a finite number of "small" subsystems: it is a kind of "infinitesimal analysis" that seems to be in good tradition of classic physical-chemistry (in the theory of thermodynamics, say). For the time being we merely note that we here have a nice illustration of how a "model" can help to elucidate the analytic properties of $P_v(t)$, namely the solution to the equation (2.4.4) as discussed in the comments on Postulate IV in the later part of Section 2.5.

It is important to realize that a model of this kind is nothing but a model (a realization of some more or less speculative thinking), and although it appeals to intuition—and might hint various things—it can never *prove* a thing. Personally I consider this kind of model as a "mathematical device", but it is then crucial to note that, when the model is applied as a mathematical device the results must be *independent* of the number n of subsystems. Of course, one application that I have in mind is the one in Appendix 3 for the judgment on a suggested ergodic principle (D^l being independent of γ).

Appendix 6
On the Stability of Stochastic Processes

A6.1 A background

The notion of stochastic process is considered difficult and there is in the literature a variety of definitions or representations of that concept, all of which are supposed to be formally equivalent but do have different flavor to intuition. To get some idea of what we are up to, let us ignore all formal details and consider the situation depicted by Fig. A6.1.1: at the time $t = 0$ a particle is released at *A* and moves to the right because, say, there are a barrier to the left of the Y-axis and a transition probability (for motion to the right) acting on the particle. The particle's position at the time $t = t_1$ is recorded. The experiment is repeated many times and at each experiment the location $X(t_1) > 0$ is observed and recorded. After, say, 100 such experiments the recorded positions of the particle form a realization of a probability distribution indicate by (a) in Fig. A6.1.1. Assume that we also record each particle's location at $t = t_2 > t_1$ as well: this gives a distribution indicated by (b) in Fig. A6.1.1.

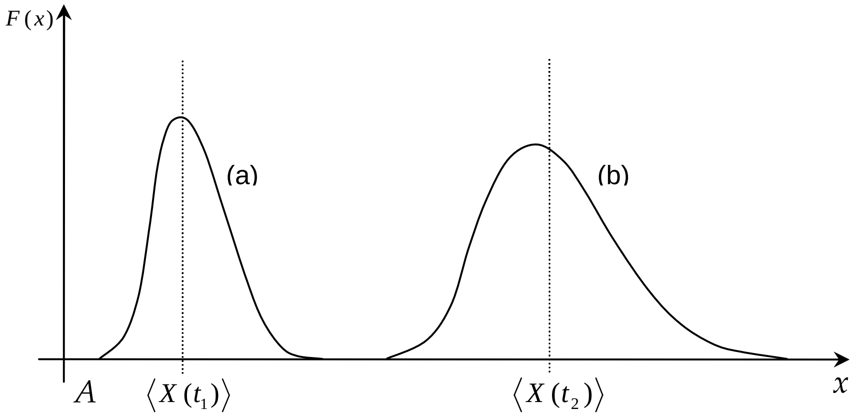

Fig. A6.1.1 A random walk as illustration of the notion of stochastic time process. Let a particle u *be released at A at the time* $t = 0$*. Because of some mechanism acting on* u *it moves to the right and its position is observed at the times* $t_1 > 0$ *and* $t_2 > t_1$*. The experiment is repeated many times and the recorded positions form two realizations of probability distributions. In general one should expect the standard deviation at* t_2 *to be larger than that at* t_1*. However, the difference is often less than one would expect, because stochastic time processes in nature usually contain a deterministic part that gives a kind of stability to the processes. It is this deterministic part that, when sufficiently dominating, allows for deterministic data analysis. Often there is a co-variance between the two data clusters (it decreases when the time-distance between the two clusters increases), and this means that data from such processes contain less information than ones intuition expects (e.g. no matter how complex the underlying physics is, kinetic data look usually nice and smooth).*

There is a wealth of questions that this scenario gives rise to, some of which are of deep mathematical significance. Here I shall however limit the attention to one basic thing, namely that we consider a variable $X(t)$ – the particles location on the x-axis at the time t – that for each fixed $t = t_1$ or t_2 it is an ordinary stochastic variable (namely the particles position at the times t_1 respectively t_2), which among various things implicates that there is a probability distribution with the density $\rho(t,x)$ for $t > 0$ such that $\rho(0,x)$ is the Dirac delta function $\delta(0)$; that is, at the time $t = 0$ the location of the particle is $X = 0$ with probability one (i.e. at the time $t = 0$ the particle is *exactly* at the location *point A* in Fig. A6.1.1).

Let us now consider the set $\{x(t_1), x(t_2); t_2 > t_1\}$. That is, we have now realizations of the ONE AND THE SAME (Sic!) particle's location at two consecutive time points. This is about the simplest possible example there is of a realization of a stochastic time process, and a typical question in that context is about the covariance between $X(t_1)$ and $X(t_2)$; for instance, if the two time points are not very far apart one would expect that if $x(t_1)$ is well to the left of the mean value $\langle X(t_1)\rangle$ it is likely that also $x(t_2)$ is located to the left of $\langle X(t_2)\rangle$.

One should thus expect that consecutive observations of a time process are usually not mutually independent. This is one of the fundamental difficulties with this kind of processes, namely that data from observations of such processes usually do not contain as much information that one intuitively would expect.

There is one aspect of this particular example that I want to emphasize. Thus, there is an uncertainty (variance) in $X(t_1)$ and it seems intuitively reasonable to assume that this uncertainty will, so to speak, add to the uncertainty of $X(t_2)$ or, differently expressed, one would expect the standard deviation of $X(t_2)$ to be larger than that of $X(t_1)$. Actually, if we follow the particle for a longer period of time $t_2 < t_3 < \ldots < t_n$ it would appear natural to expect that the standard deviation to increases with time. But usually this does not happen—there is a kind of stability present, and if one should ask why this is so there is usually no answer (it is somewhat similar to the common situation with the force of gravitation: physically speaking one has usually not known what that force is, it has only been possible to note that it is there). Commonly this "stability" has been represented by considering the stochastic process as a sum of one deterministic component and one stochastic component (Fig. A6.2.2 in the next section). And often the deterministic part is represented by the probability density $\rho(t,x)$ being a deterministic function of classic form. Personally I like to express this as: some kind of determinism is necessary for the stochastic process to have the **stability** we often can observe but cannot explain. In fact, this stability is necessary (but of course not sufficient) for the mathematics of stochastic processes to be practically useful. Therefore, practically I use to search for forms of determinism in observed stochastic processes.

One of the expressions of the stability one meets in nature is the possibility of such things as mass production of shoes and costumes (obviously the variation in human size is not totally wild); but this being so, things like complexion and fingerprint reveal the other side of the coin, namely that there is also an underlying variance of quite some significance. And it is imperative that the language we construct allows for such a dichotomy (determinism together with large randomness). Well, occasionally the observable stability one meets in nature is not totally impossible to "understand": various feedback mechanisms imply a stability that occasionally is referred to as homeostasis, and the high buffering capacity of blood is not only well known but also reasonably understood in terms of classic chemistry. An example of how nature seems to induce stability by counteracting deviations from the "normal" is the notion of regression in classic medical genetics, as exemplified by how tall parents tend to receive children of normal length, and analogous for short parents: also their children tend to be of normal length (the kids grow up taller than their parents).

A6.2 The Faraday rule

When interpreting his experiments in electrolysis Michael Faraday applied a rule that stated that the data analysis should be done in a language (grammar – often in the form of postulates – and a vocabulary representing concepts) that was such that the experiments could be discussed without making necessary any commitments to hypotheses about the unknown and the uncontrollable. For instance, initially he wanted to name the two electrodes (in an electrolytic cell) east and west but a philologist whom he consulted pointed out that this implied a hypothesis about a connection with magnetism and this violated Faraday's own principle (this was before the work of Maxwell); the eventual result of this terminology-fuss was that the two names anode and cathode were coined (Williams, 1971).

Personally I like to formulate **Faraday's rule** as:

> The language must be such that ease of use is obtained through concepts that do not *ignore* but are *independent of* the factors that cause difficulties.

This principle has been widely applied during the years. For instance, the meaning of "rate constant" in classic chemistry does not ignore but is *independent* of the complexities of chemical reactions that are dealt with in quantum chemistry. And the significance of "mean-value" in statistics is by and large *independent* of the look of the underlying distribution.

It is in the light of this rule I often like to contemplate the usefulness of the theory of stochastic processes, namely that one often can in *exact* terms speak of the probability of a physical event without getting involved in the nature of the underlying physical mechanisms. Usually such mechanisms are crushingly "complex" and an analysis in those terms necessarily becomes speculative and violates Faraday's rule (in other words, the whole thing becomes messy). That is, the various probability concepts can be given physical significance *independently* of the underlying mechanisms in terms of more classic physics. This is one of the features of probability that makes this notion so powerful.

However, to use the theory of stochastic processes for this reason carries a cost, namely that the very idea of stochastic process means that one introduces elements of randomness (that mask, as it were, the investigator's ignorance about the underlying physics). And it is here that the stability matter discussed in the preceding section becomes important, because without this stability the theory's usefulness would be lessened at design and interpretation of experiments.

In order to handle the situation it is important to note that it is very much a matter of selecting appropriate concepts. Yes, it becomes a linguistic problem as it were, and the primary task is to visualize the notion of stochastic process $X(t)$, and the empirical dichotomy "stability together with variance". And as already hinted in the preceding section, to this end I like the representation

$$X(t) = \mu(t) + \xi(t)$$

Fig. A6.2.2 *A rather general representation of a stochastic process. It is looked upon as a sum of one* deterministic *part* $[\mu(t)]$ *and one purely* nondeterministic *part* $[\xi(t)]$*, and it supplies the reader with a somewhat acceptable "feel" for how the two properties contribute to observable features of the Nature's behavior.*

where $\mu(t)$ is the deterministic part (a time function of classic sense) and $\xi(t)$ is the purely nondeterministic part (cf. Cramér and Leadbetter, 1969); often it is convenient to write the mean value $\langle \xi(t) \rangle \equiv 0$, so that $\langle X(t) \rangle = \mu(t)$. In these terms – and using standard probability algebra – it is not difficult to see how the randomness balances the deterministic feature and, above all, the converse. That is, because of the stability discussed in the preceding section it becomes urgent to recognize the determinism that obviously does exist in nature; the necessity of such an existence follows from the representation depicted in Fig. A6.2.2, together with the empirical facts discussed in the preceding section. It is of course a matter of description: any attempt to "explain" the situation would be contra productive. Again, it is very much a linguistic matter, namely how to describe and conceive the dichotomy that Mother Nature supplies us with.

It should be noted here that the treatment of random walk in standard textbooks usually shows how the variance increases with time (i.e. with increase of the distance the particle moves). The problem here is that often this variance is under some form of control, and Fig. A6.2.2 suggests one possibility to represent this phenomenon, which suggests that the influence of the deterministic element should increase with time. The formal expression of "how" is outside the present discussion; perhaps one might say that the search for deterministic factors constitutes a significant part of the work on the topic.

Appendix 7
Qualitative Analysis

A fundamental procedure of science is comparison. Perhaps one could go so far and say that the very purpose of science is comparison of systems; it could be the same "system" at two different occasions (e.g. a lake in summer and in winter) or it could just be two different systems like two rabbits. By the term **quantitative analysis** I mean that the comparison is in terms of numerical estimates of some parameter(s). An example would be a simple irreversible first order chemical reaction of the type $a \rightarrow b$ taking place in a small well-stirred vessel. If the volume concentration *C* of *a* is observed as a function of time, which according to classic theory should be of the form $C(0)\,exp\{-kt\}\big|_{k>0}$, the so-called *rate constant k* can be estimated. What has been a rather popular project is to study such a reaction at different temperatures in order to see how *k* depends on temperature: one compares the same system (reaction) at different temperatures, and those comparisons are made in terms of the parameter *k*. That is, for each selected value of the temperature the value of *k* is estimated from plots of the above function and, hence, one can see how the rate constant depends on the temperature. Thus, one can say that the reaction goes so and so much faster at one temperature than at another temperature: the value of *k* allows for a *quantitative* measure of "how much faster". From such a temperature dependency one might estimate another parameter, namely the ACTIVATION ENERGY, and then compare quantitatively different but similar reactions (e.g. $m \rightarrow n$)—different systems that is—in terms of that parameter (yes, the temperature behavior of each such reaction must be determined separately, so we are considering an investigation that usually takes little more time than just a few minutes).

There are parameters that can be numerically estimated *directly* as it were. And as we here are dealing with a well-stirred system the simple notion of temperature is an example of such a quantity. As a principle it is just a matter of placing a thermometer into the mixture and read, which in a more magnificent language sounds "making an empirical observation".

Important in that kind of quantitative analysis is that the parameters have some meaning (significance) in terms of some particular theory (Fig. 0.3.1): the comparison must mean something more than whether one numerical value is greater or less than some other numerical value. Another thing of importance is that the numerical estimates are *sufficiently good* in the sense that a comparison is feasible in terms of those values, and almost invariably this means that the statistical variance is reasonable and that the sample size is large.

It is this last demand that often causes difficulties, especially in bio-medical work where often the variances are annoyingly large and the sample sizes frustratingly small, and this to the extent that usual quantitative comparison of parameter values becomes meaningless. An example is when data are a result of an accident and the amount of data is quite limited but nevertheless can be expected to contain valuable information (e.g. in the past, radiation accidents have been a dominant source of information for understanding of man's response to radioactive irradiation). And for such situations I have earlier suggested a form of data analysis that I name **qualitative analysis** that, as discussed by Gradijan and Bergner (1972), is somewhat difficult to describe in reasonably simple terms (there is yet no general formalism), and therefore a description of the methodology must be somewhat heuristic.

Perhaps it would be correct to identify quantitative analysis with what might be called classic **analysis.** An aspect on this kind of data analysis should then be that it employs, in my personal terminology, "Fisher statistics", a characteristic feature of which is that words such as level of confidence and of significance appear in the basic terminology. A typical feature of qualitative analysis should thus be that such notions appear nowhere, simply because there is no attempt to parameter estimation.

For instance, it could be that one is able to state "data are consistent with the hypothesis that *if the quantity X increases also the quantity Y increases*", where one should note the use of the term "consistent" at the same time as nothing is said about how fast the increases are. And if we first consider this notion of consistent, one could perhaps say that in qualitative analysis one tries to collect independent consistencies (e.g. a number of independent sets of data all pointing in the same direction), and the larger that number of consistencies is the stronger is the empirical support for the contemplated hypothesis; in a way the number of consistencies replaces the common notion of significance level in quantitative analysis.

If we then turn to the matter that nothing is said about the magnitudes of the increase of the involved quantities, one may observe that a formal expression of this aspect would be:

$$X \uparrow \Rightarrow Y \uparrow$$

Personally I feel that this expression illustrates pretty well not only the qualitative character of this particular hypothesis but also the very sense of the term "qualitative" as such. See also (4.1.6b) in Chapter 4.

Another aspect would be that whereas in classic quantitative analysis one might use statements like "in agreement with data" a corresponding statement in qualitative analysis would be "consistent with data". Often this means that in qualitative analysis *equalities are replaced by inequalities*.

To illustrate this point let us assume that we observe two time processes for which we have constructed the models $a(t)$ and $b(t)$ and that the observations have resulted in the data $\{a_i\}_{i=1}^{i=n}$ and $\{b_i\}_{i=1}^{i=n}$ corresponding to the time points $t_1, t_2 \ldots, t_n$. If then for all values of *i* we have $a_i < b_i$ and $a(t_i) < b(t_i)$ we might say that the models are consistent with data; yes, it would be too much to say that "the models agree with data", and at least to me the phrase "the models are consistent with data" sounds more appropriate (more modest as it were).

It would perhaps be tempting here to say that there are *n* consistencies in this example but, as I have pointed out earlier, typical for time processes is often that consecutive values are not mutually independent and, therefore, consecutive observations are not mutually independent: it might in fact be pretty difficult to say what the number of independent consistencies actually is.

Unquestionably there is commonly a need for as many independent consistencies as possible, and application of the notion of stochastic process might then be called for. For instance, should the time processes in the preceding paragraph be stochastic it could be that the quantities *a* and *b* are mean values of the observed quantities and, hence, it would be meaningful and possible to consider the corresponding standard deviations *sa* and *sb* as functions of time: should inequalities similar to those above hold also for *sa* and *sb*, one could rather safely speak of at least two independent consistencies (with respect to the mean values *a* and *b*, and with respect to standard deviations *sa* and *sb*). This aspect has been explored by Gradijan and Bergner (*loc cit.*): the model is "fitted" to data with respect to both the mean values and the standard deviations; the variance is then not merely an annoyance but in fact a *property* of the investigated system.

To view statistical variance as a system property is old hat in today's physics but not so in biology and medicine, where the methodology has for a long time been focused on methods for making that data feature as "small" as possible. As a matter of fact, when such variance appears as a significant system property in a biomedical investigation conceptual problems arise. For instance, imagine an animal study where the toxicity of two drugs is investigated and that the drugs have the same LD_{50} but fundamentally different variance. To the best of my knowledge the average pharmacologist of today does not have an appropriate conceptual apparatus for dealing with that kind of situation; i.e. to meaningfully discriminate between the two drugs with respect to that particular data feature.

Perhaps it is correct to say that one of the purposes of qualitative analysis is to save data, in the sense of preventing data from being discarded as "insufficient" (i.e. insufficient when viewed in terms of quantitative analysis: too large variance and too small sample). But qualitative analysis is seemingly more subjective than is quantitative analysis, though the term *seemingly* should be carefully noted: for instance, often statistical tests in quantitative analysis require conditions that are assumed (Sic!) to be valid, and frequently such assumptions are quite subjective indeed.

In summary, a basic difference between quantitative and qualitative analysis is that in the latter no attempts are made to estimate parameter values. In qualitative analysis one frequently experiences this as some extra freedom at selection of models and, technically, that equalities (in quantitative analysis) are replaced by inequalities. One possible application of this kind of qualitative data interpretation is given in Section 4.1, which also demonstrates how qualitative analysis enhances the possibility to avoid hypotheses about the underlying physical mechanisms.

Appendix 8
The correspondence principle

In order to compare two quantities they must have the same denomination or, as we for historic reasons call, **physical significance**. For instance, it is meaningless to ask whether two kilometer is larger than one kilogram. Yes that's right; this is because the two quantities do not have the same physical significance. Primary concepts have no definition from which to inherit an existing physical significance and therefore each primary concept must be *assigned* one of its own. (As a principle, two quantities of different primary concepts can not be compared.)

Such thinking is at least partly behind the **correspondence principle** of quantum mechanics (cf. Merzbacher, 1961), one formulation of which could be: *if the old theory (classic physics) and the new theory (quantum mechanics) are simultaneously applicable on the same phenomenon there must be a simple and direct correspondence between the respective quantities of the two theories.* By this correspondence the new quantities are given physical significance.

Historically the correspondence principle belongs to the early development of quantum mechanics, but its validity is general. To me it seems evident that despite what new quantities are developed they must have physical significance, so that the issue of their comparability can be decided upon.

Another aspect of the correspondence principle could be called "conceptual continuation". This is an interpretation of the correspondence principle that can also be found in the literature, saying that the principle makes it possible to *understand the new into terms of the familiar*. Perhaps one could say that without that principle scientific theory would be much more fragmented than it presently is, and the communication problems would be even greater than they already are; but one could also say that today's diversity and frequent lack of mutual understanding are there very much because of the correspondence principle being so often ignored.

So in the light of this, the correspondence principle has two essential functions as it were: (1) to control the comparability of new quantities; (2) to simplify the understanding of new quantities. And it cannot be overemphasized that one inescapable consequence of this is that primary concepts must be given (their own) physical significance.

References

Bergner, P.-E. E. (1965): On stationary and non-stationary red cell survival curves *J.Theor.Biol.* **9**, 366-388

Bergner, P.-E. E. (1964): Tracer Dynamics and the Determination of Pool-sizes and Turnover Factors in Metabolic Systems. *J.Theor.Biol.* **6**, 137-158

Cramér, H. (1954): *Mathematical Methods of Statistics*. Princeton University Press

Cramér, H. and Leadbetter, M.R. (1967): *Stationary and Related Stochastic Processes.* John Wiley & Sons, Inc. New York

Gradijan, J.R. & Bergner, P.-E. E. (1972): Qualitative consequences of randomness in a linear kinetic system. *Biometrics*. **28**, 313-328

Khinchin, A.Y. (1960): *Mathematical Methods in the Theory of Queueing.* Charles Griffin & Co. LTD, London.

Morowitz, H.J.(1968): *Energy Flow in Biology*. Academic Press, New York

Orr, J.S. & Gillespie, F.C. (1969): Occupancy principle for radioactive tracers in biologic systems. *Science* **162** 138-139

Stephenson, J.L. (1960): Theory of transport in biological systems. *Bull.Math.Biophys.* **22**, 1-17

Takács, L. (1960): *Stochastic Processes* Methuen & Co. LTD, London

Williams, L.P. (1971): *The Selected Correspondence of Michael Faraday*. Volume Cambridge University Press, Cambridge. UK

Yaglom, A.M. (1962): *Stationary Random Functions*. Prentice Hall, Inc. New Jersey

Index