An information theoretic approach to system differentiation on the basis of statistical correlations between subsystems

Jürgen Jost, Nils Bertschinger, Eckehard Olbrich,
Nihat Ay, Sidney Frankel
Max Planck Institute for Mathematics in the Sciences,
Inselstr. 22, D 04103 Leipzig, Germany

11.10.2006

Abstract

We develop an analysis of complex systems in terms of statistical correlations between the dynamics of its subsystems as a formal framework within which to understand processes of system differentiation.

1 Introduction

Durkheim[9] introduced integration and differentiation as basic concepts of social theory. In one of the key works of modern sociology, Luhmann's theory of social systems [14, 16], three types of differentiation are distinguished:

1. Segmentary differentiation
2. Stratificatory (hierarchical) differentiation
3. Functional differentiation

Segmentary differentiation is exemplified by a society consisting of independent groups or tribes that operate in parallel and essentially independently of each other, with no division of labor. Stratificatory differentiation is characterized by some central control; historically, it emerged at the transition from small neolithic settlements to centrally organized agricultural societies in Mesopotamia, Egypt, and elsewhere. The control intertwines economic, religious, military, and other aspects. The differentiation between center and periphery that is often emphasized by other authors can perhaps be subsumed as a special case of stratificatory differentiation. In functional differentiation, subsystems fulfill particular functions for the system. On that basis, they develop some kind of autonomy, that is, configure their dynamics according to internal goals of (sub)system identity. In attempting to achieve closure from the rest of the system, they relegate the function performed for the system at large to a means for achieving their own ends, see [13]. Because of the relative autonomy of these subsystems, central control is no longer possible, and coordination between the subsystems has to be resolved locally. Luhmann [14, 16] argues that the third type, functional differentiation, is characteristic of modern European societies and of contemporary world society. For a survey of this and other approaches to social differentiation, see [19].

It is the purpose of this contribution to develop formal quantities, more precisely (differences of) temporal correlations between subsystem statistics, that can characterize and quantify different stages of system differentiation and put them into the perspective of a general theory of the evolution of complex systems, as an interplay between integration and differentiation processes.

Let us consider a very simple example, military organization. The first stage consists of a bunch of ferocious fighters that may charge the enemy in a group rush, but then fight independently and autonomously. The best fighters are distinguished as heroes. Military training, if it exists at all, emphasizes individual skills. Gradually, fighters become soldiers whose movements are coordinated by a central commander. The Roman army achieved much of its superiority by a highly trained synchronization of the operations of the soldiers. Military training emphasizes a structure of obeying commands that are passed down from a strategist to the individual soldier via a hierarchy of officers. The army leader becomes the hero that takes the credit for a victorious battle. On the basis of this, some division of tasks between different groups becomes possible, forward and rear, infantry, cavalry, technical troops, fortification and siege specialists. The Roman army, to consider the example once more, established a superior logistic planning. Still, the control rested essentially with a central commander. In modern fighting, we see semi-independent groups operating according to some general plan or strategy, but more detailed decisions are taken locally. These groups still need to draw upon central logistic and other resources, but much of these resources are distributed according to local requirements. There are no longer individual heroes to whom the success is attributed. − Of course, the military system has ceased to play a leading role in the evolution of modern societies, and modern societies operate in a rather different manner than armies. The military system therefore does not constitute the most appropriate example for understanding social structures. This example has been chosen here only because of its conceptual simplicity, and not because of any claims of strongest relevance.

2 Function and structure

The term "function" has (at least) two different meanings, both of which will be relevant for us:

1. The purpose of a subsystem or component as assigned by the encompassing system and the operations contributing to that purpose

2. A dependency relation between elements

We investigate them in turn:

1. 2 aspects:

   (a) A part Z of a system S carries out a function F for S when that operation is beneficial for S, for example by contributing to its survival or reproduction

   (b) The function originated for that purpose, for example by a process of selection

   The second aspect is sometimes problematic. In evolutionary biology, Gould and Vrba [11, 10] emphasized the distinction between historical origin and current utility (as had Nietzsche [18] earlier) and developed the concept of exaptation, that is, a structure employed for a current utility for which it had not been originally developed. One can construct examples for either the inclusion or the exclusion of that aspect in the definition of a function. Overall, perhaps inclusion is preferable. This is related to the problem of formalizing the notion of operational closure. In any case, it is important that Z is a part of S, in order not to assign system functions to external circumstances.
   Thus, the point is to relate a function to a system purpose. That purpose can be internally defined or stipulated by an external observer or designer. The function may be hypothetical, that is, anticipate certain external or internal risks, like a spare tire.
   Whereas the purpose or aim of the system is globally determined, functions can be locally assigned.

2. When a system consists of elements $\alpha = 1, \dots, m$, their states $z^\alpha(t)$ will satisfy certain functional dependencies. We first consider the deterministic case. We can then write (not necessarily uniquely) the dependencies as

$$z^\alpha(t+1) = \phi^\alpha(z^1(t), \dots, z^m(t)). \quad (1)$$

   Typically, however, not every element will be functionally dependent on every other element. This leads us to modularity and aggregation, concepts that have discussed in the literature, for example by Simon-Ando [22] and Shpak et al. [20, 21]. We present here the following version: A deterministic system (1) is modular when $\phi^\alpha(z^1(t), \dots, z^m(t))$ depends at most on z^β for $\beta \in I_\alpha$ where I_α is a proper subset of $\{1, \dots, m\}$ and $I_\alpha = I_\beta$ whenever $\beta \in I_\alpha$. In order to achieve that symmetry, however, for a given α, the dependence set I_α might not be the smallest possible one because within a module, not every element need directly depend on every other one.
   Aggregation means that

$$z^\alpha(t+1) = \phi^\alpha(z^1(t), \dots, z^m(t)) = \Phi^\alpha(Z^{J_1}(t), \dots, Z^{J_\alpha}(t)) \quad (2)$$

where (for suitable functions f^J)

$$Z^J = f^J(z^{\alpha_1}, \dots, z^{\alpha_J}) \tag{3}$$

are aggregate variables. Obviously, this can be iterated over different system levels.
When the element states are modeled as random variables, for modularity we would require (assuming for simplicity a Markov condition) that, given $z^\beta(t)$ for all $\beta \in I_\alpha$, $z^\alpha(t+1)$ is conditionally independent of $z^\gamma(t)$ for all γ not contained in I_α. We point out, however, that this does not necessarily imply that the states $z^\beta(t)$ for $\beta \in I_\alpha$ causally determine $z^\alpha(t+1)$.
For the aggregation, we construct a Bayesian network with nodes $z^\alpha(t)$ and newly introduced aggregation variables $Z^{J_1}(t), \dots, Z^{J_\alpha}(t)$ which depend on (possibly subsets of) $z^1(t), \dots, z^m(t)$ and become the parents of $z^\alpha(t+1)$. Formally speaking we require $(z^1(t), \dots, z^m(t)) \to (Z^{J_1}(t), \dots, Z^{J_\alpha}(t)) \to z^\alpha(t+1)$ to be a Markov chain. $Z^{J_1}(t), \dots, Z^{J_\alpha}(t)$ are meaningful aggregation variables if they provide a more compact representation of the relevant part of the original state space, i.e. the entropy $H(Z^{J_1}(t), \dots, Z^{J_\alpha}(t)) < H(z^1(t), \dots, z^m(t))$ and at the same time preserve all the information that is needed to determine the next state $z^\alpha(t+1)$, i.e. $MI(z^\alpha(t+1) : Z^{J_1}(t), \dots, Z^{J_\alpha}(t)) = MI(z^\alpha(t+1) : z^1(t), \dots, z^m(t))$.[1]
We note that modularization as defined here constitutes an ideal type (in the sense of Max Weber) that cannot be realized within an integrated system because such systems are not completely decomposable. In terms of correlations, however, we can distinguish between strong correlations within a module and relatively weaker ones between modules, that is, the conditional mutual information $MI(z^\alpha(t+1) : z^\gamma(t)|I_\alpha(t))$ for γ not in I_α need not exactly vanish, but only be small.

Differentiation is related to both function concepts. In terms of functional aspects (in the first sense), a system function can be decomposed into subfunctions, and these can be carried out by corresponding subsystems. In terms of relational aspects, such a differentiation is then supported by a modularized structure that specializes in carrying out that function.

3 Information theoretic concepts

We consider a system X whose state at the discrete time $t \in \mathbb{N}$ is $X(t)$ and that consists of subsystems $X_i, i = 1, \dots, m$, with states $X_i(t)$. We also write $(X \backslash X_i)(t)$ for the state of the system without the component i. We consider the $X_i(t)$ as random variables and apply concepts from information theory; for an introduction, see e.g. [17].
The uncertainty about the system state $X_i(t+1)$ is given by the entropy

$$H(X_i(t+1)). \tag{4}$$

[1] A mathematical formalization would involve filtrations of sigma-algebras.

Given the system state $X_i(t)$, the uncertainty about $X_i(t+1)$ reduces to

$$H(X_i(t+1)|X_i(t)). \tag{5}$$

The mutual information between subsequent states of X_i measures the (expected) reduction in uncertainty about $X_i(t+1)$ when $X_i(t)$ is known; it is given by

$$MI(X_i(t+1) : X_i(t)) = H(X_i(t+1)) - H(X_i(t+1)|X_i(t)). \tag{6}$$

Given the full system state $X(t)$, the uncertainty about $X_i(t+1)$ reduces to

$$H(X_i(t+1)|X(t)). \tag{7}$$

This then is the part of the uncertainty about $X_i(t+1)$ that is coming from outside the system, that is, the uncertainty about $X_i(t+1)$ not reduced within the system X. Consequently, the amount of information about $X_i(t+1)$ determined within the system X then is the mutual information

$$MI(X_i(t+1) : X(t)) = H(X_i(t+1)) - H(X_i(t+1)|X(t)). \tag{8}$$

Next, the allo-determination of X_i within the system, that is, the amount of information about $X_i(t+1)$ not determined by $X_i(t)$, but by the state of the rest of the system is

$$H(X_i(t+1)|X_i(t)) - H(X_i(t+1)|X(t)), \tag{9}$$

and the self-determination of X_i then is

$$H(X_i(t+1)|(X \backslash X_i)(t)) - H(X_i(t+1)|X(t)). \tag{10}$$

Modularization as defined above in §2 should then increase the self-determination. We also have the integration of a system consisting of components i:

$$I(X)(t) := \sum H(X_i(t)) - H(X(t)) \tag{11}$$

which is always non-negative.

4 Differences

We now consider the difference between (6) and (10), that is,

$$\begin{aligned} & H(X_i(t+1)) - H(X_i(t+1)|X_i(t)) \\ & -(H(X_i(t+1)|(X \backslash X_i)(t)) - H(X_i(t+1)|X(t))) \\ = \quad & MI(X_i(t+1) : X_i(t)) - MI(X_i(t+1) : X_i(t)|(X \backslash X_i)(t)). \end{aligned} \tag{12}$$

This is the mutual information between subsequent system states minus the self-determination. This then is positive and large when the state of i itself contains a large amount of information about subsequent states and when much of

this information can also be obtained from the rest of the system.
Conversely, this quantity is negative when $X_i(t)$ and $(X\backslash X_i)(t)$ are complementary. An example is when $X_i(t+1) = \mathrm{XOR}(X_i(t), (X\backslash X_i)(t))$ (see 5. below).
We can rewrite (12) by rearranging the differences as

$$\begin{aligned} & H(X_i(t+1)) - (H(X_i(t+1)|(X\backslash X_i)(t)) \\ & -(H(X_i(t+1)|X_i(t)) - H(X_i(t+1)|X(t))) \\ = \quad & MI(X_i(t+1) : (X\backslash X_i)(t)) - MI(X_i(t+1) : (X\backslash X_i)(t)|X_i(t)). \end{aligned} \tag{13}$$

This is large when the mutual information between i and the rest of the system is large while the information flow into i is small. The latter is, for example, the case when i can model the rest of the system.
We note that in the analysis of (12) and (13), we have employed different interpretations of the correlation between i and the rest of the system. In the first case, we have interpreted it as a control of i by the rest of the system while in the second one as a modeling of the rest of the system by i. Of course, correlations give no indication of the direction of causality, an issue to which we shall shortly return.
In summary, we can distinguish several types:

1. Correlation (between X_i and $X\backslash X_i$): the second term on the right hand side of both (12) and (13) is small. Thus, (12) = (13) will be large when the first term is large, that is, when a lot of information is propagated from time t to $t+1$, and it does not matter whether that information comes from i itself or from the rest of the system.

2. Independence: (12) = (13) is small (in absolute value, that is, close to 0). In (12), this comes about because both terms on the right hand side are (almost) the same even though they can be individually large. In (13), by contrast, each of the terms will be small because there is little correlation between the rest of the system and i.

3. External control: Again, (12) = (13) is small, but now the two terms in (12) are small because $X_i(t)$ has little influence on $X_i(t+1)$, whereas the terms in (13) can now be large (because $(X\backslash X_i)(t)$ controls $X_i(t+1)$), but of similar value.

4. Cooperation: A type that is intermediate between the two previous ones. Here, the terms on the right hand sides of (12) = (13) are all large and of comparable size. i and the rest of the system cooperate to determine the state $X_i(t+1)$. (12) = (13)

5. Complementarity: (12) = (13) is negative because the state $X_i(t+1)$ is being determined jointly by $X_i(t)$ and $(X\backslash X_i)(t)$ whereas each of them, without the other, contains little information about $X_i(t+1)$.

The preceding quantities can be compared to the ones introduced by us in the context of a system with states $S(t)$ and its environment with states $E(t)$.

In [7], we started with the assumption of non-heteronomy,

$$H(S(t+1)|E(t)) > 0. \tag{14}$$

The basic autonomy measure then was

$$A := MI(S(t+1) : S(t)|E(t)) = H(S(t+1)|E(t)) - H(S(t+1)|S(t), E(t)), \tag{15}$$

under the assumption that the system could not influence its environment. When, in contrast, the system can control the environment, the quantity

$$A^* := MI(S(t+1) : S(t)) = H(S(t+1)) - H(S(t+1)|S(t)) \tag{16}$$

was more suitable. The difference

$$A^* - A = MI(S(t+1) : S(t)) - MI(S(t+1) : S(t)|E(t)) \tag{17}$$

was characterized in [8] as the non-trivial information closure $NTIC$. This becomes large when the information flow into the system is small while at the same time, the mutual information between the system and the environment is large. Thus, we regarded it as a measure of the extent to which the system models its environment. When we pursue the analogy between $X_i(t)$ and $S(t)$, as well as between $(X \backslash X_i)(t)$ and $E(t)$, then we see that (17) is analogous to (12). The interpretation just recalled, however, would be rather analogous to the one given to (13).

We continue our analysis of the differences. The difference between (8), measuring what is determined about i within the system, and (10), the self-determination of i, is

$$H(X_i(t+1)) - H(X_i(t+1)|(X \backslash X_i)(t)). \tag{18}$$

This quantity measures the amount of information about the state of i that cannot be gained from within the system without knowing i, that is, what is determined either from the exterior (environment) of the system or by i itself. This is a useful quantity when i is considered as part of the internal environment of the rest of the system. For example, in Luhmann's theory [14], a psychic system is part of the internal environment of a social system.
In contrast to the difference (6) − (10), the quantity (9) $= H(X_i(t+1)|X_i(t)) - H(X_i(t+1)|X(t)) =$ (18) − ((6) − (10)) does distinguish between causalities and correlations. It does not detect the difference between causality and anticipation, however; for example, it can be large when $X_i(t+1) = X_j(t)$ for some $j \neq i$ where both depend on some external factor.

5 Iterated differences

As the next step, we compare differences at the system level and at the subsystem level, that is, consider differences of differences. The prototype is

$$\begin{aligned} & H(X(t+1)|X_i(t)) - H(X(t+1)|X(t)) \\ & -(H(X_i(t+1)|X_i(t)) - H(X_i(t+1)|X(t)) \qquad (19) \\ = \; & H(X(t+1)|X_i(t)) - H(X_i(t+1)|X_i(t)) \\ & -(H(X(t+1)|X(t)) - H(X_i(t+1)|X(t)). \qquad (20) \end{aligned}$$

When there is some central control system $\star$, this quantity becomes small for all $i \neq \star$ when $\star$ directs or regulates all subsystems in the same manner, but it can get large when they are all regulated differently.
The control system $\star$ can also be identified or characterized via statistical correlations. We have the integration (11)

$$I(X)(t) = \sum_{i \neq \star} H(X_i(t)) + H(X_\star(t)) - H(X(t)), \qquad (21)$$

the conditional integration

$$I(X|\star)(t) = \sum_i H(X_i(t)|X_\star(t)) - H(X(t)|X_\star(t)) \qquad (22)$$

and the sum of conditional mutual informations

$$\sum_{i \neq \star} MI(X_i(t) : X_\star(t)) = \sum_{i \neq \star} (H(X_i(t)) - H(X_i(t)|X_\star(t))). \qquad (23)$$

Since $H(X(t)) = H(X(t)|X_\star(t)) + H(X_\star(t))$, we have

$$I(X)(t) = I(X|\star)(t) + \sum_{i \neq \star} MI(X_i(t) : X_\star(t)). \qquad (24)$$

Since the left hand side of (25) does not depend on which component is labelled as $\star$, that component $\star$ can alternatively be characterized as the one for which the sum of mutual informations $\sum_{i \neq \star} MI(X_i(t)|X_\star(t))$ becomes largest or the one for which the conditional integration $I(X|\star)(t)$ becomes smallest. Of course, in the extreme situation where every correlation between subsystems is induced by a central control $\star$, the conditional integration vanishes.

We should observe here, however, that the quantity (20) is already large when the system consists of many independent components with a deterministic time course. In that situation, $H(X(t+1)|X_i(t))$ is large while the other expressions in (19), (20) vanish. In order to distinguish the case of independent components from the one of a differentiated control, in §7, we shall look at a sequence of steps through which system complexity and differentiation can be increased.

6 Integration and reduction of complexity

A decrease in

$$\sum_i H(X_i(t+1)|X(t)), \tag{25}$$

with the absolute entropies $H(X_i(t))$ not affected, indicates some simplification of the system. This can occur for different reasons:

1. The different components i of the system become more correlated, that is

$$H(X_i(t)|X_j(t)) \tag{26}$$

decreases for some $j \neq i$. This could be any j, as in the case of synchronization, or it could be some particular $j = \star$, the same for all i, $\star$ being a controlling center for the system. Of course, when $H(X_i(t)|X_\star(t))$ decreases for all i, then also $H(X_i(t)|X_j(t))$ might decrease for all j because $\star$ introduces correlations between the various components of the system. In order to distinguish such correlations from causal influences, we might rather consider a decrease in $H(X_i(t+1)|X_\star(t))$. Here, $\star$ tells the other components i of the system what to do in the next step, and those components obey that command so that, given $X_\star(t)$, there is little uncertainty about $X_i(t+1)$. Still, however, this does not distinguish between causality and anticipation. Instead of a commander, $\star$ could simply be a good predictor. For a more profound analysis of the issue or causality, see [6]. – Of course, also intermediate cases between synchronization and central control are possible. A decrease in (26) represents coordination within the system. This is some form of integration.

2.

$$H(X_i(t+1)|X_i(t)) \tag{27}$$

decreases. This means that the dynamics of the component i becomes more regular.

3. In §7 below, we shall see another instance where (25) decreases, complementarity.

The point we are going to explore is that such a coordination or regularization can provide the basis for a subsequent complexification and differentiation of the system, as argued in [1]. This can be iterated.

7 Complexity and differentiation

We need to distinguish two types of differentiation,

1. the internal or temporal differentiation of a component, and
2. the differentiation between the components of the system.

These two types of differentiation are in a certain sense complementary to each other. Temporal differentiation of the component i is considered via the quantity

$$H(X_i(t+1)) - H(X_i(t+1)|X_i(t)) = MI(X_i(t+1) : X_i(t)). \tag{28}$$

When the states of i are temporally independent,

$$H(X_i(t+1)) - H(X_i(t+1)|X_i(t)) = MI(X_i(t+1) : X_i(t)) = 0. \tag{29}$$

A step of temporal differentiation then makes $X_i(t+1)$ more independent of $X_i(t)$, that is, increases $H(X_i(t+1)|X_i(t))$ and thereby decreases $MI(X_i(t+1) : X_i(t))$. When i develops temporal regularities, (29) increases, because $H(X_i(t+1)|X_i(t))$ decreases as the present state contains information about the future states. This is a necessary preparation for an increase of the complexity $H(X_i(t))$ at later times insofar as it enables i to use information contained in its state at time t to develop structure at time $t+1$.

An aspect not pursued here is the extension of correlations to longer temporal scales that was proposed in [12] as a mechanism for increasing the complexity of social systems.

We are interested in differentiation between the components of the system. This might require an increase in the complexity $H(X_i(t))$ and therefore possibly a reduction of the temporal differentiation of the individual components, as high complexity can typically only be maintained on the basis of structural regularities.

We now consider a sequence of steps from a simple and undifferentiated system to a complex and differentiated one:

1. The complexity $H(X_i(t))$ or the number of the components i increases. When they are mutually independent

   $$\sum_i H(X_i(t)) - H(X(t)) = 0. \tag{30}$$

   When they become correlated, (30) increases, because the entropy $H(X(t))$ of the total system decreases. [3] developed this into a theory of pragmatic structuring. Also, this can be iterated at different system levels. This issue has been systematically analyzed with the concepts of information geometry in [4].

2. The first differentiation step then makes the states $X_i(t)$ of the components i different from each other, without necessarily increasing their complexity. Therefore, $H(X(t))$ increases, thereby decreasing (30). This simply means that the redundancy present in the system through similarities between the components is used up to render the system more complex, as described in [2].

3. The next step uses the rest of the system in order to pass information from $X_i(t)$ to $X_i(t+1)$. This increases (10), that is,

   $$H(X_i(t+1)|(X\setminus X_i)(t)) - H(X_i(t+1)|X(t)) = MI(X_i(t+1) : X_i(t)|(X\setminus X_i)(t)). \tag{31}$$

This then decreases (12), that is,

$$MI(X_i(t+1) : X_i(t)) - MI(X_i(t+1) : X_i(t)|(X \backslash X_i)(t)). \qquad (32)$$

This quantity can even become negative, as discussed above. This is the situation of complementarity between i and the rest of the system.
In particular, we again see a decrease in (25), but this time this is somewhat more subtle than the instances discussed in §6, coordination and regularization.

Thus, in the course of the differentiation process, (32) can oscillate between large positive and negative values. A differentiated system state might then correspond to a value of (32) that is small in absolute value, but with the terms $MI(X_i(t+1) : X_i(t))$ quite large, that is, with complex components that pass a lot of information along in time.
An example where the relevance of the quantity (32) can be seen is economic exchange. In a situation where the state $X_i(t)$ of an agent i at time t records that i possesses an item A and needs another item B. When i chooses not to produce A himself, but rather tries to obtain it via exchange from another agent, his state transition $X_i(t) \rightarrow X_i(t+1)$ becomes more dependent on the state $(X \backslash X_i)(t)$ of the rest of the system as he needs a trading partner, another agent j willing to exchange the item B that he possesses against A. This is a case of complementarity, meaning that (31) is high and (32) low. In contrast to a simple exchange economy, in a money economy, this dependency on the state of the rest of the system is reduced. i now only needs that there is somebody willing to buy A and another agent ready to sell B. Thus, the system state needed to affect the transition from $X_i(t)$ to $X_i(t+1)$ is less specific, and therefore, (31) gets reduced, and (32) can increase. In a credit economy, this effect is even further enhanced because then i need not first sell A to obtain the money for buying B, but can rather buy B on credit. Thus, the required system is even less specific, and the achievement of the desired system state $X_i(t+1)$ becomes less dependent on $(X \backslash X_i)(t)$.
Thus, when passing from a primitive economic system where each agents produce what she needs, to an exchange economy, and then to a money and credit economy, we first see a decrease in (32) and then a subsequent increase. Therefore, such a decrease in (32) can represent an intermediate step in a transition to a higher level of system complexity.

8 Kinetics

The question of how the system can sustain its complexity increases needs to be addressed, that is, how an integration mechanism like coordination, regularization or complementarity can provide the resources for a subsequent complexity increase. When we look at hierarchical differentiation, or the one between center and periphery, we can discern some duality. $\star$, the top level in the hierarchy, or the center, provides the rest of the system with information in the sense that

$$H(X_i(t+1)|X_\star(t)) \tag{33}$$

is small. In turn, however, the rest of the system provides $\star$ with energy or material. Therefore, the information flow from $\star$ into the system is complemented by an energy flow from the rest of the system to $\star$. The point then is that the information flow from $\star$ can make the energy production in the rest of the system, and the flow into $\star$, more efficient. The usual explanation given for this phenomenon is in terms of economies of scale, that is, that a differentiated production of larger quantities becomes more efficient than the parallel production of many small quantities. In our framework, we can analyze this phenomenon further. We separate the production effort from the control or coordination effort, and we express the latter in terms of entropies, that is, quantified information. The point then is that this information need not increase at the same rate that the number or size of the controlled components is increasing. In that manner, a more complex system can become more efficient than an independent collection of small simple systems and free resources for additional complexity increases.

What is missing then is the kinetic theory for the energy or material flow in the system. (In fact, this is entirely absent in Luhmann's theory; for example, in his analysis [15] of the economic system, he only considered the flow of payments, which would be some kind of information flow in our setting, but not the flow of economic goods or services in the reverse direction.)

9 Conclusion

In a segmented society, the units operate independently and in parallel, and the integration measure (11) is very small. Hierarchical stratification or any other organization with a central control $\star$ introduces correlations between the components and increases the integration (11). The conditional integration (22) gets small when all information is channeled through that central control. Division of labor leads to mutual dependence in terms of complementarity between subsystems and thereby decreases (11) and, more importantly, also the measure (12)=(13). In particular, the central control disappears. When the subsystems strive to regain some degree of autonomy and reduce the dependence on the other subsystems, this then increases (12) again. Thus, functional differentiation is characterized by a comparatively small value of (12), representing a balance between correlations and control on one hand and dependence through complementarity resulting from specialization. At the same time, (6) is large, expressing relative autonomy of the subsystems.

References

[1] F. M. Atay, J. Jost, On the emergence of complex systems on the basis of the coordination of complex behaviors of their elements, Complexity 10,

17–22, 2004

[2] H.Atlan, Self creation of meaning, Physica Scripta 36, 563–576, 1987

[3] N.Ay, An information geometric approach to a theory of pragmatic structuring, Ann.Prob.30, 416–436, 2002

[4] N.Ay, Information geometry on complexity and stochastic interaction, MPI MIS Preprint 95, 2001, to appear in IEEE Trans.Inform.Th.

[5] N.Ay, E.Olbrich, N.Bertschinger, J.Jost, A unifying framework for complexity measures of finite systems, accepted for ECCS 06

[6] N.Ay, D.Polani, Information flow in causal networks, SFI Working paper, 2006

[7] N.Bertschinger, E.Olbrich, N.Ay, J.Jost, Autonomy: an information theoretic perspective

[8] N.Bertschinger, E.Olbrich, N.Ay, J.Jost, Information and closure in systems theory

[9] E.Durkheim, De la division du travail social, Presses Universitaires de France, 1930

[10] S.J.Gould, The structure of evolutionary theory, Harvard Univ.Press, 2002

[11] S.J.Gould, E.Vrba, Exaptation – a missing term in the science of form, Paleobiol.8, 4–15, 1982

[12] J.Jost, Formal aspects of the emergence of institutions, Structure and Dynamics: eJournal of Anthropological and Related Sciences Vol.1, No.1, Article 6, 2004, http://repositories.cdlib.org/imbs/socdyn/sdeas/vol1/iss1/art6

[13] N.Luhmann, Zweckbegriff und Systemrationalität. Über die Funktion von Zwecken in sozialen Systemen, Suhrkamp, Frankfurt/M., 1973

[14] N.Luhmann, Soziale Systeme, Suhrkamp, Frankfurt/M., 1984

[15] N.Luhmann, Die Wirtschaft der Gesellschaft, Suhrkamp, Frankfurt/M., 1988

[16] N.Luhmann, Die Gesellschaft der Gesellschaft, Suhrkamp, Frankfurt/M., 2 vols., 1997

[17] D.MacKay, Information theory, inference, and learning algorithms, Cambr.Univ.Press, 2003

[18] F.Nietzsche, Zur Genealogie der Moral, in: Werke II (K.Schlechta, ed.), Hanser, München, ^{6}1969, pp.761–900

[19] U.Schimank, Theorien gesellschaftlicher Differenzierung, Leske + Budrich, Opladen, ^{2}2000

[20] M.Shpak, P.Stadler, G.Wagner, J.Hermisson, Aggregation of variables and system decomposition: applications to fitness landscape analysis, Theory Biosc.123, 33–68, 2004; erratum in Theory Biosc.124, 87-88, 2005

[21] M.Shpak, P.Stadler, G.Wagner, L.Altenberg, Simon-Ando decomposability and fitness landscapes, Theory Biosc.123, 139–180, 2004

[22] H.Simon, A.Ando, Aggregation of variables in dynamic systems, Econometrica 29, 111–138, 1961