Chapter 1 – Niklas Luhmann's Social Systems Theory – Concepts

This chapter introduces the concepts of Luhmann's Social Systems Theory. Luhmann's books provide explanations of the central concepts of the theory; three of them present comprehensive definitions (Luhmann 1995, 2007, 2013). In the following paragraphs we use definitions taken from those three sources.

Luhmann provided definitions for a number of concepts that already existed; the concepts therefore acquired meanings they did not previously have. In this chapter the concepts are presented in simple terms in glossary format. In subsequent chapters, where the concepts are used in the discussions, richer and more detailed definitions are given.

To begin with, we need to consider that the overall architecture of the theory is based on the distinction of autopoietic and non-autopoietic systems. Among the autopoietic system we find three types of systems: the biological, the psychic and the social system. Luhmann's sociology is obviously concerned exclusively with social systems. The details come next.

System Luhmann says: "system is the difference between system and environment" (Luhmann 2013, p. 44). This definition gives us the key notion that the system–environment distinction is the fundament of the creation of a system. The systems appear and are therefore observable as a result of this distinction, by which the observer can assign the place of observation either inside the system or in the environment. In other words, this distinction is the key constitutive step of social systems – systems come to exist as distinct from environment. A social system does not appear without its environment; for instance, health systems are orientated towards diseases happening in its environment. Once a disease is detected, communications about it become part of the health system. A system can internally consider its constitutive distinction, making for itself an internal representation of the environment that concerns

it. Within the environment, anything irrelevant for system's operations is not a matter of concern for the system. The system can also recognize the presence of other systems in the environment.

Autopoiesis Imported from the works of the biologists Humberto Maturana and Francisco Varela (1974) and applied by Luhmann to Social Systems, autopoiesis is understood as the driving basal orientation of social systems. It denotes the condition of existence of social systems, consisting in self-reproducing through the means they themselves produce. A social system either performs its autopoiesis or does not exist as a system; and no system can take care of the reproduction of any other system. Social systems are therefore orientated towards assuring the preservation of their self-reproduction, which implies preserving the consistency of generating capabilities of the means of reproduction. Autopoiesis is also a permanent drive, performed at each operation of the system, and no system can afford to "take a break" in that regard. Where there is a social system, autopoiesis is at work at all times.

Communication Communication is, so to speak, the "building block" of social systems. Luhmann uses the term communication with meanings distinct from the traditional transmission model of communication, where "Alter" sends a message that is then received by "Ego", thus completing the communication link. For Luhmann, in contrast, communication has three components: content (information), utterance and understanding. Utterance is the physical emission (sound vibration, visible printed characters, light signals, tactile braille, etc.), and content is the conveyed information (within a given shared semantic universe of significances). Understanding is the unit of content and utterance, and is what make possible the interlacing and exchange with the subsequent utterances and contents. Without such interlacing, which may consist for example in a request for clarification, we cannot say that communication has occurred. In this sense, understanding includes misunderstanding. In Luhmann's terms, social systems are made up of communications and nothing else; in other words, without communication there is no social system. Communication has attributes of recursive confirmation and self-validation, interlacing past, present and future communications, stabilizing meanings and the systems that rely on them. Furthermore, for Luhmann society is the totality of the communications taking place, and no communication takes place outside society.

Contingent and double contingency Contingent means something that could be different, i.e. in its current forms and attributes it is neither necessary nor impossible. All communications happen in a condition of double contingency, which means that on both sides Alter and Ego, while engaging in communication, know that they are performing their own selections over what is said, listened and replied to. The selections Ego makes can be different from Alter's and vice versa. The selections include the possibility of rejection of the communicated messages. All social systems have to live with the improbability of understanding, given the contingent selections, and also the open possibility of rejection.

Operational closure A key-structuring feature of a social system is the closure; a system can only deal with the information it internally produces. No information in the environment can be taken in directly by a system. What is observed in the environment becomes information once the perceptions are selected and internally processed. Operational closure correlates with autopoiesis in the sense that the selection and validation of observations and communications are prerogatives (and survival matters) of the system, and no other system can insert information inside another system; in the same way, a mind (a psychic system) cannot put a thought inside another mind (another psychic system). If that was possible, the boundaries of the two systems would collapse and the system/environment distinction would no longer be valid. Operational closure nevertheless allows for a system's capacity not only to observe itself but also to observe another.

Three types of social systems In his grand sociological theory, Luhmann differentiates three types of social systems: function systems, organizations and interactions. Each type is defined in terms of the specific ways they communicatively operate within their closure. *Function systems* are based on specific binary codes of communications (see next point); *organizations* are systems based on membership and decisions (a specific type of communication); *interactions* are short-lived systems constituted by face-to-face communications. A function system, like the health system, does not exclude anyone in the society as either user or provider can take part in it at some point. Differently, organizations select those it can identify and can identify themselves as members (employees for instance), and only members can take part in valid decision-making communications. Interactions are communication systems that develop where two or more people meet (face-to-face or virtually) and once the communication is

over the system ceases. Obviously society members may take part in several social systems, sometimes simultaneously.

Binary codes Binary codes are essential for identity building of *function systems*. We do not refer here to the other two types of systems as explained previously: *organization* and *interaction*. Each *function system* has its own specific binary code (legal/illegal, healthy/sick, government/opposition, art/not art, etc.). The binary codes mark and are fundamental to all communications as communications of the specific *function system*. No *function system* has the capability and legitimacy to use the code that belongs to another *function system*. The society does not recognize a communication stating that something is legal or illegal if such communication comes from any *function system* but the system of law. The words legal and illegal can be used in any circumstances but they will not carry the weight, legitimacy and consequences of, for instance, an adjudication of a court of law. Society members with professional attribution inside the health system would clearly distinguish the communications on health matters that are valid. *Function systems* are therefore based in simple binary codes that nevertheless can provide infinite possibilities of ramifications in increasing levels of complexity.

Structural coupling Complementary to the concept of operational closure, structural coupling recognizes the possibility of systems observing each other, and by doing so achieve some level of coordination, nevertheless keeping their closure. In Luhmann's terms, systems organized under their respective constitutive closure do find ways to observe others and by doing so "irritate" or "are irritated by" the other, creating expectations and reacting to other systems without losing their distinctive separation from the environment and the other systems in it. *Structural coupling* is the term Luhmann uses to describe such operations; it allows for coordination between different social systems, like health and education, health and law, etc., as each system can observe the others, selecting what is relevant for each, and in the process operate in a coordinated manner as seen from the observer's point of view.

Social differentiation In Luhmann's grand theory, the current stage of evolutionary transformation of societies' structures is characterized by the existence of several operationally closed and differentiated *function systems*. From previous stages of *segmented* and *stratified* societies, historical evolution has arrived at the advanced stage of social differentiation that started in the eighteenth

century. Differentiated *function systems* (law, politics, economics, health, religion, art, media, etc.) create addresses whereby inclusion of individuals is open to all society members, preserving the possibility for anyone to be concurrently included to varying degrees in several function systems. Differentiated function systems strive to keep their characteristics and specificities with socially precisely recognizable boundaries between them, mainly based on their binary codes and communications. At the *segmented* society stage, individuals were assigned to social structures by their place of birth or life. *Stratified* societies were more complex and individuals were assigned to a stratus the society created (such as castes, social classes, artisanal groups, etc.). In the modern society, differentiated in *function systems*, no function system has the central role and preponderates over the others; even the political functional system is one functional system among the others.

Distinctions and observations An important turn in the development of Social System Theory is Luhmann's incorporation of the works of the mathematician George Spencer-Brown. In the *Laws of Form*, Spencer-Brown (2015) asserts the inseparability of observations and distinctions. To make observations, one needs to draw distinctions. Distinctions are forms with two sides: the marked and the unmarked. A distinction is thus a unit of difference. Observations are made according to the adopted distinctions. To carry out an observation, the observer takes the marked side, which is of interest – for instance, disease instead of health – and leaves the rest out, on the unmarked side. These are essential selection processes by which systems can become self-referential, self-observing and self-organizing (see these definitions later). Furthermore, all observation has blind spots[1]; the blind spot is often the deployed distinction itself. A distinction can also be observed, but that requires another distinction assigning the distinction to be observed to the marked side.

Second-order observation Another important reference in the development of Luhmann's theory is the works of Foerster (2014), who conceptualized the cybernetic qualitative step by which an observing system can observe observers and, respectively, the distinctions they use and the observations they make. This includes self-observation. All social systems perform first- and second-order observations.

1 Observers need to distinguish themselves from what they observe, in doing so, to carry out the observation, they stay in the "blind spot".

Coding and programming Coding is the way of orientating communication. Binary codes offer the two sides of the distinctions the communication refers to, while always electing one side. Programmes guide the selection of themes and semantics, supporting communication connectivity within the chosen side of the binary distinction.

Complexity Complexity is a feature of observation, not a reality in itself. It denotes that both a system and its environment have elements and relations between elements that surpass the system's observation capacity. A system has to reduce complexity, by making selections and focusing on relevant internal and/or external elements, excluding the rest from consideration. Complexity therefore refers to the unavoidable limits in the capacity of making observations. Contingency is intrinsic to complexity, because while reducing complexity a system has to make selections, which by their turn could be different, i.e. are contingent. The environment is more complex than the systems.

Having briefly described these central concepts, there is moreover a relevant set of specific concepts. Luhmann's studies in the field of management science have attracted attention among those interested in organizations as social systems. Health systems have many organizations and therefore it is relevant to understand how this type of system (organization) operates. The conceptual understanding of organizations includes the following:

Decision Decision is a particular type of communication crucial for the existence of organizations. Without decision there is no organization. All matters of concern for an organization as social system are objects of decisions. A decision communicates the side of the distinctions the subsequent communications should be connected to, therefore orientating and creating the premises for the decisions to be taken next.

Membership The differentiation between who belongs (members) and who does not belong to an organization is of vital importance for any organization. All organizations are based on decisions and membership, and only members take legitimate decisions, or, in other words, only decisions taken by members are recognized as valid and relevant for the organization.

Uncertainty absorption A decision already made does not have to communicate the uncertainties that surrounded it before it was taken, which include the ambiguities or doubts in relation to the evidence and inferences guiding the

decision. A communicated decision does not communicate that it is also contingent, i.e. the possibility that the decision could be different. Once adopted, a decision thus absorbs uncertainties, which are then excluded from further consideration. The absorption of uncertainties achieved by decisions is therefore crucial for the management of any organization.

Decision premises Decision premises is a functional characteristic of decisions that make former decisions the justifications for subsequent ones, thus becoming the basis of organizations' life. Decision premises help to solve the paradoxical nature of any decision; the paradox of undecided decisions which derives from the fact that if an option is obviously better than all the others, there is no decision to be made; conversely, if all options are equally good, a decision cannot be made. So, decision premises attract and justify the choice in terms of the decisions previously made. Luhmann identified three types of decision premises: programmes (with stabilized sets of prescriptions and expectations), personnel (with stabilized definitions of members' roles) and communication channels (also understood as organizational culture, setting courses for routine formal and informal communication flows). Decisions based on decision premises have the advantage of not having to exhaustively demonstrate the reasons for taking them.

There are a number of additional important concepts with rather complex formulations. They are briefly mentioned and explained as advanced topics in the Annex.

Having presented the central concepts of the Social System Theory, the next section discusses the differences between Luhmann's formulations and the more traditional views that Luhmann's theory does not incorporate. Three topics are discussed: 1) the *input–output model*; 2) the notion of a *system as bigger than the sum of its parts*; and 3) the concept of the *adaptive system* and its evolution.

These topics are not required for understanding Luhmann's theory and therefore readers, particularly those following health systems thinking approaches, may want to be aware of the reasons behind Luhmann's theoretical decisions, and their implications. Those who are only familiar with those commonly held views about systems are particularly encouraged to read the section, and familiarize themselves with Luhmann's explanations for not including these views in his conceptual architecture. The discussion is rich and instructive, and builds on a number of the concepts presented in the first section of this chapter.

1. The input–output model

"The systems thinking approach goes beyond this 'input-blackbox-output' paradigm to one that considers inputs, outputs, initial, intermediate and eventual outcomes, and feedback, processes, flows, control and contexts" (WHO 2009, p. 34). In common with this understanding, Luhmann's theory moves away from the input–processing–output model predominant in the 1950s and earlier.

In his theory, however, inputs are not taken in, processed and the resultant "products" sent to the environment as outputs. Inputs are creations of the system. The system does not take in what it observes in the environment; it interprets and translates what is observed into meaningful information, consistent with the meanings the system has at its disposal and works with.

Luhmann says that there is no information as such in the environment. Information does not penetrate the system; it is instead generated inside the system. Surely, though, a system relies on the sources of energy and materials in the environment. This, however, is a basic distinction that needs to be considered, separating the physical world and the world of communications and meanings.

For any biological organism, energy and materials do enter the body, and the waste is subsequently thrown out. The same does not happen with information. As said earlier, information does not enter; it is elaborated inside the system. A system may make available the results of its information processing to other systems; but it can neither insert the information into the other system nor enforce it, if the other system does not observe it or consider the information relevant.

Luhmann therefore uses the words input and output with very specific connotations; inputs and outputs are recognized as such inside the system, in the way it internally and communicatively processes the information it generates itself. These points have implications for the theoretical conceptualization of the closedness and openness of systems.

A number of questions therefore become pertinent. For instance: how open can a system be and still preserve its boundaries and identities? How close can a system be and still acquire energy, and make observations for its information processes without running the risk of being annihilated by the environment?

The theory offers the following considerations. Systems can only translate what they observe in the environment into information because they are operationally closed. If information could penetrate the system, it would be the

same as someone inserting their own thoughts inside someone else's mind, as mentioned previously. The distinction between the two minds would collapse and both minds consequently would have the same fate. Although they cannot take information from the environment, the systems can take energy and materials from it; they are therefore both open and closed.

Thus understood, operational closure, deriving from the basic system/environment distinction, is the very condition for the system to process observations and elaborate information. As the environment is incomparably more complex than the system, the system does not have what is conceptually denominated as "requisite variety" (Luhmann 2013, p. 121) by which it could establish complete representation inside the system, relating one to one all the elements in the environment. Therefore, the system needs to make selections of what it is going to process and observe in the environment. This reduces environmental complexities as far as the system is concerned, without affecting the environment, which remains as complex as before.

If these notions are accepted, it is understandable that the system would be destroyed by the overwhelming volume of information the environment could potentially generate. If information as such were coming from the environment and entering the system, the system would not be able to distinguish what belonged to it as opposed to the environment. Information generation therefore is a process in which systems are actively engaged as a matter of autopoietic survival.

Still in relation to the input–output model, the notion of "non-trivial systems" (Foerster 2014) is relevant to consider. A non-trivial system may process and generate diverse types of outputs with the same observation it makes in its environment. This system can observe itself, and therefore it can reflect on its own processes and results, and introduce changes, which as a result may render the generated outputs unpredictable for an external observer.

Compared to the simple rather outdated input–process–output model, these models of self-referred systems dynamics grant a conceptual architecture with better tools to approach the complexities of social systems such as the health system. Translating these understandings into observations of health systems' operations we can see as follows.

A case of a suspected disease remains an unknown event in the environment among all sorts of events taking place in it. Once it is detected by a health professional, which then communicates about it inside the health system, the suspected disease becomes a point of reference so that subsequent communi-

cations can take place in the system. Before it was detected, it was not information waiting to be disclosed; it was just an unknown.

Disease detection is carried out by procedures of observations whereby health professionals deploy the distinctions they were trained to use. The health professionals write the reports or communicate orally what they observe. From that moment on, the event "unknown disease" acquires the form of "symptoms and suspected diagnostic", which can be recognized, read, checked, registered, etc., becoming in the process a communicated reality within the system.

The information can then be retrieved, coded, double-checked, criticized, complemented, and so on, according to the programmes the system communicatively deploys. The former "unknown" therefore acquires "life", so to speak, as a disease inside the system. Effectively, by becoming communicatively inserted in the system, the disease actually becomes part of the system as a recognizable event/object.

The self-reference (including self-observation and self-organization) functionality of the system endows it with the capacity to adjust its observational capacities, refine its interpretation and service delivery capabilities, and proceed with the autopoietic reproduction of these abilities. All that will be in correspondence with the (internal and external) level of complexity the system can handle.

The reproduction of the system therefore can only be self-reproduction, because only the system understands and can use the communications that are its prerogatives. The input–processing–output model does not give a proper account of these highly relevant systemic functions.

2. The system is bigger than the sum of its parts

Systems are more than the sum of their parts is perhaps the most commonly held view about systems. At first glance, it seems there is nothing wrong with that, as the statement calls attention to functions emerging out of a collection of elements, which individually and separately do not show those specific functionalities.

However, this concept of the whole and parts cannot easily accommodate the notion that social systems are made of communications. While the idea of parts physically and functionally distinct and the whole that encompasses them is easily perceived in systems that have precise concrete existence, such

as ecological systems, nervous systems, urban transport systems and so on, it is less evident when comes to intangible realities, such as social systems based on communications. The whole and parts concept does not seem to give a good account of the set-up of the system.

The argumentation that follows looks at the *whole/parts* from the Social Systems Theory perspective, which includes discussions on the validity of the scheme considering that: 2.1) systems are made of communications; 2.2) communications cannot be summed up like "bricks"; 2.3) communications interlace each other making stable "meaning pools";[2] 2.4) communications not only make the system but also reproduce it; 2.5) communications make the internal differentiations and sub-systems; 2.6) communications differentiate a system from other systems in its environment; and, finally, 2.7) the *whole/parts* scheme is analysed as a potential observation device deployed by a system for self-reference and self-organization.

2.1 Systems as communications

Questions concerning a system's parts and whole acquire new difficulties when we consider that systems are made of communications, as proposed by the Social Systems Theory. How can parts of a system add up to a whole if they are identified as communications?

A nervous system can be dissected and its parts isolated even in organisms still alive. If a part is severed, the reaction at the level of the whole system can be monitored. The parts can be identified and counted; the connections between them can be cut or preserved. Some parts may play a more important role than others in terms of the overall emergent functionality of the system.

However, if we take communications as the parts of the system, a piece of isolated communication no longer constitutes part of the system, because what makes a communication part of the system is the fact that it is connected to other communications belonging to the system; the connectivity is crucial.

Communications are events connected to other communicative events. If a nurse observes that the temperature of the patient is rising and does not communicate that to the other nurses and doctors, the communication event that does not happen has not become part of the system.

But the system does not have an inventory of communications already established, representing the aggregation of all parts of the system. Commu-

2 These terms, "bricks" and "meaning pools", do not come from Luhmann.

nications cannot be dissected and counted. The virtual inventory of communications, if we may keep this expression, is being built as communications happen, as opposed to the stable list of components of a system with discrete countable identifiable parts that can be regularly inventoried.

The observation the nurse had made may never be incorporated into the universe of events constituting the health system. Similarly, a communication may happen and remain isolated, as for instance a note in the patient's dossier that was never considered and ended up forgotten in the archive (or database) of the hospital among hundreds of thousands of other records. In spite of communications that may not happen or have been forgotten, the system carries on its autopoiesis with all the other communications taking place. No single communication has the crucial role for the existence of a social system, as is the case in certain systems with core function such as the nervous system.

Unlike a system with a countable limited number of components, a system made of communication has an unlimited number of components and cannot correspond to any prescription of amounts of communications that need to take place and the time for that. In communication-based systems, the parts cannot be assembled together in a predefined permanent, stable combination, although routine communications may occur regularly. New communication parts will always be required.

These explanations should give a clear understanding of the specificities of social systems. Although themes and topics (such as those in the universe of specific medical specialties) create sets of communications related to each other and recurrently claiming continuity and consistency, the field of communication has to remain open because of both the nature of communication as events in time and the limitlessness resulting from the complexity of human health. Some additional considerations are still pertinent in relation to communication and the parts/whole scheme.

It must be clear now that a communication-based system is not composed of the elements communicating with each other inside it. It is not meaningful to count the communicating elements as if they were the individual parts of the system. We may see communication as emerging (or otherwise not existing) and also comprising a network-like web of recurrent and forward-moving connected communicative events, reproducing themselves, in potentially infinite forms, in the system being built by them. Communication emerges from communications, not from the material base where it happens.

In view of that, the incongruity of comparing the constitution of a whole with the sum of its parts becomes clear. With some irony, scholars who stud-

ied Luhmann's work (King and Thornhill– 2003, p. 10) say: "the whole is less than the sum of its parts". This may be understood from the perspective that at any given moment, only a limited number of communications actualizing the system is in fact occurring, compared to the countless repertoire of possible communications and the incalculable numbers of previous communications of the same system.

The point to be retained in this discussion is that the system is comprised of the communications that are making it, and communications are continuously linking past communications with current and future ones. Only communications can make a system perform its communicative functions. In Luhmann's words "only communication can communicate" (Luhmann 2002). In the case of health systems, this means communicating and reproducing the healthy/sick binary code and the programmes sustained by it over the course of communicatively realized diagnostics and treatments.

Furthermore, the whole of the system cannot be equated to the addition of communications that have already taken place. Past communications only contribute to building the system if they are currently linked to actual on-going communications; otherwise they are irrelevant. The system's whole would thus be an artificial abstraction of the entire set of communications taking place in each given moment. Such abstraction is of no relevance.

2.2 Building blocks metaphor

Not all individuals identifiable as operating within the system are actually engaged in communications at all times. While the majority may remain "silent" and a small minority actually communicates at any given time, this does not mean that the system becomes smaller or bigger because there are fewer or more silent individuals.

The metaphor of "building block" implies the wrong impression of quantifiable accumulative unitary bricks. Communication cannot simply be reduced to countable discrete units of exchanges between individuals. Communications are more complex than that; as already mentioned, they imply connectivity with other communications.

It does not make sense either to compare parts of the system without communications, and parts of the system with communications, because there is only the system where communications occurs (whatever the time required). The parts of a communication-based social system are the communications

taking place, not the members that could possibly communicate with each other.

The system does not include the communications that did not occur, even if they remained potential and plausible. Possible communications are not retrievable as recorded memory of the system and cannot interlace new communications. Consequently, if one still wants to use the simplistic "building block" metaphor, one needs to focus on the actual communications happening inside the system at any given time.

Of course communications vanish as soon as they happen, but this does not mean that systems disappear with them. Elements remain latent whether communications are taking place or not. For instance, the semantics, codes, grammars, syntaxes, symbols, semiotics, associated meanings, memories, etc. deployed during communications are kept as reservoirs of shapers of an infinite number of possible communications. The operators inside the system draw from those common stable memory pools the rules and signifiers needed to compose their messages, enabling them to be understood by the recipient of the messages, who uses the same pools for communicative purposes. Individuals' insertion in a system is preceded and maintained through continuous learning of those pools, allowing them to engage in meaningful communications inside the system. Memory plays a crucial role in this.

However, anything in those reservoirs that is not used becomes irrelevant as constitutive of the system. In the same way, information a health worker does not communicate back to the system's counterparts cannot be considered part of the system. The reservoirs are not accumulation of "building blocks"; the system thus should not be confused with those reservoirs and potential communications. A social system made up of communication does not have to make exhaustive uninterrupted use of all communications possibilities that can be produced with what is available in the pools. If that were the case, the system would be destroyed instantaneously. For that matter, the system makes selections. Any communication implies selections from both sides (sender and receiver), and cannot be otherwise. Selections are an inextricable part of communications, and therefore also constitutive operations of the system.

2.3 Interlacing communications

As explained in the first section of this chapter, communications are comprised of utterance, content (information) and understanding (including misunder-

standing).[3] They interlace with components of other communications, becoming part of a system. In isolation, an utterance or content is not integrated into a system. Communications always have the possibility of interlacing other communications on the accepting (yes) or the rejection (no) side, and through the content and/or the utterances. The recipient may emphasize either of these components while participating in communication.

By interlacing with others, a communication becomes part of meaningful sets of communications. The interlacing is thus fundamental to keep the system alive, and involves connectivity through both utterances and content. What we are trying to say here is that the importance of interlacing makes the cut-off of communication in discrete pieces irrelevant. It is rather artificial to cut a communication with the purpose of counting the numbers when the very fact of being meaningfully interlaced is the crucial aspect of communication.

2.4 Autopoiesis

Taking autopoiesis of systems into consideration, obviously a system does not carry out its self-reproduction without the concurrence of its parts. Clearly, when we speak about parts here we are talking about communications, nothing else. So, there would not be autopoiesis if the system did not reproduce the distinctive communications that characterize the system.

The communications are thus making the system, and at the same time realizing its autopoiesis. The communications thus reproduce previous communications and open connections for the subsequent ones. In this way, standards and semantics are meaningfully preserved and reproduced, assuring the continuity of the communicative operations of the system. Social systems reproduction is achieved through reproduction of its communications; communications reproduce communications, we may say.

2.5 Internal differentiation

Let's consider now that communication can become specialized and differentiated inside a system. There may be sets of related communications, using specific semantics and selections, which may become characteristic of some

3 Luhmann says "understanding" is the unit of the distinction between "utterance" and "content".

identifiable sections, programmes or sub-systems of the encompassing system. These sub-systems become constitutive parts of the system, making it more complex.

A way of approaching the system/sub-systems distinction may use the concepts of integration and differentiation. Differentiation is one of the central concepts of Luhmann's theory; it refers to both the differentiation of systems in the same environment, and the internal differentiation of components inside a system. Integration, on the other hand, is not a key concept in the theory and Luhmann summarizes it as reduction of degrees of freedom of any system, while entering structural coupling with other systems.

Differentiation across systems is an outcome of the autopoiesis processes each system undertakes, while preserving their operational closure. By making their individual reproductive selections, any system becomes increasingly different from the others also performing their own autopoiesis in the same environment.

Internal differentiation, on the other hand, has to be understood as part of the individual dynamic process of autopoiesis itself. The system creates internal divisions in its processes of reproducing itself.[4] Internal differentiation increases a system's complexity and consequently the complexity the system can address in its environment.

A system may have internal parts involved in the processes of communicating the core business of the system (for instance, diagnostics and treatments in the health system). But it may also create parts specialized in self-observation of the system (monitoring and evaluation of health programmes for instance). The system may also have sections specialized in observing the environment (epidemiological surveillance and infectious diseases controls), and sections fulfilling normative roles, assessing if internal communications and actions correspond to the patterns and standards the system adopts (professional councils and ethical boards, for example).

Internal differentiation does not imply the creation of a distinct part with the attribution to rule the entire system on all matters. Some rules may be enacted, but only become effective when the parts incorporate them. The parts themselves perform coordinated acts to a large extent independently from central observation and control. Just think about how patients' referral systems may operate across health facilities, with facilities communicating

4 Reproducing internally the system/environment distinction.

with each other, sending and receiving patients without resorting to any centralized mechanism of authorization or subordination.

What should be retained from this discussion is that a system surely allows for the development of differentiated forms of communications between its components. But, in short, it is not relevant whether internal differentiation makes the whole bigger than the sum of its parts or otherwise. What matters is that it makes the whole internally diverse and better able to deal with the complexities of the environment, while preserving its unity and maintaining the distinctive character of the internal communications.

2.6 System–environment distinction

Perhaps crucial for understanding how Social Systems Theory deals with the parts/whole scheme is the notion that the system is, in Luhmann's words, "the difference between system and environment" (Luhmann 2013, p. 44). A system exists from the moment it becomes distinct from its environment. As a consequence, the maintenance of the system implies the maintenance of its differentiation vis-à-vis its environment.

The reproduction of a social system is the repetition of its system/environment difference and respective limits inside the system in all new sub-systems created, which individually repeat the original distinction, whatever specific additional functionalities the sub-systems may acquire. This is an advanced complex topic we can deal with later.

But, in this sense, a system is found where there is a system/environment distinction, not where there are parts comprising wholes. Social Systems Theory thus replaces the paradigm of the whole and parts difference with the difference between system and environment.

Parts can be arbitrarily aggregated without constituting a unit with a singular self-reference, clear delimitation in relation to the environment and without a cohesive process of self-reproduction and self-preservation. Think of a crowd in a busy train station – the collection of individual-parts does not make that a system.

To further illustrate this conception consider, for instance, a large bazaar with hundreds of independent stalls selling thousands of different products to thousands of buyers on a daily basis. If we take the whole and parts paradigm, we can say that there is a system here: the bazaar is the system and the shops its parts.

However, if we look at the collection of shops, consumers and products thus described and try to understand it from the point of view of the system/environment distinction we reach a different conclusion. We would have difficulty finding the operations the bazaar performs that sets the boundaries between it and what can be identified as its environment. If we exclude the managerial services of the bazaar, in charge of keeping the structures in working condition and renting out the stalls, there are no other overarching functions bringing a sense of unit.

As we see when we visit a bazaar, there is space for tens of thousands of daily buying/selling transactions. The administration of the bazaar can be identified as an organization (a type of social system). The stalls are not part of that organization. They act independently, renting the space and conducting their buying and selling operations as autonomous entities. Some coordination for determination of prices and selection of products may or may not happen; it will largely depend on the individual decisions of the stalls and administrators. However, they do react to what they observe taking place around them.

In this sense, the bazaar would rather be the environment where customers appear, observe products and prices and are observed by the stalls. Negotiations may happen and may evolve into a final selling and buying transaction. An analogy can be made, comparing customers in the environment of the bazaar and patients coming from the neighbourhood to get treatment in a primary healthcare centre. The communications involved in selling and buying in an economic system are functional equivalents to the communication operations involved in treating and receiving treatment in a health system. In both cases the system/environment distinction is more relevant in analytical terms than the parts/whole distinction, because the system configures itself vis-à-vis and in contrast with its environment.

In radical contrast with the whole and parts paradigm, the analysis from the point of the Social Systems Theory would conclude that the bazaar is in fact the environment instead of a system. Furthermore, we can also say that the bazaar does not perform autopoiesis; the environment does not do self-reproduction. Autopoiesis is a specific attribute of the systems.

2.7 Parts, wholes and observation

A final consideration on the parts and wholes topic suggests that the scheme can be understood as an observation strategy. In this sense, there is no ontolog-

ical or technical prescription setting the sizes and limits of the parts to be considered. The questions about how big/small a part should be to be considered as a part do not have a definitive answer. The observer draws the distinctions, makes the definitions and proceeds to observe, describe and count the parts, according to his/her decisions, which are all contingent (i.e. could be different).

In object-orientated ontology, Graham Harman (2018) says an object can be approached and described by cutting it into its component parts or by relating it to other objects. These approaches can also be employed jointly. The limit is that the object can never be apprehended in itself. Parts and wholes are thus deployed observational techniques. The question whether the whole is more than the sum of its parts depends on the understanding of the attributes being observed and their relevance or irrelevance as attributes of the parts or of the whole. So, the results of cutting into parts are contingent, and do not reveal the essential nature of the object.

We think these conclusions sit comfortably with the conclusions Social Systems Theory suggests to us. If the parts of the object system are communications, they cannot meaningfully be summed up to the totality of a single whole that is graspable as the defined object within which all its parts are accommodated. The system object would not be apprehended in itself.

Furthermore, for observation purposes, the notion of parts implies the assumption that they are homogeneous or at least have common uniform qualities that all parts show in some strictly comparable manner. But, as Luhmann (1995, p. 7) says, "there were hardly any theoretically proven criteria for homogeneity".

Hospitals, public health programmes, professional bodies, etc. can be considered as sub-systems of a health system. For that, they may only have the fact of being orientated towards and mobilized by the communication concerning healthy/sick binary code in one of its multiple presentations and focus. In other words, the fact that these sub-systems are included in the universe of communications recognized as exclusively belonging to the health system attests to their insertion in the overall health system.

Therefore, sub-systems are part of the health system as legitimate makers of health communications, regardless of the size or quantity of each subsystem's communications. The semantic order the sub-system is inserted into does not require specific homogeneity of their messages, or specific sizes of their sentences, or specific effects of their utterances, or specific complexity of the information conveyed, or specific quantity of exchanges, or comparative relevance or frequency in relation to communications of other sub-systems.

Any measurement of communications in such ways would be meaningless. The semantic order, by means of autopoietic self-reference and self-reproduction, attests the validity of the communications as part of the system.

For its self-reproduction, the system may see advantages in adopting the parts/whole observation scheme, when the technique may facilitate the system's internal communication about itself. In this process, semantic spaces are created, configuring sub-systems, as might be the case for example in the creation of communications specific to medical specialities. The self-reference thus achieved and maintained contributes to the autopoiesis of the system. The question whether the whole is or isn't bigger than the sum of its parts remains irrelevant from the system's self-reproduction and self-reference perspective. In its self-reflection, a system will look for ways of reducing its own complexities; the more complex a system becomes, the less it will try to create a copy of itself that reflects all its complexities, it will therefore settle with self-descriptions that summarize the main features the system is selectively electing as its priorities, vis-à-vis the environment from which it distinguishes itself. In conclusion, the parts/whole scheme should be employed according to internal self-observation aims, if this deployment is to have some advantage.

3. The notion of adaptive systems

Health systems are "complex adaptive systems" (Olivier, J. et al. 2017, p. 21). Throughout the health systems thinking literature one finds references as such to health systems as complex adaptive systems (CAS). This is a potentially controversial topic, if one considers what the Social Systems Theory says about the concept of adaptation.

If adaptation only means learning and incorporating new information into its communications, social systems are indeed adaptive; this is partially consistent with the Social Systems Theory. Adaptation in this case refers exclusively to internal changes in the system, without necessarily linking it with better chances of survival in a given environment. Learning is operationally achieved by the necessary connections between past, current and future communications, establishing new and more complex memories and repertoires of themes, which the system uses for processing information.

Nevertheless, if adaptive refers to adaptation to an external environment (however it may be defined), it is necessary to consider that the theory highlights the system's environment as a construction of the system itself; the sys-

tem selects in the environment the elements it observes and considers relevant for its autopoiesis, ignoring the rest. In the process, the system is aligned with the environment itself is recurrently confirming and constructing. The understanding of the founding role of the system/environment distinction in the theory is therefore crucial to approaching the adaptation issue.

The following example may help this discussion. Take a hospital as a system (an organization system in line with the theory), and its emergency ward overwhelmed by the number of cases arriving at its door. The hospital may decide to set up a triage procedure, by which cases are sent through to one of the following four options: immediate care; referral to more complex centres; stay and wait; or send home. Some patients may voluntarily select themselves out and go to another hospital or home. This may be the case on a particularly busy Saturday night but might also be a trend observed over a number of months. With the triage, the hospital is selecting what it is observing in the environment (its doorstep).

What an external observer may see as system adaptation to a higher demand (the triage), an internal one sees it as a selection operation. For the hospital, it is not of primary relevance whether the procedure represents an adaptation to the environment or not. The procedure is a solution to the pressure the hospital-system is internally detecting; the hospital needs to safeguard its internal processes and communications against possible collapse. The hospital is primarily concerned with the integrity of its communication, its autopoiesis. Intervention in the environment beyond its doorstep is not its concern. The environment at large may carry on producing more and more cases (or not); the hospital is neither more nor less adapted to the case production capacity and its determinants in the environment. It is only carrying on its operations within the limits it recognizes as important for its operations; it is adjusting internal communicative operations, not aiming at achieving better adaptation to the environment. As an organizational system, the hospital is selecting from its environment the elements that it considers of interest and finds itself capable of handling. The crucial point is that internal adjustments do not necessarily represent adaptation to the environment.

The hospital could have continued its internal routines without the triage it has set up, maintaining ad hoc spontaneous triage as before. If the new procedure is considered an adaptation, it was an "adaptation to itself", learning and incorporating new elements, as noted at the beginning. The adjustments were intended to achieve better coordination, releasing tensions and stresses among the staff, and pressures on hospital resources. If any adaptation is hap-

pening, it is to the internal environment of the hospital. It can still be considered that these changes could be conceived and incorporated anyway, independently from the pressures of the external environment. The hospital could set up a triage for the advantages in handling the arriving cases and for preparedness, in case it expected to face a possible surge in demand in the near future.

In more theoretical terms, in *La Sociedad de la Sociedad*, Luhmann (2007a) speaks of an excess of possibilities the systems internally develop and try, which are not related, conductive or justified by the necessity of adaptation to the environment. The independence of the systems from the environment, allowing freedom for internal communications, endows systems with the capability to develop new distinctions, new forms for new observations, new ways of communicating and so on. This wealth of possibilities and subsequent adoption of changes leads to differentiations inside and among the systems, which do not necessarily rely on some sort of "natural selection" to survive and reproduce. The emergence of the peacock tail could not be explained if all a system could do (biological ones included) was to strictly adapt to its environment.

Luhmann (1995) also says no system has the "requisite variety" to internally represent the environment in all its complexities and control the "natural selection" adaptation. He says: "The system can only compensate the lack of knowledge of its environment with intern excess of possibilities, to say, can compensate the lack of determination in the environment with its own lack of determination". (p. 182).

He continues by explaining that the capacity of the system to draw its distinctions and use them to make observations in the environment (and itself) is a function that does not have any correlate in the environment. Distinctions are not imposed by the environment and do not exist as such in it. The capacity to independently draw distinctions endows the system with the possibility to create an excess of possible internal elements and configurations that it can adopt regardless of what the environment has.

Furthermore, where the system decides to closely couple with elements of the environment (other systems, for instance, as the health system couples with the political, economic, legal, etc. systems), it does so with its operational closure and autopoietic reproduction, where it sees the advantages of doing so. In other words, it selects the way it couples with other systems according to the calculations and information processing it is capable of.

On the other side, the overarching, all-encompassing environment is supportive (i.e. not strictly selective) of many possible forms the systems living

in it autopoietically incorporates in themselves. The environment tolerates, so to speak, large variation of new individual and collective forms, no matter whether they are "fine-tuned" or not to the environment.

To be sure, selection may happen in the environment, but the environment does not perform autopoiesis. As the system can produce an excess of forms for its autopoiesis (not only the strictly determined ones), the environment has excess potentialities and expanded unused capabilities to absorb ever-increasing diversity of autopoietic systems with unknown limitations. As already said, the environment is not governed by autopoietic drive, so it does not "intentionally" make selections.

In conclusion, adaptation is a concept that does not grant the system privileged chances to perform its autopoiesis. A system cannot increase or decrease its autopoiesis or its chances to perform that. A system that stops its autopoiesis also stops its adaptation because it is already dead. In its autopoietic drive, a system is doing more than what the adaptation expects; it is varying its processes of reproduction and creation of the means to reproduce itself.

What tells a system whether the tested and adopted new operations and respective means of doing so should be maintained or not, is not the adaptation but rather the way the system recognizes it as feasible to be repeated and maintained, given, first of all, its internal communicative capabilities to do so. If adaptation to the environment is achieved – meaning, if the system's new ways of successfully reproducing itself do not meet strong negative reactions in the environment – this might be an additional argument for keeping going and inventing new forms.

It is something well established in biological sciences that although adaptation to the environment assures good chances of survival, it does not explain the huge variations of forms created, where surely some of them are in fact costly and represent a sub-optimal adaptation strategy; members of some species would have better chances of survival without some of the characteristics they have acquired, although the specie as a whole carries on faring well.

Having said that, a few additional reflections are necessary. The word "adaptive" has strong denotation associated with Darwinism and natural selection. It expresses the idea of "adaptation" of species (and more recently of psychological and social systems) to their environment. It is hard to use the word without such connotations. Adaptation is therefore understood as the driving process by which a system becomes better able to live in a given environment and by doing so becomes more resilient. The system–environment

relation in such a notion is of determination, whereby the environment has the causal factors, which the system has to recognize and adjust accordingly.

Luhmann's Social Systems Theory departs from such notions in correspondence with his conceptualization of operationally closed systems built by communications. As already said, if there is adaptation, and Luhmann would not dismiss this possibility, he would qualify it as adaptation to an environment that the system represents internally, meaning "adaptation" to the represented environment with the selected features seen as relevant by the system, and not the environment itself in all its complexities.

Furthermore, not all transformation a system goes through is necessarily intended to make it better adapted to its environment. The autopoiesis of the system, by which the system reproduces itself through its own means, may generate internal structures of communication that are understood by the system as relevant in themselves for the systemic operations it keeps running.

For instance, when a health system hit by an epidemic elaborates and puts in place new structures and routines, it is adjusting itself to the internal tensions and pressures it observes in consequence of the increased number of cases of the disease. Is this system becoming better adapted to its environment? It is possible to say yes, because the pressure is coming from the environment and the system is trying to respond to it. However, whatever the system internally does cannot be summarily described as adaptation to the environment. Through its operational units the system is adjusting functionalities as is feasible within the units' capabilities and resources, and their competences to self-observe and select what needs to be changed (increased, or stopped, or reduced, or started anew, etc.). Each unit, as distinct organization systems (hospitals, clinical facilities, etc.), redesigns part of its internal functions according to its self-observation and self-organization capabilities. In that, each unit offer limited responses, taking into account and seeking coherence with all the other operations the unit has to keep running. It has, so to speak, the keys to the control room and the codes that can operate any transformation intended to offer better responses to the new pressures from the environment, as well as to better preserve the existing functionalities of the system, considering its regular routines, protecting existing functions from the new interferences.

The prerogatives and the priorities for keeping the system running and performing its autopoiesis overrule the circumstantial pressures coming from the environment. If not for that, the environment could destroy the system; the system must keep its selective capabilities – the ones it has acquired over the years, even if they are not optimal in face of changes in the environment.

Furthermore, if the current complexities of the system represent successful self-organization the system has adopted throughout its history, that does not mean the system could not be different today, and that the decisions previously taken couldn't have been different. Contingency is the concept that Luhmann often brings into these discussions. The decisions are contingent, meaning they could be different or even the opposite – a certain decision is neither necessary nor could it not be different. In this sense, if there is adaptation, it did not follow any directionality imposed by the environment. The system selected its new operations and the selection process is contingent, i.e. we repeat, it could be different.

This changes significantly our understanding of adaptation – not as deterministic but as a field of possibilities, with some having a lower probability of being selected but nevertheless being chosen and implemented, leading to a future state of the system that is perfectly functional even if circumstantially less cost-effective.[5]

Understanding adaptation to the environment as a major driver of a system misses the point of the high contingency of the decisions a system can make and their subsequent success. "Adaptive" is therefore an adjective that can be easily used in reference to health systems but, first, it is not an object of clear scrutiny and specification, and second, even worse, it obstructs the sight of health systems' huge diversity, internal complexities and the contingencies of their decisions.

5 A large hospital complex may decide to install high-tech new equipment that can hardly bring investment returns, but nevertheless puts the hospital in the forefront of the advancing technologies that eventually will become routine; the hospital will then have gathered the necessary experience.