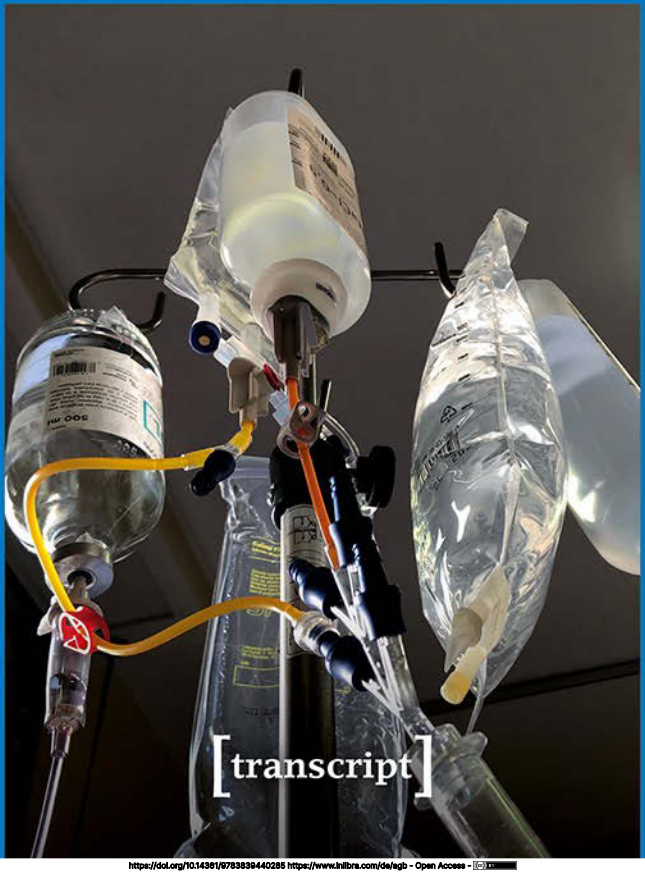
HEALTH, COMMUNICATION AND SOCIETY

JOÃO COSTA

THE CONSTRUCTION OF SOCIAL HEALTH SYSTEMS

HISTORY AND SELF-REFERENCE OF MEDICINE AND PUBLIC HEALTH



João Costa
The Construction of Social Health Systems

João Costa is a public health specialist and health economist. He has a doctorate degree from the London School of Hygiene and Tropical Medicine, where he also worked as research fellow. He is also an independent consultant and associate researcher of the Swiss Tropical and Public Health Institute. He has approximately 45 years of working experience in development projects in several countries, always focusing on strengthening health systems. His international profile includes project management, research, teaching, consultancy, and high-level advisory positions – working with academic institutions, consultancy companies, international development banks, international development agencies, and others.

João Costa

The Construction of Social Health Systems

History and Self-Reference of Medicine and Public Health

[transcript]

With kind support of:

R. Geigy-Stiftung
Die Stiftung des Swiss TPH

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://dnb.dnb.de

transcript Verlag | Hermannstraße 26 | D-33602 Bielefeld | live@transcript-verlag.de

Cover design: Maria Arndt
Cover illustration: João Costa
Printing: Elanders Waiblingen GmbH, Waiblingen
https://doi.org/10.14361/9783839440285
Print-ISBN: 978-3-8376-7954-0 | PDF-ISBN: 978-3-8394-4028-5
ISSN of series: 2940-1828 | eISSN of series: 2940-1836

Printed on permanent acid-free text paper.

To Angela, my mother
(In Memoriam)

Contents

PART II

PART III

PART IV

Appendix

Acknowledgements

First and foremost, I would like to express my gratitude to the R. Geigy Foundation and the Swiss Tropical & Public Health Institute, with special thanks to Director Juerg Utzinger, for the financial support that enabled the publication of this book. Additionally, I am very thankful to Marcel Tanner, from whom I received the first insights a decade ago into the relevance of systems approaches for understanding and addressing the issues related to health systems.

I am also thankful to Marcel Tanner and Lukas Meier for reading previous versions of the manuscripts and giving me precious advice. Marcel on the topic of health systems and Lukas on historical themes. I am also very grateful to Elena Esposito for her comments on Luhmann and the subject of social systems during the first stages of the manuscript.

While their contributions were very relevant for shaping the final format and content of this book, I bear sole responsibility for the ideas and interpretations presented herein.

I can add here a list of persons who, perhaps inadvertently, encouraged, contributed, or inspired me in chats and informal conversations, even when they were not aware that I was taking mental notes for later on. In no particular order they were: Christian Morgner, Maria Thereza, Till Mostowlansky, Felix Roth, Kaspar Wyss, Clara Thierfelder and Marcel Braun.

I also express my appreciation to my editor, Jakob Horstmann for his continuous encouragement since the early stages of the two and a half years process of writing this book, and to Joan Dale Lace for her revision

of the last manuscript, without which this book would not be as reader friendly as I believe it has become.

Preface

Books are read with expectations. The title, the blurb on the cover, a word or two glimpsed, they may trigger associations, presumptions and anticipations. But authors' intentions and readers' expectations are not always aligned.

It is worth talking about this in these first pages. In this foreword we can say what the book is about; its general framework and the audience it was written for. We can talk of passionate motivations to tell a story that is very interesting from the authors' point of view.

We can also suggest how this book may be read and what the reader will encounter. We may also need to say what the reader will not find. This exercise should help to avoid expectations the book might disappoint.

This book is an application of the Social Systems Theory. As the application of a theory, it is both a test of the theory's capacity to shed light on the phenomena it is applied to and a test of the possibility of using it for framing practical matters and finding solutions.

Most of the book is about the first test. We do believe the Social Systems Theory offers the concepts we need for understanding something as diverse and complex as health systems.

For the second test, we dedicated only a short section in the conclusions chapter. We believe that if we understand how health as a social system came into being, how it emerged and became what we see today, this can be a valuable reference for understanding health systems in general and specifically.

This book is about health as a social system; about the emergence of health as a *social system*. A historical perspective therefore frames the

book. As we speak of emergence, we speak of history; we speak of a time in the past when health systems did not exist, and thereafter we talk about the unfolding of events and conditions, leading to the eventual establishment of health systems.

We revisit a time when medicine as we know it did not exist at all. Healing practices were not separated from magical and religious beliefs. Public health did not exist either. There were no institutions or public coordination intended to preserve or improve people's health. An institutionalized knowledge domain concerned with treatment of individuals and prevention of diseases in the population was thus non-existent. Healing practices were everyone's and anyone's businesses. The strength of soldiers and workers, and the habitability of environments, were of differing levels of concern to administrations, with no distinct institutions having a health mandate. Certainly, it was a very different world from the one we now take for granted.

Therefore, health as a social system did not exist. It is indeed difficult to imagine such a context and we are tempted to inadvertently superimpose current structures and functions on to the past when we try to describe a context in antiquity. We may even imagine doctors at the bedsides of their patients, with some of the ubiquitous characteristics we are familiar with, while in fact the differences must be huge. So we need to be careful when we are looking at the past.

We have explained in our book *Health as a Social System* why health is a social system (Costa, 2023). In that book we applied Niklas Luhmann's Social Systems Theory to the practices and knowledge of the field of health, and explained why the health systems we currently know have the characteristics of *social systems*.

In this book we do not return to those themes. We take the previous book as our point of departure. The intent now is to add a view on the advent of health systems to the understanding of health as a social system – that is, to tell the histories of social configurations that made health systems possible. By doing that we should be able to apprehend the historical making of a social system, and gain a better understanding of health as a social system.

With such an intention, we needed to study the history of medicine and the history of public health, as medicine and public health are essential parts of health systems. Nevertheless, we do not revisit the history of medicine or public health; this book does not replace the numerous books telling these histories. We do something specific here, aimed at health systems.

But we do need to talk about medicine and public health, which we do throughout the book. The reader will see, from the beginning to the end, that the book talks about medicine and public health with a broader understanding of these domains.

We consider medicine and public health as sub-systems of the health systems. We work with the distinction between medicine and public health, so it is good for the reader to bear in mind that this distinction is a central tenet of the book.

We understand the focus of medicine is on the individual. The binary distinction "healthy/sick" is the fundamental building block of the communications making the health systems, and particularly the communicative operations of the medicine sub-system.

Likewise, in our broader understanding, public health is the sub-system of the health system orientated towards populations, to the collective as opposed to the individual. The orientation of the communications of the public health sub-system is the basic binary distinction "at risk/not at risk". This distinction is fundamental to and encompasses all public health initiatives, whether initiated at governmental institutions or any other organization, addressing a population's risks of losing or not recovering their health. We will talk a lot about that distinction in this book.

And what about intentions and expectations? We certainly are interested in health as a social system and in using the Social Systems Theory to analyse the development and emergence of that system. We seek to validate the theory and at the same time acquire valuable insights about the history and the actual *modus operandi* of health systems.

Validation of theories can happen first in the theoretical space, where the consistency and coherence of the concepts and their articulation is verified by the basic knowledge of the reality. Secondly, it can happen

with empirical observations and testing, and with measurements as precise and accurate as possible. However, we are mostly dedicating this book to the assessment of theoretical consistency.

But we must clarify that this book is far from being about abstract systems theory. We do not want to discourage potential readers with the suggestion that they will find here an arid and sterile discussion about abstractions. Our historical references are as concrete as can be found in any book on the history of medicine. We look for ways of "vindicating" the theory and the clarity it can bring about.

Still, we fully acknowledge that the audience we are aiming at are those struggling in one way or another to understand the health systems they are in charge of or are entrusted with the responsibility to improve. We are also aiming at those who dedicate their energies to observing, describing and suggesting improvements to health systems across the globe. In the audience we imagine also those who have a special interest in social systems and construct methods and approaches to observe and describe them. All of these readers are in many diverse institutions with largely different aims and objectives. In one way or another, though, they are all interested in health systems.

We think the book needs to hold something valuable for all of them. But we need to be clear that it does not offer guidelines, recipes or ready-made road maps. We do not talk about methods of observing or conducting interventions. The book does not contain case studies and does not use illustrations of specific countries and contexts. But we are sure the reader will be able to see in the text the health systems they work on or have experience with.

Nevertheless, we assembled draft instructions, which the reader will find in a section of the concluding chapter (Chapter 8), with the kind of "take-home" message readers can keep at the back of their minds while observing the health systems they are interested in.

We call this "ways of looking". We suggest to readers how they can focus on the system and see the peculiarities we attempt to highlight with our historical approach. We believe the reader will find reflections in the peculiarities of each country, of the path humanity travelled, even with-

out medicine, public health and health systems as we now know them, looking for ways to learn about health.

If we understand how health systems came into being, how they emerged and became as they are today, we can find valuable references to help us understand the different stages of development. We can identify what were, and still are in some cases, the key missing structures and functions.

To facilitate the reading process, the chapters indicate their essential messages. Some chapters are optional and we also show what the reader will find in them. We added an Appendix called "Advanced Topics", which goes deeper into theoretical discussions and might be of interest for some readers.

The idea is also to make each chapter as far as possible a "standalone" piece, possible to read independently. However, our suggestion is that Chapters 2, 3 and 4 are necessary for a full understanding of what follows. Readers familiar with the application of Luhmann's theory and general information about the history of medicine may feel comfortable reading just the subsequent chapters.

We forewarn readers that some themes are recurrent and reappear throughout the book in several different sections. Concepts such as system's self-reference and complexity, for instance, are deployed frequently in the text. Therefore, although always adding additional nuances, some concepts are explained repeatedly to facilitate the work of the reader, reducing the need to return to previous sections to refresh the understanding of the concepts. We hope this works for the benefit of the reader. These are concepts at the heart of the message this book tries to convey and thus clearly need to be mentioned often. Still, we added a short glossary to offer quick clarification of the meanings of the theory's concepts to make it easier for the reader.

Chapter 1: Introduction

This book is about the emergence of health as a social system. Ancient medicine is where it all started. The history of medicine tells us the story of the development of the nucleus, the embryo around which future health systems were progressively built. This organic image may help us to see medical knowledge developing, with its ramifications, constructing and shaping health as a social function, concerned with illnesses occurring in the body and the respective risks to societies.

The history of medicine has been thoroughly researched and the literature is vast. The themes recur across publications, with documented or presumed facts being well established without much controversy. The materials are widely available in comprehensive textbooks; we therefore did not need to research original historical documents. We did not intend to bring new facts to light or new versions of the facts.

Our interest is in health as a social system, its rise and particularly the construction of its self-reference. We understand that Niklas Luhmann's Social Systems Theory opens new possibilities for studying the development of medicine and health systems. Luhmann proposed a grand theory explaining how social systems developed over the centuries. His theoretical perspectives are presented in this first chapter.

Through the lens of Luhmann's theory we can see nascent structures leading to the systems differentiation seen in contemporary societies, where social systems achieved a ubiquitous presence. Our aim is to bring together two fields of knowledge: the historical narratives of the evolution of medicine and public health, and the theoretical view of social sys-

tems development. In this way, we believe we can trace the evolution of health systems and understand them better.

Today, healthcare services across countries show characteristics of social function systems, as defined by Luhmann (see Luhmann, 2016; Costa, 2023). Functional differentiation – that is, the structuring of societies into function systems – was consolidated from the eighteenth century onwards, with health systems becoming one of the function systems, among others such as politics, education, science, art, law, the economy. Each function system has its own history, maintaining its differentiation from all the others.

The Social Systems Theory also calls our attention to complexities. Systems are complex and the historical development of a system is also the historical development of its complexities. The evolution of medicine unfolded with progressively more sophisticated models of structures and functioning of the body and its diseases. The complex understandings of the body evolved, with subsequent and correspondingly complex treatments. These were processes whereby complexities increased in all domains; therefore, in our attempt to trace and explain the emergence of health as a social system, we pay attention to complexities.

In the process of reaching the current advanced stage, the history of medicine is the history of discoveries of human bodies' structures and functions, and the history of findings about the changes produced by diseases and treatments. In this history of discoveries, we can see the evolution of medical thinking unfolding with the continuous emergence of new paradigms.

Gradually, medicine as a knowledge discipline became able to revisit, revise, discard or change its own paradigms, developing new ones. In this process we see self-reference in action. Self-reference is a distinctive feature of social systems and also a crucial concept for our analysis in this book.

Since its early stages, medicine has managed its own self-reference, by selecting and deciding on the meanings to keep or discard. With the capacity to internally communicate about itself, medicine acquired autonomy, clearly marking its differentiation from other fields of knowledge. This process developed over the course of centuries, preceding the

advent of health as a social system. This is what we would like to examine in this book.

To give a "road-map" of the book, the text is organized as follows:

- The first part has Chapter 2, containing a summary of Luhmann's concepts relevant for the analysis, followed by Chapter 3, with a discussion about medicine and public health self-reference informed by the concepts of the theory.
- The second part also comprises two chapters. The first (Chapter 4) divides the history of medicine into four periods, in line with our theoretical perspective. The subsequent chapter (Chapter 5) offers presentations and discussions of the ideas of Michel Foucault and Georges Canguilhem, two authors who studied the crucial period in the history of medicine and, from our point of view, for understanding the start of health as a social function system.
- The third part also has two chapters; the first (Chapter 6) is called "'Anatomy' of Public Health"; it discusses health indicators as the backbone of public health self-reference, and the central relevance of the concept of risk. The second chapter in this part (Chapter 7) presents and discusses health as a social system in light of the previous chapters.
- The last chapters bring the conclusions of the book (Chapter 8) and presentations of advanced topics (Chapter 9), in terms of the theoretical elements incorporated into the discussions.

As mentioned earlier, the primary orientation of this book is first to narrate the evolution of scientific medicine, a long historical path of about 2,500 years, during which the knowledge about the body and diseases became increasingly precise. These historical processes offer important clues about the establishment of health as a social system as we know it today.

At some point on that long historical path, but more precisely in the last three centuries, a specific field of knowledge became progressively distinct, addressing not diseases in individual bodies but the common causal factors that affected the collective risks of becoming ill. The con-

cerns were thus with populations and causalities that could be attributed to the environment in a broader sense, not only the ecological and physical one, but also the cultural, economic, political and social environment.

With slower development due to its dependence on the progress of medical explanatory models, the then new field of knowledge eventually arrived at what we now know as public health, with its own sets of models and immense literature. Although throughout the evolution of medicine the sub-disciplines called epidemiology and/or social medicine appeared on several occasions in the past, the field remained underdeveloped until health became a political and legal matter to be dealt with by nation states. From then on, we understand, public health evolved as a field of knowledge with a vast array of models and explanatory paradigms with its own self-references.

In short, more specifically, this book discusses the distinctive origins of these two branches of the "same tree" – medicine and public health – both departing from the same binary distinction of health/illness, with one focusing on individual bodies and the other observing populations.

Furthermore, our main intention is to show that the self-reference of medicine was the core which public health, with its own self-reference, complemented, co-evolving together from then on, generating health as a social system. Health systems are therefore sites of the confluence of medicine and public health, comprising all communications of both subsystems.

As we noted in the foreword, we understand the focus of medicine is on the individual. The binary distinction "healthy/sick" is the fundamental building block of the communicative operations of the medicine sub-system. Likewise, in our broader understanding, public health is the sub-system of the health system orientated towards populations, to the collectives as opposed to the individual.

The orientation of the communications characterizing the public health sub-system is the basic binary distinction "at-risk/not-at-risk", the foundation of all public health initiatives, governmental or otherwise, addressing a population's risks of losing or not recovering their health.

Here we feel the necessity to provide the reader with a few explanations about our understanding and use of the concept of risk, which is crucial in our conceptualization of public health as a sub-system. Binary distinctions are at the heart of the function of social systems in Luhmann's Social Systems Theory, and we will talk more about that in Chapter 2.

It is important to keep in mind that we use the term risk in a way similar to its use in epidemiological studies. In very simple terms, "risk" denotes a ratio between cases and a population denominator, in which the cases are observed. For instance, the risk of getting a certain disease is the incidence of that disease – that is, new cases in a given period and population (or segment of the population), sharing the attributable exposure or risk factor(s) plausibly associated with the causation or contribution to the occurrence of the disease.[1]

Thus understood, the concept of risk can be deployed to many relations with similar structures, indicating probabilities of suffering an undesirable outcome among a given population and time. For example, risk can be observed in reference to risks of being exposed to pathogenic factors; risks of getting sick after exposure; risks of not accessing needed healthcare; risks of not getting healthcare of the necessary quality; risks of negative treatment outcomes; risks of being incapacitated; risks of dying; and so on. Public health orientation is therefore always intended to bring about prevention and protection of populations against identified health risks.

Furthermore, the notion of risk (as well as for the healthy/sick distinction) implies a time dimension through which any collective or individual can cross from one to the other side of the distinction. Time is

1 We acknowledge that as expressed in a mathematical ratio format, risk has an "analogical" outlook rather than the "digital" binary health/sick distinction. Between the extremes yes/no of estimation of risks or exposure to risk, for instance, there are continuous degrees of calculable ratios. However, risks may not necessarily be expressed in numbers. Anyway, if the numbers are available, what matters for public health decisions is the comparative proximity or temporal trends towards one or the other extreme of the continuum.

required for anyone at risk to become sick; likewise, time is needed for someone not at risk to cross the line to become at risk, or vice versa. Children who complete the vaccination schemes move from the at-risk to the not-at-risk position for the preventable disease. We will also talk more about the relevance of time in the universe of public health meanings in chapters to come.

Essentially, the focus on risks orientates the communications and respective actions to address the communalities of the direct causes of diseases as well as the contributing factors, and the interventions to tackle them. With such notion of risk, and seeing the collectives subject to it, public health pays attention to the "forest" beyond and surrounding the individual "trees". We believe this metaphor helps us to grasp the notion.

Having said that, we also understand that while trying to reduce a population's defined health risks, public health employs concepts of a range of knowledge domains not strictly speaking related to medical practices. Among these domains we can mention: management, economics, sociology, systems science, anthropology, law, political sciences, psychology.

Similarly, as the meanings of the universe of medical communications increased over the centuries, the elements and relations now composing the universe of public health meanings have also increased enormously. With the incorporation of those scientific domains into public health communications, their respective complexities were also imported with them.[2]

Nevertheless, despite the dimensions of these huge universes of meanings (medicine and public health), the tasks to locate meanings in one or the other side of the distinction remain relatively easy. We therefore believe the use of a distinction between medicine and public health is appropriate. Even if controversies may still appear, this distinction should facilitate the apprehension of the message of this book about the historical emergence of health as a social system.

---

2 The reader will find an in-depth discussion of the theme of *risk* in the concluding remarks of Chapter 6 and in the last sections of Chapter 9 on advanced theoretical topics, at the end of the book.

To summarize, the book addresses the following thesis:

- With the development and eventual consolidation over the course of approximately 2,500 years, medicine developed the functionality of self-reference fundamental for the genesis of health as a self-referential social system. By consolidating its self-referential *closure* (by which only medicine could judge what belonged to it or not), medicine made possible the formation of health systems.
- Health systems, as differentiated social function systems, have a history of around 300 years. For their genesis, the health systems needed the fully grown structural systemic features medicine had already developed. This includes: medicine's singular exclusive codes and semantics; the closure of its communications; its self-assessment and self-reproduction; and the distinctive way of approaching and observing its object (the human body).
- The health systems hence developed around the core structures of medicine, subsequently expanding and encompassing the health of populations, establishing public health, comprehending not only observation and communications about the sick but also health risk factors affecting populations comprised of people not yet sick but potentially at risk of becoming sick.
- With these two central concerns, the health systems came to include complementary knowledge also linked and orientated by the binary *healthy/sick* codification. With these developments health systems expanded, incorporating practices that were not strictly medical, but befitted the system as supporting, supported or relating to medical treatments, therefore likewise orientated by the same general *healthy/sick* code. This includes nursing, dentistry, clinical laboratories, pharmacology, nutrition, physiotherapy, and psychotherapy and logo therapy.
- Specifically, public health as a key sub-system of health systems, elaborates health systems' own *self-reference as a social system*. The very idea of health as a social system is within the scope of public health concerns and attributions. More explicitly, while medicine reflects on itself vis-à-vis individual patients diagnosed and treated, public

health reflects on its meanings, concerning recognition and decrease of health risks of populations by operations carried out by the health system. The health systems is a comprehensive whole, with dual complementary self-references: medicine and public health. Public health already had its tentative incipient beginnings when ancient dominant classes and governments paid attention to epidemics and particularly the health of specific groups such as soldiers and workers.

- By paying attention to social horizons, in recent centuries, public health meanings became instrumental for health communications with the political and legal systems, while medicine became too complex for communicating with other systems beyond its limits. In turn, though, public health also became progressively more complex with the influx of concepts, paradigms and research methods from other disciplines such as political science, statistics, economics, sociology, anthropology, psychology, communication science, management science and systems science, thus being able to reach broader audiences, increasing its own complexity in the process.

We can say in short that we believe that by studying the history of medicine we can see the unfolding of these developments. The questions we aim at answering throughout the book are thus:

1) How does the history of medicine show the development of self-reference? For that we discuss:

- The development of singular exclusive codes and semantics;
- The closure of medical communications;
- The establishment of medicine's command of its self-assessment and self-reproduction;
- The consolidation of medicine-specific ways to approach, observe and communicate about its subject (the human body)

2) How did medicine self-reference become the foundation of the self-reference of health systems?

3) How did the coupling of the self-references of medicine and public health create the health systems?

In Chapter 7 we present a summary of the answers to these three questions. Furthermore, relevant to our reflections, as already mentioned, is the understanding that public health also has the function of constructing for the health system the identity of the health system as a system, a function social system of which medicine is part. In this sense, this book has been written within the universe of meanings of public health, bringing inputs from Social Systems Theory in an attempt to construct descriptions of health systems for those health systems, more precisely to bring conceptualizations of the descriptions of health systems for those who study and/or work in health systems.

PART I

Chapter 2: Social Systems Theory

In the first section of this chapter we give a brief overview of the key concepts of Luhmann's Social Systems Theory, particularly those to be used in the subsequent discussions.[1] The second section of the chapter talks about elements to consider when carrying out the task of observing social systems, particularly the health system with its two sub-systems, medicine and public health.

Concepts

Pervasive in contemporary societies, health systems are straightforwardly recognizable; those who work in them as well as those who look for healthcare assistance have clear expectations about them.

Those studying health systems from an academic perspective observe them as external observers. Nevertheless, observers, health profes-

1 A number of books have comprehensive explanations of the concepts mentioned in this chapter. It is advisable therefore that the interested reader consults the attached references for more in-depth presentations of the concepts. The book we published in 2023, *Health as a Social System* (Costa, 2023), has discussions of the concepts and their application to health systems. However we recommend Luhmann (1990, 1995, 2007, and particularly 2013) for extensive detailed conceptual presentations. Furthermore, glossaries of Luhmann's concepts can be found in Baraldi, Corsi and Esposito (2021), and Seidl and Becker (2006).

sionals and patients depart from similar understandings of the systems as a comprehensive singular all-encompassing distinctive whole.

Independent from the types of institutions it is comprised of, and types of expertise available, funding arrangements, structures, distribution of services, organizations, and so on, an observer of systems sees health systems as distinct from the others systems in a society.

The distinction arises primarily from the codes and semantics used by the system. Health systems communicate internally about health, and more specifically about diseases and treatments. The identity of the system does not leave much ground for doubt about its main general social purposes, aims and justifications.

According to Luhmann (2013, p. 44), "system is the difference between system and environment". A system exists vis-à-vis the environment where it strives to keep its distinctiveness, unit and identity.

Everything that is not part of the system belongs to the environment. The system exists because it is distinct from its environment. The distinction system/environment[2] is a foundational concept according to the theory. If this distinction is lost, the system disappears.

For being different from its environment, a system has limits marking the separation between what is internal to the system and what is external to it.

However, a system can observe itself as well as its environment. Particularly, out of the incommensurably vast diversities of elements and relations in the environment, the system observes what is relevant for it and ignores the rest.

In the case of a health system, the attention is primarily directed to human bodies. The human bodies are part of the environment. More specifically, the bodies' internal structures and functions, and what happen to them when diseases affect them, are crucial elements of the environment of the health system.

To deal with health and disease in its environment, a health system deploys the semantics and meanings that constitute the characteristic

2 Throughout the book we denote distinctions by separating the two opposite sides with "/" as in "system/environment" mentioned above.

communications belonging to a health system; internal communications occur many thousands of times on a daily basis in any health system in the world.

According to Luhmann (2013), a social system is made of communications and nothing else. Without communication, there is no social system. In written or oral forms, communication is comprised of utterances and information, generating understanding (including misunderstanding) between those engaged in it.

Without understanding, there is no communication. Understanding can be verified by the self-referential functionality of communication, through which those communicating can recursively return to the information and utterances already made and confirm (or not) the agreement on commonly employed meanings.

Recognition and validation of meanings therefore takes place through communications. Only meaningful communications are part of the system; the rest is discarded as not being matters of concern or as just noise. The system needs to continuously operate with validated communication otherwise it loses its self-reference.

For that matter, social systems maintain operational closure, by which their meanings are only communicated and validated internally. Only internally can the codes and semantics employed by the system be corroborated and understood. Health professionals, and only them, have the prerogative and responsibility to make meaningful statements recognizable and acceptable by other health professionals operating in the system.

Operational closure is a fundamental step in the constitution of a social system. A social system can only exist after having established its closure – that is, after creating limits for exclusive meaningful communications distinct from communications of other systems in its environment.

The history of a social system indicates the processes by which the closure took place and the system started to rely exclusively on the semantics of its internal universe of communications.

Without operational closure, a system loses distinctiveness and its communications can be appropriated by other social systems. In such

cases, there is no difference between system and environment. Because of that, each social function system in the society preserves and communicatively reproduces its unique distinctive codes and semantics.

This process is called *social differentiation* (Luhmann, 2007). By that, each social function system (the political, legal, economic, education, religious, art, media, health, and so on) differentiates itself (its communications) from the others, erecting clear boundaries that cannot be crossed by the other systems.

A function system cannot communicate inside another function system, otherwise one would be destroyed; so, by the same token, a function system cannot be subordinated to another and no function system occupies the centre or the hierarchical top of a society structured according to function systems differentiation.

Function systems can only communicate internally using exclusive semantics. However, function systems can enter into *structural coupling* (more on this later) by observing each other. Organizations, though, are social systems that can communicate with each other, because they have inside them sections that belong to the same function system, and those sections can understand each other's semantics (Knudsen, 2012). (See more on this topic later in this section.)

The functional differentiation of contemporary societies started around the eighteenth century. Then function systems became the prevalent form of social structuring, as they still are in current societies. Functional differentiation allows any member of the society to participate in any system; this is another important feature of functional differentiation.

The systems are open to the society in the sense that anyone can become, for instance, a health professional or a patient at some point in their lives, independent of characteristics such as place of birth or family background. In the same way, anyone can on one side be appealing to a legal system or, on the other, be a legal professional dealing with the case. In this way, legal systems are open to all members of the society. Likewise, anyone can be part of the political system, either as governing or governed, or the economic system as a buyer or seller.

Before societies became structured according to function system differentiation, the primary differentiation was according to segments. In segmentary societies, the social location and identity of any member was determined first by the place of birth/residence and second by the family/tribe the individual belonged to. Any individual would be socially recognized (as well as self-identified) as being from a certain place and family. Nothing else would be necessary for creating a social identity.

Then, segmentary societies gradually became internally differentiated according to strata. In stratified societies, any person could be either a member of the dominant stratum (aristocrats and monarchs), or a member of the lower stratum, as typically represented in the Middle Ages by the feudal order of noble landowners and peasants. In that case, a person would be assigned to one or the other stratum by birth.

With progressive differentiation of members of the lower stratum into intermediate categories of craftsmen, artisans, skilled doers, soldiers, public servants, and so on, possessing valuable skills and the ability to accumulate property, new structures for society membership emerged. These new structures progressively developed into the function systems of contemporary societies.

But this progress was not the result of accumulation of stratified structures; the notion of accumulation does not explain the emergence of functions. A function such as political, legal, economic or religious is distinct from all the others; no function is subordinated to any other or has its creation decided by other functions.[3]

---

3 Our description of social differentiation in this book is brief and simple, it is far from giving the full breath of Luhmann's conceptual architecture on this topic, which is explained in details in a 144-page chapter of his book *La Sociedad de la Sociedad* (2007), where he also talks about differentiation between centre and periphery and the overlapping of differentiations, as we can see today forms of structural differentiation coexisting in the same society, with functional differentiation being the primary one. We understand that those macro-social differentiations were not the driving force for the constitution of health as a social system. We rather see differentiations as contextual macro-structural factors favorable and supportive for the development of the health system. In this book, we try to explain that the social differentiation of health as a social

The function system is in charge of itself and its own reproduction – that is, reproduction of the meanings that distinguish it from all the other meanings. A system thus differentiated, as a social function system, is solely responsible for keeping its function. We will return to this point later in this section.

Social structures and meanings offer the elements for conducting mutual observations of/by society members. Observations deploy distinctions by which society members can observe each other and each self. Social identities are thus created and become part of individuals' self-references.

To exist, a social function system relies on observations it makes in its environment as well as observations of its internal operations. To carry out observations, a system deploys distinctions that can also be communicated. Communications themes are both about the observations made and, implicitly or explicitly, the distinctions employed for the purpose of observing. In short, observations and the correspondent communications are the life of the social systems.

To remain alive, a social system needs to preserve its internal communication possibilities, reproducing them through the validated recognizable semantics. The reproduction of a social system can only be self-reproduction, as the social system has exclusive use of its distinctive semantics. No other system can do that for it.

In technical terms, the self-reproduction of a social system is called autopoiesis. This means the reproduction of the system by the means the system itself produces. Only the system can produce the means (meanings) for its own reproduction.

Furthermore, Luhmann identifies three types of communication-based social systems: *interactions*, *organizations* and *function systems*. Most of what has been said before are characteristic of function systems; however, it should be emphasized that *function systems*' communications develop from the base of a binary code that is uniquely used by the respective system. For instance, the legal system fundamentally operates

---

system was mainly built from within the health system itself. This is enough on this topic for the moment; we talk a lot about this theme throughout the book.

with the code legal/not legal, as the health system operates with the code healthy/sick. One side of the code leads to further communications within the system; it is the connective side, linking one communication to the next, while the other remains as the reference, reflection side. Within a health system, in the medical sub-system, there is not much to communicate about a healthy individual; the communication in these cases stops. However, within the public health sub-system, healthy populations can be further distinguished with the deployment of the at-risk/not-at-risk distinction, being the at-risk connecting side of the binary code. Public health keeps central to its concerns and communication the identification of populations at risk and consequently the preventive measures to be deployed to reduce the risks. The not-at-risk side remains as the reference/reflection side. We will return to the topic of binary codes in other sections; for the moment, we revisit the types of social systems.

Interactions are systems established by communication between two or more people where presence is required, in a face-to-face encounter or using electronic media. Once the communication ends, the system also ends. Those participating in the meeting may or may not, precisely or not, recall it and keep memories of it, which can be brought out in future meetings; but this is not a necessary condition for the existence of such system. The participants may never meet again.

Organizations are systems characterized by two things: membership and decisions. Typical of an organization are the selections of those who are members in contrast with all the others. Additionally, organizations perform a type of communication characterized as decision. Only members can make decisions in an organization. Without the combination of these two elements, decisions and members to make decisions, there is no organization.

Having gone through these basic concepts of the theory, the final two to be added to this very brief description are the concepts of *self-reference* and *complexity*. These themes are central in our discussion in this book and will appear frequently in most chapters.

Self-reference is constructed through processes of making distinctions and selections, and their deployment in observations and commu-

nications. Luhmann (1990, p. 123) says: "Self-referential systems are able to observe themselves. By using fundamental distinction schema to delineate their self-identities, they can direct their own operations towards their self-identities". A function system is both a first- and second-order observer. It is a first-order observer by observing the environment and the system's internal dispositions. It is a second-order observer as it performs observations of observers, including observing itself as observer as well as observing observers observing it.

Self-reference thus adds the complexity a system needs to survive among systems that also perform self-observations. Self-reference is established on one side of a distinction where the other side is hetero-reference – that is, what *is not* the system. With that, a system can observe that other systems also have their own self-references.

The capacity of observing other systems creates the possibility of what the theory calls *structural coupling*, by which different systems can achieve a certain degree of coordination without communicating with each other, just observing and acting in correspondence to what they independently observe.

A system capable of self-reference produces self-descriptions, creating therefore a structural predisposition for observing and communicating about itself and its environment. Self-descriptions guide the selections the system makes, organizing them in a more systematic way. The descriptions can be handed down and more easily remembered and connected to each other. We can say that preservation of self-descriptions is the same as maintenance of identity. A system surely dedicates efforts to keep the descriptions it makes of itself; however, the system may also change them, when the need arises. Self-descriptions are contingent i.e. they are neither necessary nor impossible and can be different (this is Luhmann's definition of *contingency*).

Studying systems' self-references, one may observe that self-descriptions may change over the course of the life of the system. The contingent character of self-description thus becomes evident. But a system cannot afford to dismiss all self-descriptions; on the contrary, while maintaining self-descriptions, the system corroborates that those are the only valid ones and are permanent; the contingency is noted by

observers, while the system on the other hand strives to confirm and preserve its assumed definite "nature". This way contingency retreats from sight and becomes latent.

When discrepancies between the self-descriptions and the actual observations the system make of itself and its environment become problematic (when self-descriptions for instance become obsolete), the system needs to make an effort to strike a new balance between the descriptions it still needs to preserve and the ones it needs to change or discard. The history of the self-reference of a system is therefore the narrative of those instances of changing and/or preserving self-descriptions. This leads us to the theme of complexity.

The complexity of systems capable of self-reference is prone to increase, as they are not, as called in cybernetic jargon, "trivial machines".[4] Trivial machines deliver the same results when they face the same inputs. They are predictable. Non-trivial machines, on the other hand, by being able to observe their own processes of producing results from observed inputs, become highly complex. Correspondingly, they become unpredictable, with open possibilities of generating unforeseen results. However, as an evolutionary advantage, with the increase in complexity, a self-referenced system can progressively address additional complexities observed in its environment.

To clarify the concept of complexity in Luhmann's theory, in a simple way we refer to complexity when the number of elements and relations between the elements surpasses the capacity of the system to recognize and deal with them. The system may realize that there are elements and relations that are not well known, as there are also known unknowns and unknown unknowns. The system is thus aware of the information deficit it lives with and the respective gaps in information. We will use these references to characterize complexity in the next chapters, emphasizing the configuration of complexity in terms of elements and relations among elements, particularly causality relations.

4 In his books, Luhmann often refers to Heinz von Foerster in relation to cybernetic concepts and particularly "trivial machines". Technical explanations of "trivial machines" can be found in Foerster (2014).

We can conclude the section by explaining that complexity is a relative attribution of what is observed in correspondence to the system's observation capabilities. Something is temporarily complex because the observing system does not command the necessary observational distinctions and capacities. Complexity is acknowledged where there are insufficiencies in competences to observe. Complexity is not an object in itself; it is a consequence of the temporary limitations of the observer.

Given the incommensurable complexity of the world, any system carrying out observations of its environment needs to reduce complexities by focusing on the elements and relations, and the respective meanings, the system can select to work with. The system needs to represent the environment inside itself. The systems create simplified models of the environment to be able to meaningfully approach it. Meanings are thus embedded in complexities and complexity reduction attempts; a meaning construction requires selecting out all other possibilities apart from the elected one. To be meaningful, something needs to be separated and distinguished from the complexity it is immersed in – a "sea of noises". We will return to the topics of complexity and meaning at some points in the book where the explanations can help to clarify the text being discussed.

Observing systems

The history of medicine is a saga of overcoming limitations. To confront the human body and its diseases, the medical observers faced complexities of unknown magnitudes. Indeed, the pull to know better was continuously present alongside the recognition of the limits of what was already known. The unknown was immersed in the insurmountable complexities of the environment of the system.

A system of meanings cannot have direct access to the complexities of the environment; therefore it needs to represent the environment inside the system. For that, the system can observe the environment and internally develop models to make sense of what it has observed. Models reduce the complexities of the environment to manageable sizes. The

system then deals with the representations itself made with the meanings it constructed.

To elaborate those models, the system makes distinctions and selections of the features it considers relevant. Relevance is judged by the system in relation to its own self-references and self-reproduction, and the supposedly acceptable justifications thus put together.

Nevertheless, the representations the system creates of the environment also have their own level of complexities. The system therefore needs to keep the complexities of its observations as well as the complexities of its models to convenient sizes, discarding what the system cannot properly acknowledge or relate inside the constructed model.

In the process of dealing with complexities a system may not have what Luhmann (1995) calls the "requisite variety". He says: "Systems lack the 'requisite variety' that would enable them to react to every state of the environment" (p. 25), and adds that there can be "no point-for-point correspondence between system and environment". He also says:

> "No society can bring about the 'requisite variety' or corresponding degree of complexity of the environment. However complex its linguistic possibilities and however subtle the structure of its themes, society can never make possible communication about everything that occur in its environment" (p. 182).

Let's take an example. From ancient times up to the Middle Ages, to deal with diseases, practising doctors had available to them what from today's perspectives were very simple models, describing the human body as consisting of basic functions and parts affected by the balance between the four humours (blood, yellow bile, black bile and phlegm). Essentially, the four humours was the only explanatory model available; it represented a huge reduction of the complexities of signs and symptoms seen during disease progression.

As any model, the four humours performed the reduction of complexities expected from it by: limiting the *diversity* of elements and relations to be observed in the human body; presenting possible *causality relations* between observable potential causes linked to effects (such as

imbalance in humours due to excess of exercise, food, sex, weather conditions, and so on, predicting the progress of the disease); indicating actions to achieve the expected desirable *effects*. This model could help to make prognoses and create expectations.

The adoption of models were not necessarily accompanied by the expected results, but fundamentally the models made available a structure of meanings orientating the communications among those who were participating in the activities of identifying and treating diseases. Such a base for communication was the initial step towards the closure of the semantic field, to be eventually achieved when meanings started to be used only by initiated practitioners sharing common understandings. The four humours framework was common knowledge and anyone in ancient societies could play and speculate with it.

Fundamentally, the systematization of observations, and collective adoption of the same distinctions to make observations was a major achievement of Greek medicine. The documents written and circulated had a powerful influence in creating the preliminary conditions for the unfolding of what many centuries later would become health systems.

A disease such as tuberculosis, for example, has been around since time immemorial. The descriptions of the symptoms made by Hippocrates and his disciples allow twenty-first-century doctors to recognize them as observations made on patients who almost surely had TB (see Fox, 2022). The distinctions used to carry out the observations are still considered valid. Obviously the arsenal of resources now available to confirm TB in any patient is incomparably more complex. The doctors can see now far more intricate symptoms and causal links, and deploy a range of interventions no one could even have dreamt about in ancient times.

Someone may say that "TB has changed in comparison with what it was then", but we can instead say that the complexities of what is represented inside the system reflects the complexities the system can now observe in the environment. The complexity of the system has increased and therefore also the complexity of the representation the system makes of that particular object in the environment, TB. The

object TB may still be the same; what has changed is the way the system now observes and communicates about it.

What is remarkable, and therefore won for Hippocrates the title "father of medicine", is the setting in motion of the self-maintained process of writing and reading about diseases, and finding in actual patients correspondences with the texts, and adjusting, complementing and enriching the texts for continuous subsequent observations; in that we see the embryonic system dealing with and developing the complexities of its systematic observations.

They would carry out observations on the body of the patient, and with the semantics then available would communicate their understanding to the patient, to patient's families, to assistants and other doctors involved in the process. All these elaborated meanings were part of the beginnings of the complexity of the system starting to take form. Treatment actions could also be observed and translated into communications using meaningful semantics, and the new meanings were thus incorporated into the repertoire of communications and memories of the evolving system.

A new case could add new complexities to the system's representation of the body, if the patient did not react according to the predictions of the causality models. But if there were no substitute meanings, the new complexities appearing in the environment and acknowledged within the system could not destroy the meanings already in use. So the system had to perform operations to somehow turn what appeared to be meaningless into something meaningful; the system, as a communication-based system, could not afford meaninglessness.

Cases could be dismissed, models could be revised, new semantic terms could be created, and new observations could be performed, or any other strategy for assuring the survivability of the system and reproduction of its meanings, preventing disruptions. As far as possible, internal complexity as well as the internal representation of external complexities, had to be kept at manageable and consistent levels.

With these notions in mind, it is possible to trace the evolution of the health system through the stages of social differentiation, analysing the

configuration of systems' internal and external environmental complexities as represented by the system inside itself.

An observer of a system should therefore look at: the *semantics*, identifying the sets of meanings deployed in the system's internal communications; the distinctions and procedures the system deploys in its *observation* of its internal as well as external environments; and the range of *communications*, in any form, that the system maintains.

Semantics indicates the complexities of the concepts used. Semantics orientate observations and are modified by them; observations and semantics have to appear in communications, otherwise they do not exist for the system. So, these three dimensions – *semantics*, *observations* and *communications* – are interlinked and maintain coherence and stability within the sets of meanings the system operates. It is important to remember though that an observer of a system will not be able to reproduce all the complexities of the system they are observing; a social system is always more complex than an external observer can figure out. Nevertheless, the observer can see and analyse the relevant features of the topics of their interests.

In this chapter we have advanced a bit on the topics of Hippocrates and the orientation towards observations in Greek medicine. These points were mentioned here for illustrative purpose. In Chapter 4 on the history of medicine, from its beginnings in Greece in the fifth century BC, topics of self-reference, semantic closure, complexity and systems' self-reproduction will be thoroughly discussed.

As a light-hearted warning, we emphasize that this brief introduction to some of Luhmann's concepts by no means pretends to be comprehensive and complete. Luhmann's conceptualization is rich in details and requires dedicated study to grasp the full extent of his insights. Here we only offered a few hints using simplified versions of the concepts. The reader, we hope, will feel motivated to dig deeper into the Luhmanian world.

Chapter 3: Self-Reference

The self-reference of a health system is built using meanings and communications about the human body, about diseases and treatments, and about their distribution in the population. As we explained previously, self-reference is made using the meanings a system generates and reproduces in relation to its internal and external environments, and distinctively communicates internally. Self-reference guides the selections of themes, observations and communications according to their pertinence to the system.[1]

Despite the fact that over the centuries many different representations of the human body and disease were proposed and became part of the repertoire of medical meanings available for communications, the human body as a potential "repository" of all signs, symptoms, structures and functions remained a permanent signifier, a symbolic medium supporting the meanings (what is signified) with which the systems communicated.

In the first section of this chapter we talk about consistency. We talk about how throughout history a certain way of seeing and communicating has been preserved continuously. We talk about the preservation of the symbolic representation of the human body and health, and the binary distinction healthy/sick. If we have a history of medicine it is because in one way or another these symbolic representations have been consistently used and universally preserved across the world over time, maintaining the base of what medicine is.

1 The discussions in this chapter are mostly based on Luhmann (1990 and 2022).

Symbolic medium

We can see the human body as a medium whereby, throughout the history of medicine, medical forms (anatomy, physiology, pathology, biochemistry, and so on) progressively took shape. As a medium, the body allows the forms of disease to be distinguished, observed, constructed and communicated about.

It is important to bear in mind that we are talking about a *symbolic medium*. The recognition of something that is not going well in a body is represented in a symbolic realm, where the body is represented as well as the disturbance. Only representations at the level of this symbolic realm can be communicated. Without symbolic representation, there cannot be communication about it. Doctors meet at this symbolic realm where they can communicate and refer back to a concrete body where they observe signs and symptoms. Signs and symptoms are symbols, so to speak; symbolic representations referring to reality and addressed as real. Without symbolic representation there cannot be medicine.

The symbolic nature of medicine is confirmed in its history, where we can see several interpretations of reality succeeding and replacing one another in the search for better "tuned" but still just representations of the real. According to Luhmann, a system cannot directly apprehend its environment; a system can observe the environment, select the elements it wants and can observe, and thus internally construct the meanings the system deals with. The system creates internally a representation of its environment and can then use the meanings to inform and mediate its actions on the environment.

Having said that, now we come back to diseases as forms in the medium of the body. In other words, a communication about a disease portrays it as a form in that medium. As a medium, the body is the full potential, and any specific disease is one of many possible forms that can appear in it (as a piece of wood as a medium may contain an almost infinite number of possible sculptures).[2] If we accept the idea of the

2 However, as opposed to a concrete medium such as a piece of wood or marble, which are exhausted once a sculpture is carved, a symbolic medium is not

body as a medium, the history of medicine is the history of the discovery or construction of forms, i.e. structures, functions and pathologies in that medium.

At one time, diseases took specific anatomical and physiological forms. As time went on, new forms were uncovered in cells and tissues. Subsequently, they appeared in microstructures in the cells, chemical molecules, electric impulses and pathological arrangements of multiple expressions. And the progress continues. Forms have been systematically described since the "birth" of medicine; the historical development, we can say, has consisted in identifying and studying form-constitutive elements (building blocks) and their arrangements.

An ever-increasing repertoire of forms is now available to all practising doctors in their quest to find diagnoses and treatments. These forms are above all communicable forms – forms that can be explained and talked about. Doctors must be able to adequately deploy the meanings of those forms in communications among themselves.

The self-reference of the health system is therefore based on the symbolic medium of the body. Medicine focuses its attention on it and there is no pertinent question that medicine needs to seek elsewhere. However, another symbolic medium of specific nature is also fundamental. Luhmann uses the terms *symbolically generalized medium of communication*, which is a concept Luhmann borrowed from Talcott Parsons. The term denotes the notion of a medium similar to how we talked about the body in the previous paragraphs.

However, differently, the symbolically generalized medium of communication implies common understanding and shared meaning among those communicating with it, linking both selections of what is communicated and motivations to accept the selections. In less abstract terms, they are symbols that orient the choices and make the acceptance of what is communicated more probable. Messages about health, for instance, are more likely to draw attention and potential acceptance if

---

consumed in the process of creation of forms; new forms can continuously be made intelligible in such a medium.

they come from health professionals rather than professionals of other fields.

Luhmann (2007) presents several examples of symbolically generalized medium of communication, among them particularly power, law, money, love, art and religion. To these we can add health. The motivational aspect of the symbolically generalized medium of communication can be illustrated with the example of communications using the symbolic medium of power. A government decree or an act approved by a Parliament carries the symbolic determination of a deed of power and accordingly motivates compliance and acceptance.

Outside the symbolic domain of power, the communication of a document with the same text as a decree would not elicit any motivation to comply. It is important to also notice that as medium, power can support countless forms (as in the more concrete example of sculptures in the medium of wood), including all sorts of decisions, orders, commands, instructions, regulations, and so on coming from recognized power holders.

Likewise, as for all symbolically generalized mediums of communication, we can see the same. Money elicits motivation for acceptance on both sides of the transaction, the seller and the buyer. Law motivates obeisance among those collectively bound by it. Communications in the medium of art can result in a certain object being accepted or rejected as art. In the medium of love, some messages may be taken as corresponding (or otherwise) with the standard of expected reciprocity. In the symbolic medium of a particular religion, certain communications can motivate recognition of validity. All these mediums operate in symbolic domains. They simplify communication and reduce the complexities that would otherwise expand with lengthy and probably unconvincing specifications, therefore preventing acceptance.

We may say that health entertains the same symbolic type of functionality, presupposing acceptability in certain communicative contexts. The description of a disease form by a doctor surely refers to the symbolic medium of the body but is built in and communicated through the also symbolic medium of health. The description is likely to be accepted and

mobilize further interest in advancing the communication regarding the corresponding treatments.[3]

Medicine is essentially concerned with the human body as a medium, as mentioned earlier. Health, on the other hand, as a generic symbolic medium, is attributable to any living organism. However, inside the health system, there is high willingness and motivation to keep communications about the body and human health going; they are absolutely vital for the existence and reproduction of the health systems. This sounds like a rather obvious thing to say. But let's consider a few more thoughts.

Is there a difference between these two symbolic mediums of the body and health? Are they referring to different things? The answer is yes, there is a difference in the kind of communications they support. The body refers to a concrete, tangible object, conveying the sense of specificity of determined structural and functional configuration. Health, on the other hand, refers to several aspects and types of phenomena. Health may for example refer to conditions of the body; the institutions providing healthcare; observational strategies (health surveys); contextual sets of factors and risks (occupational health, environmental health, and so on); policy orientations. Health as a medium entertains the structures and communication functionalities of social function systems, and the correspondent comprehensiveness of all forms constructed within it, while the body is a medium for the identification of (and communication about) meaningful biological forms.

To sum up, when we study the history of medicine we see that the meanings communicated among health professionals have radically changed many times over the centuries. The second part of this book will take us through those major changes. But we have to admit that the symbolic representation of the body for which the fundamental healthy/sick distinction could be employed has been retained throughout history. Not the body of aesthetic ideals, or sports achievements, or countable in the formations of infantry phalanx, or a legal entity endowed with

3 Additional presentations of the *symbolically generalized medium of communication* topic can be found in Costa (2023) and more extensively in Luhmann (2007).

rights, or a religious sanctuary of divine nature, or some other representation, but the body as the medium where diseases take shape. The very basic binary health/sickness code, from where all subsequent medical codifications sprouted, has been carved onto the preserved millennia-old symbolic medium of the body.

The consistency of the preserved symbolic medium and the instrumental binary code for "sculpturing" it has allowed the systemic progress of medical knowledge and the eventual advances in health systems (including public health). This consistency made possible the global reach of medicine, expanding from its start, projecting a unifying and coherent universal understanding.

Wherever medicine was and is practised, it could and can be recognized and validated as such. The use of the universal binary code and its reference to the symbolic medium of the body are the base of the recognition, which includes self-recognition and self-validation of medicine. We can say the self-reference of medicine endows it with universal values; at its core, it is a universal self-reference.

However, that consistency has also been a dynamic process, sustained by the fundamental support of the unique symbolic structure. A hugely complex building has been constructed on a rather simple but extremely stable symbolic base. The building expanded in many unforeseen directions. The history of medicine as it is usually written and taught is the history of discoveries and breakthroughs. It is a history of a discipline progressing within the scientific system. We can say it is the history of a science discipline evolving as it increased its repertoire of information on its subject, developing new ways of approaching it.[4]

4 The last section of Chapter 9, "Advanced topics", presents some additional explanations on the symbolic medium of the body and the symbolically generalized medium of communication.

Medicine self-reference as science and as practice

We now direct our attention to the more operational and pragmatic side of medicine progress. Surely there have been time lags between what medical science has produced and what became incorporated into medical practices. Therefore we need to talk about a distinction between medicine as a distinctive discipline in the scientific domains and medicine as a core component of health systems, providing actual healthcare (diagnosis and treatment).

While medical scientists throughout history have been concerned with describing what they uncovered in their research, medical practitioners as providers of healthcare have always been concerned with what could be done for the patient they had in front of them. Treatment of actual patients hardly offers the optimal conditions for scientific study; treatment decisions are to be made with what is available, independent of the best knowledge on the matter. Uncovering new scientific knowledge is not a priority in such a context. The settings of these two exercises, medical science and medical practice, are distinct, although they may overlap, for example when treatments are delivered in clinical trials, or when records made by doctors in their daily work are used for statistical analysis in exploratory and epidemiological studies.

Nevertheless, during its history, as a knowledge discipline and as a field of practice, medicine exercised its self-references. Here we try to explain self-reference in simple terms and examples. As a scientific discipline, medicine progressively acquired and exerted its competences to revisit, revise, reconstruct, discard or change its settled notions, developing new ones, in a rational process of knowledge selection.

In comparison, in the practice of treating patients, medicine has always been constrained by what is feasible in the circumstances where healthcare has to be delivered. Even where doctors have scientific inclination and comprehensively search the literature for the issue they need to decide on, the limitations of the situation and the time constraints are determinants of what can be done.

Any doctor, at any point in history, would be able to say whether they were exploring scientific horizons or doing what they could to treat a

given patient with the knowledge and conditions available. The distinction would be and still is clear for any professional. The communicative operations of medicine in the social function systems of science and in the social function system of health are also distinct, reflecting the aims they pursue and the understandings they strive to achieve. The science system primarily is orientated at deploying the binary code true/false as scientific knowledge (Luhmann, 2013), while the health system, as mentioned earlier, is primarily concerned with the healthy/sick binary code (Luhmann, 2016).

In line with these observations, we can say that since its early stages, be it in scientific explorations or in daily practices of treating patients, medicine has managed its own self-references, selecting and making decision on what to keep or discard. This is a common feature of self-reference construction of any discipline, whether as applied knowledge or as a scientific exploratory domain. With the capacity to internally communicate about itself, medicine acquired autonomy, strengthened its self-references, making complete its differentiation from other fields of knowledge (we will return to this in many following sections).

Public health

If we now look at public health, we can say that the same distinction between scientific and service provision can be made. In the history of medicine, public health appeared as a concern with conditions that affected populations. If at the beginning it had specific population groups in mind, such as soldiers for the benefit of keeping armies fit and ready to fight, or workers in mines, construction sites, plantations or other activities involving large human concentrations, the concern with collectives was also acute, particularly when epidemics were frequent and constituted recurrent threats to countries' entire populations.

Directing attention to collectives rather than single individuals, population-wide phenomena required decisions by the governments in place, in whatever form they existed at the time, even if the only anti-epidemic measure known during long periods in history consisted of

simply isolating entire cities and villages, forbidding foreigners to enter and residents to leave.

From the start, while medicine concentrated on the human body, tracking diseases in their hidden sites, concerns with the health of populations expanded wider and wider, incorporating all sorts of knowledge related to the collective. However, historically, the scientific public health perspective had to wait for the development of medical knowledge, particularly for the understanding of etiological links such as transmission, risk factors and vectors. The scientific conceptualization of risk started to be formulated by the eighteenth century (Luhmann, 2008b). Furthermore, the emergence of public health also depended on the constitution of national states and respective governments, applying methods of counting and statistical analysis addressing population's dynamics.

In the last three centuries, the public health field advanced, tackling many common causal factors that affected the risks of becoming ill. The concern with causalities attributable to the environment in a broader sense began to include not only the ecological, biological and physical domains but also the cultural, economic and social realms. The focus on populations eventually arrived at what we now know as public health, with its own sets of statistical and epidemiological models.

As noted earlier, although throughout the evolution of medicine the concerns with epidemics, epidemiology and/or social medicine appeared in several occasions in the past, the field remained underdeveloped while the medical knowledge advanced, but the decisive change occurred when health became a political and legal matter for national states. From then on, public health evolved as a field of knowledge with a vast array of explanatory paradigms with its own self-references. Public health became instrumental for the coupling of health systems with political systems, as political systems faced legitimacy challenges related to the health of their populations.

In short, these two branches of the "same tree" – the medical and public health sub-systems – both based on the same basic binary distinction of healthy/sick, with one focusing on individual bodies and the other looking at risks at population level, evolved together over the last three centuries. Endowed with the initial self-reference of medicine at

its core, which public health complemented with its own self-reference, health as a social system became the configuration and co-evolution of the two sub-systems together.

As with medicine, the scientific and service provision sides of public health are also distinct. The scientific orientation of public health is observable in public health research institutes and departments of medical schools throughout the world. The scientific production is vast. Besides that, public health is the driving orientation of Ministries of Health and public health authorities everywhere. Decisions concerning healthcare service production, health investment, healthcare coverage, effectiveness, equity, efficiency, and many other indicators with focus on needs and delivery of benefits to collectivities, are matters of daily communications by health authorities.

In contrast with the optimization often secured in scientific endeavours due to the specific limited focus and scope of the studies, public health offices have to address as many of the population's health issues in any given country, making optimization of healthcare distribution a very difficult task.

In fact, public health inescapably acknowledges that the difficulty is not only related to the distribution of healthcare but also to the distribution of all sorts of risk factors. We can formulate risks in many different forms – for instance, the risk of becoming exposed, becoming ill, not accessing necessary care, not getting the required quality of care, not achieving the best outcomes, not surviving the treatment. Such objectives, more or less specific or general, fall within the universe of public health concerns. In the sense we use here, risk is broadly also understood as the chance or probability defined for a collective or a given population of not obtaining defined desirable health-related events (be it: prevention of diseases, access to care, access to quality care, cure of the diseases, avoidance of death, and suchlike).

Risks are attributable to populations, not individuals. An individual who belongs to a certain population with a given risk of illness cannot be said to have the same risk of becoming ill; factors affecting a population do not affect every single one of its individuals in the same way. But while

treating a patient, a doctor will consider information about the exposure to risk factors existing in the patient's community.

Furthermore, and very importantly, public health incorporates and further develops the self-reference of the health system as a system. By scrutinizing everything related to needs, planning and evaluation of healthcare provided by the system, public health creates the narratives about the operations of the system itself. The self-observation of a health system is thus performed by public health. In this sense, public health creates the key narratives of the self-reference of the health system. In doing so, public health also generates the visions for the future of the health system and its environment, composing the scenario of diseases, their relevance, risks and priorities, as well as overall aggregated pictures of the burden of the diseases for the society.

As, historically, the self-reference of medicine endowed medicine with the power to revise its own paradigms and pursue better answers to the questions it formulated concerning individuals' sickness, likewise the self-reference of public health pursues the optimization of the tools it uses and its capacity to explain itself as the function of the health system to address (identify and tackle) the health risks of the population, thus explaining to the health system what the health system is about.

Public health always should have something to say in all matters of healthcare decisions for benefiting populations, with estimations of risks of becoming sick, being diagnosed, being treated and obtaining favourable outcomes (or not), and so on. These themes of interest can refer to any level within the health system and the population.

The two "branches" – medicine and public health – grew in different directions, while relying on each other in several respects for what they address and decide. A kind of partnership (in technical terms, the theory calls it coupling) developed between them, observing each other and communicating. As both exist within the health system, they understand each other's codes, sustaining nevertheless the specificity of their respective domains.

When, for instance, a public health sub-system carries out a health survey, it will rely on medical communications about diagnosis and treatment occurring in a defined population. The health system, through

its public health sub-system, can then organize interventions informed by the results of the survey, and medicine can benefit from the knowledge of the health of the collectivity that public health makes available. It may be sufficient to say that description of any particular disease includes epidemiological assessment of the distributions of causes, occurrences and treatment outcome probabilities.

Still, in contrast with medicine and its concrete symbolic medium of communication, the human body, public health does not have a similar material medium with the same level of "materiality", so to speak, and has to rather use the society as a symbolic medium – making forms appear in the medium of collectivities (as indicators of prevalence, incidence, mortality, lethality, coverage, equity, effectiveness, and so on). In that, different from medicine, it often requires both higher levels of abstraction and numerical expressions.

Health systems

A health system has to cope with the fact that these two self-references are not always aligned, and address issues in complementary fashion without friction. For example, a not uncommon situation, while doctors may wish to have at hand all the items necessary for treating their patients, from a public health perspective prioritizing and rationing is often necessary, perhaps saving resources for patients with better chances of being cured.

Within the distinct closure of each self-referential sub-system, on one side doctors report and are accountable to their peers, who can judge the appropriateness of their medical conduct, while on the other side, within the public health sphere, the internal pressures to deliver according to adopted indicators and the respective political accountability are of high stakes.

Such differences are often not easy to reconcile because the two self-references work with different dimensions of time, perceptions of need, expectations, urgency and sense of obligation. For instance, the omnipresent potential of conflicts between insurers and health-

care providers, or insurers and healthcare beneficiaries, or healthcare providers and beneficiaries, are well acknowledged in health insurance arrangements (Costa, 2011). Similarly, but with perhaps lower potential for conflict, in contexts where instead of insurers there is public funding, conflicting interests also cannot be ruled out.

Although comprehensible from a health system perspective, the observation and understanding of the dual nature of health systems' self-references still does not guarantee stable permanent solutions; tension and conflict have to be solved as they appear. The possibility of a subsystem undermining or trying to undermine the other is real. This often requires a third perspective of another observing system, such as the legal system, or the science system, or the political system.

However, while an external system observing the scene can have some degree of influence on public health decisions (as for example when the political system weighs in on decisions on prioritization of health programmes based on trade-offs of cost-effectiveness versus equity), no external observing system can decide on medical matters that are the absolute prerogative of medical professionals.

In this sense, the core of the health system where medicine is located remains untouchable by any other system, preserving the fundamental system differentiation, distinguishing social health systems from all other social systems in the society. This has profound significance for public health self-reference, because public health sits in the convergence of external pressures (from the political system, for instance) but also the internal pressures from medicine itself (assertive in its sovereignty). Figuratively speaking, we can say that public health self-identifies with a buffer zone, cushioning conflicts, but also being a conveyor of tensions and a communication channel of pressures, acting in both directions, internal and external to the health system.

With the theoretical references explained in this and the previous chapter, we can say that our purpose of looking at the history of medicine and public health through the lens of Social Systems Theory is to reflect on the following points:

1) The primary orientation is to look for the revealing stories, where we can see medicine constructing its self-reference as a distinct discipline, establishing procedures to approach diseases and treat them.
2) The historical narratives should allow us to assess the types of social system that could exist at each stage.[5]
3) We should see medical practices appearing historically in the *organization systems* format. Organizations may have had diverse purposes, such as bringing doctors together to protect the craft (guilds); as initiatives for organizing provision of medical care in hospital facilities; as training arrangements between students, masters and, later on, in universities; as bodies such as collegium, councils or associations, with self-regulatory roles.
4) In its historical development, medicine preceded the establishment of health as a social system. Medicine developed its specific semantics well before the emergence of health systems.
5) We should see medicine at some point becoming legally, politically and socially accepted as a legitimate field of knowledge and practice. That included the recognition of the prerogative of using medical semantics.
6) We should be able to observe that the restricted use of medical semantics did not require the assignment of responsibilities to any specific overseeing medical institution. We may find the exercise of valid medical communications taking place in many different sites, wherever medicine could recognize itself in line with its self-recognizable rules of communication.
7) Evolving from that, however, we will be able to see the correspondent semantic closure, by which medicine came to be fully in charge, to ex-

5 To briefly recall, there are three types of communication-based social systems: *interaction systems*, *organization systems* and *function systems* (see Chapter 2). Conditions at each stage of historical development may have been adequate for the existence of one type of system but not for the others. Function systems could only appear when functional differentiation became the way societies structured themselves (which happened around the second half of the eighteenth century).

clusively ascertain the validity of medical communications. For that, university courses, as well as the guilds, surely played the crucial role.

8) In that process, we see the embryonic social health system establishing uniform stable regulations for medical training, licensing and professional recognition.
9) The historical assessment should thus reveal the stage when and how medicine became part of *social function systems*.[6]
10) We should gain insights into *health* achieving *systemic form* with the recognition by other systems (particularly the political and legal systems), with medicine as an integral part of the established *health system*.
11) In reference to the theme of complexities, each historical period should be characterized by the complexity of practices and knowledge. Historical evolution is the history of the advent of increasingly complex models, practices and settings.
12) At the same time, evolution also required reduction of complexities in order to avoid crossing threshold beyond which internal and related external complexities could become overwhelming for the system. A system's self-reference reflects the level of complexity the system handles and the strategies adopted to keep complexities manageable, expanding or reducing it accordingly.
13) Therefore, we should identify the expedients for reduction of complexities in the historical development of health systems. Among such strategies, internal regulations play a key role. The components of the system should exercise regulatory roles in accordance with their self-references.
14) The historical narratives also talk about medicine and the advent of public health as sub-functions of the emerging health system. The construction of the health system, we will see, corresponds and is sustained by the combination of those two sub-systems' self-references, one concerned with individuals (medicine) and the other with

6 Generally speaking, a function system appears with its differentiation from other function systems, whereby each system has the respective orientation, distinctive code and self-reference.

collectives (public health). The self-reference of the health system is comprehensive and reconciles this dual self-reference.

15) Correspondingly, the advent of a public health sub-system brought the social into the health system. The health system became a *social health system* as the health of societies became part of the system's semantics, and thus part of the communications that the health system sustained and reproduced;
16) Furthermore, the health system should continuously couple with other *social function systems*. For that, we try to see how the public health sub-system acquired the crucial task of developing the self-description and self-identity of the health system, explaining internally and externally what the *health system* is about.

These points are covered in the subsequent chapters. We hope that the text is clear on these subjects, which are often rather complex.

The reader will find occasionally that we need to bring in some of the theoretical topics presented in this first part of the book. We may add here and there theoretical elements to explain better the issues at stake.

Nevertheless, the themes of complexity and self-reference do reappear often in the discussions. This is essential, as they are the key concepts for our discussions.

PART II

Chapter 4: The History of Medicine in Four Acts

If we take the Hippocratic texts of the fifth century BC as the starting point, as is widely accepted, medicine has a history of more than 2,500 years. However, with the registers of attention to diseases and curative acts described in ancient Egypt, China, India and Mesopotamia added, medicine is perhaps more than twice as old.

The consensus on the Hippocratic texts as the birth of Western medicine is justified due to the originality of the deployment of principles of systematic observations and rational considerations about the human body and its diseases, without resorting to Gods, magical influences or flimsy causality assumptions. The medicine invented in the fifth century BC is therefore scientific medicine, as we still understand it today: empirically based and rational.

The 2,500 years of history is rich in breakthroughs and remarkable advances, replacing improvements of previous periods but also continuously selecting and preserving accepted knowledge and practices as deemed appropriate.

There are many ways of telling the history of medicine or dividing it into periods. The history we are about to tell is divided into four periods: the first from the fifth century BC to the fifth century AD; the second from the sixth to the fifteenth centuries; the third from the sixteenth to eighteenth centuries; and the fourth from the nineteenth to twenty-first centuries. The reason for these divisions is our perspective of analysis; these periods offer clear distinctions when linked to the evolution of societal structures as proposed by the Social Systems Theory.

The first two periods happened within societies with *stratified differentiation*, firstly, in ancient Greek and subsequently Roman societies (first period), and secondly in the Middle Ages. The two periods unfold from the pinnacle of Greek civilization to the fall of the Roman Empire, and after that under the domain of the Catholic Church. Although within the same form of social differentiation, these two periods present important differences on how medicine was viewed and practised.

The third period is the transition from stratified to functional differentiation. In this period it is possible to see medicine acquiring features that would subsequently lead to the establishment of health systems as fully differentiated function system, in societies where several function systems coexist in each other's environment.

For each period we describe the central themes of medical self-reference, giving summary descriptions of what those themes entailed. More specifically, for each of the four periods we give: (1) a brief characterization; (2) the system's features, in line with the Social Systems Theory; (3) a discussion on its complexities; and (4) a summary of medicine self-reference of the period.

As usually told in textbooks, the history of medicine is the narrative of scientific progress in the medical field. The relevant events and players are the breakthroughs and the scientists who made them. Many but not all were doctors, but their discoveries had far-reaching consequences on the practice of medicine as social function and professional field dedicated to the treatment of disease and the preservation of health.

In Chapter 1 we addressed briefly the difference between the function system of science and the function system of health. These systems started to be constituted as social function systems around the seventeenth and eighteenth centuries, a period when societies were not yet structured according to the function differentiation that would fully emerge in the early nineteenth century.

The separation between practitioners and scientist doctors was never fully marked before the sixteenth and seventeenth centuries. The science and practice of medicine were undistinguishable until then, although only those who had scientific orientation towards the sys-

tematization of new knowledge became recognized in the history of medicine.

The two realms, science and medical practices, overlapped and showed couplings; and they still do, as when, for instance, experimental treatments are performed in controlled trials, and when practitioners make daily records of their patients as data for health surveys or epidemiological studies.

As mentioned in the previous chapter, the differences are nevertheless clear, when on one side the doctors deploy the knowledge already established, with the means available to them at the point of care, and on the other side the doctors/scientists try to expand the overall knowledge of specific diseases or bodies' functions. The time horizons and intended effects are different, one being of immediate consequences for an actual patient, and the other in principle aiming to benefit the whole human species.

Although the history of medicine is the narrative of sequences of scientific progress, the actual history of medicine is also the history of service provision and part of the history of the evolution of health systems, from embryonic stage to full development. The history of medicine is also the evolution of the self-reference of medicine, leading eventually to the self-reference of medicine within the self-reference of health systems.

First period (from the fifth century BC to the fifth century AD). Craftsmen's medicine, observing the surface of the body at the bedside

Characteristics of the period

The period from the fifth century BC to the end of the Roman Empire (fifth century AD) starts with the Greek texts attributed to Hippocrates from Kos and anonymous authors whose works comprise the Hippocratic Corpus collection. The texts strongly influenced the practice of

medicine in Persia and Alexandria and later in the Arabic-dominated regions, as they were translated into those respective languages.

Among the ancient societies that developed the capacity to write complex observations and thoughts, the Greek civilization showed remarkable achievements in many areas.[1] This is also where the interest in making systematic records of observations made at the bedside of those suffering from illness first appeared. It is therefore called the birthplace of medicine as we know it: a rational practice based on observed evidence seen in the human body.

It all started, then, with the collection of countless written records of patients being observed and treated by those doctors. Hippocrates formulated the style of attention and objective recording of observations, setting in motion a definitive innovation from previous healing practices. No longer seeking explanations in or summoning help from supernatural forces, the declared Hippocratic intention was a bet on the future, hopeful of better understanding of what was being precisely recorded.

In his outstanding scholarly study of the progress of Greek medicine (on which most of our comments in this section is based), *The Invention of Medicine, from Homer to Hippocrates*, Robin Lane Fox (2020) focuses on Books 1 and 3 of the Epidemics volumes in the Hippocratic Corpus, which he sees as following the same style, suggesting they were written by the same author, possibly Hippocrates himself. The Corpus is a collection of about 60 books, written over a period spanning perhaps 10 decades, from the mid-fifth century BC to the mid-fourth century BC, admittedly written by several authors with different styles.

About those two books, Fox observes the consistency of the simple and objective language. The texts acknowledge that causes and treatments of diseases were not yet known and all that could be effectively

1 We make a small digression to suggest that perhaps linguists may have an explanation, linking the versatility of written Greek and its potential to create textual forms translating complex observations as well as abstract ideas, something that, perhaps, the hieroglyphic pictographic representations would not permit. This is just a hypothesis, to be discussed or dismissed.

done was the written record of everything that seemed to be relevant from what was observed on the surface of the bodies of patients. The observations could not go deeper because the knowledge of human organs and physiology was poor – and basically wrong. Furthermore, dissections of human bodies were forbidden and the only knowledge of anatomy available was acquired from observation of animals or soldiers wounded in combat (as described in Homer's poems before Hippocrates).

Nevertheless, Fox presents translations of cases and we can appreciate the concision and objectivity of the records. Based on those simple, superficial descriptions of signs and symptoms, some cases can be tentatively classified with the current nosology.

The texts became fundamental for those interested in the practice of healing. Printing did not exist and books were written by hand. They were very expensive and hard to find. Teaching activities were organized around individuals recognized as knowledgeable doctors, who would have copies of the texts in which students could read descriptions of the progression of signs and symptoms. Medical training was mainly available to those who could afford to pay the master. Because literacy was not widespread, access to the medical texts was in fact the privilege of a small minority who therefore kept the knowledge within well-off families and groups. As in any other craft, medicine required a long apprenticeship.

Historically, once Roman hegemony was established, Greek medicine migrated to Rome, which did not have a developed medicine of its own. Rome and Alexandria in Egypt became the centres of medical knowledge of the Western world. The pattern of direct observation for prognosis initiated by the Greeks was thus preserved.

As valued practices performed by craftsmen trained by other craftsmen, medicine was an individual business service, mostly sold to those who could afford it. Anything similar to organizations with the exclusive purpose of providing healthcare would consist of no more than a small group of doctors, often from the same family, working together.

Medicine was practised at the bedside in patients' home, although some doctors had their own facilities. That period also saw the estab-

lishment of what could be called rudimentary "hospitals", set up for the treatment of wounded soldiers, and in some cases for containment of diseases considered dangerous ("leprosy" for instance), those infected were kept confined but without treatment.

As already noted, dissection was forbidden and the understanding of anatomy was extremely basic and often wrong, based on observations of the surface of the body. In second-century Rome, Galen (who upheld most of Hippocrates' teachings, approaches and principles) added anatomical features not seen in the Greek texts, with a remarkable although still rudimentary understanding of the circulatory system. However, his works contained many misrepresentations and gaps, such as the absence of reference to the nervous system.

The survival of the distinctive Hippocratic approach of systematic direct observation and descriptive writing of the sequences of events, day after day, to the end of the case (mostly with death), and the limited ambition and scope of the interventions, is indeed remarkable, considering the social context where religion was a strong presence in the daily life. Charlatanism was widely accepted and practised with mixed prayers, religious beliefs and alleged healing practices, with no regard to the written texts. The mixing of religious beliefs and healing practices undifferentiated the medical craft, therefore delaying the closure of systemic self-reference (more on this later).

Medicine self-reference was almost non-existent, in the sense of regulations and supervision of practices. Anyone could do anything. Knowledge was not controlled and organizations did not try to assure consistency across practices and reliability of those calling themselves doctors. The acquisition of "respected doctor" reputation was a process to which the existence of powerful patrons (such as kings and nobles) and the recognition of a certain posture in relation to patients and diseases were fundamental.

Nevertheless, to become a doctor in the tradition established by Hippocrates, the aspirant would be orientated to voluntary adoption of the Hippocratic Oath and observation of Hippocrates' Aphorisms. In a context without regulatory authority controlling or overseeing practices, practitioners were left with individual or small group self-

observation of the principles of the craft; "Do no harm" was one of the most famous. But charlatans could also pretend to follow the oath.

There are references to "schools of medicine" in some books on the history of medicine. For instance, there are references to the "empirical school" and the "dogmatic school" and also to the "methodic school". We may interpret that as a way of explaining in simple terms the separation of different orientations. However it is widely accepted that the term school by no means resembles the current use of the term. Any differentiation between schools would be precarious and with little basis in distinct observations of the body or distinct therapeutic methods, never mind regulations. The therapies available were basically the same and the explanations about the cases were strongly dominated by simple frameworks such as the four humours

Indeed there are records of authors from the same period who drafted different notions, such as Alcmaeon of Croton (fifth century BC), considered the first philosopher of medicine (who carried out animal dissection and possibly, it is speculated, of humans too), and emphasized the importance of the solid parts of the body, not only the fluid humours. However, even with the formulation of distinct frameworks (like those related to seasons, or atoms, or "critical days" in the prognosis of the disease), these views would share the same ignorance of the complexity of the body and its reactions. There is no basis to see those alleged currents of thoughts as being the same as what may have developed centuries later as "schools", created around distinct theories and empirical studies.

Treatments were simple and in the common knowledge of everyone. The recognition of a person as a doctor was based on the fact that he (medicine was almost exclusively practised by men) knew some healing procedures, and had used them before. They were practising their craft on a daily basis as opposed to the rest of the population, which was only occasionally faced with the need to treat diseases. However, the items and activities prescribed by doctors were commonly available and used by those who could not pay for doctors' consultations and the prescriptions they sold. The temples where the God doctor Asclepius was worshipped were open to all who preferred to resort to prayers and offerings.

Bloodletting was perhaps the only procedure that required a doctor, although a charlatan could certainly perform it too. Hippocrates did not have a favourable view of this procedure though (see Craig, 2021).

It is important to emphasize that the medicine of the first period had very few therapeutic resources to deploy. Basically, everything a doctor could prescribe or do was widely known: plants, baths, massages, purging, exercise (or its reduction), diet, abstinence from sex, alcohol or specific foods, and so on – simple prescriptions in the public domain. The doctors nevertheless were those knowledgeable about the diversity of resources. They had seen many sick people and could therefore know better what seemed to work in particular situations. Anyone who could not afford to pay the doctors could still use similar treatments according to popular understanding. Everyone had easy access to items employed in treatments by cheap charlatans. In general terms, it was also widely accepted that cure was nature's business and no treatment could succeed against its determination.

System's features

We cannot speak of health as a *social function system* in this period. Medical practice was unregulated, open to any craftsman with the inclination to claim competence in healing people. The only identifiable social system could be characterized as *interactions*, which, as explained in Chapter 2, consists of social systems that require face-to-face direct communication between the people involved. Interaction systems start and end as soon as the conversations finish. The doctor could make records of his observations, but that would not necessarily go into standardized registries or archives to be used in other communicative processes. When made, the collection of texts were intended to help with similar cases the doctor might encounter, and to aid other doctors in the cases they were called to treat, but they were not medical files as we know them today, kept for a variety of purposes such as managerial, legal, aide-memoire for follow-up.

We can identify the following aspects as being distinctive of a social health system, in an embryonic stage, taking its first steps towards social

differentiation: (1) the vocabulary (semantics) that was then becoming common currency among those who could perform the craft, endowing them with a distinct way of talking about diseases; (2) the social identification of the craftsmen as performing a valuable and distinct service; and (3) the very first initial step to circumscribe and fence the practices by promoting a self-referential code of conduct. The adoption of the Hippocratic's Oath was indeed an extraordinary self-regulatory orientation with powerful self-referential effect, revealing a self-conscious category of craftsmen motivated to preserve a distinctive character, which corresponded to the elitist nature of the literacy and expensive training required.

Still, on the treatment side, none of the promoted and practised therapeutics was enough to establish health as a social system. It can be said that in fact they would be rather a hindrance, given the fact that the treatments were of common use, making it difficult to turn them into specialized exclusive knowledge, as necessary to establish a social *function system*.

As *interaction* was the prevailing social system structure, and organizations were not set up for exclusive provision of care or control over the use and practice of healthcare, the overall configuration was in this sense loose, with widespread presence of charlatans, who did not read the texts and did not follow the observation approaches recommended in the texts.

This configuration thus fitted well with *stratified social differentiation*, in which, between the upper (dominant) and the bottom (dominated) strata, there was space for individual unregulated craftsmanship and provision of services to those who could pay for them. Among carpenters, metal workers, bakers, builders and others, doctors could find their niches and sell their services in the markets and fairs, as well as on their own premises or in patients' homes.

In conclusion, there was no health system as a functionally differentiated system. Society was not yet differentiated according to the functional principle, as it would only start to be from the eighteenth century onwards; therefore, the correspondence to existing social structures from which health could copy its social functional outlook were missing.

Complexities

Doctors were confronting a diverse range of complex signs, equipped with only simple explanatory theories, such as the four humours framework, helping them to link evidence with a rudimentary explicative model of causation of diseases and functioning of the body.

To reduce the complexities of patients' observed signs and symptoms, that simple framework was nevertheless helpful. The four humours framework also introduced the "interactive process" of moving from complex evidence seen in the body to the simple framework and vice versa, submitting the framework to "stress tests" at the same time that the observed complexities were "tamed" (or "hammered") into the schematic explanations. Such discipline of approaching reality and communicating about it turned into effective methods of dealing with complexities; the explanatory framework as well as any subsequent frameworks would have to pass the same sort of pragmatic tests, with outcomes confirming or not the correctness of the suppositions.

As autopsy could not be performed, the knowledge of internal organs was limited to what could be observed in wounds and animals; consequently, suppositions about internal organs affected by the diseases could not be disproved or confirmed after death. Medical studies were therefore left with what they could observe outside the body and make their inferences based on the four humours scheme, adding ad hoc references to possible effects related to the environment, seasons, weather conditions, behaviour excesses, and so on. The complexity of medical knowledge therefore was limited for lack of opportunities to increase it by examining the insides of bodies.

Nevertheless, the systematic recording of observations of the body and the course of the diseases produced a large quantity of records with rich and complex information. That complexity was further increased with subsequent records made by Galen and his followers, including later on the incorporation of records from the Arabic and Eastern worlds (such as India and Persia).

The complexities of the observed manifestations of the diseases are noticeable in written texts. Specific vocabulary had to be developed. De-

spite the paucity of knowledge and tools, and the limitations of what could be done with patients during examinations (touching, listening to the body's sounds, measuring were very restricted), the list of what needed to be observed comprised a long inventory. The list included, but not exclusively, voice quality; colour and texture of tongue; body temperature increases with or without shivering (thermometers obviously were not available); skin appearance (including rashes, swelling, colour, temperature, jaundice); temperature of extremities; rhythm and quality of breaths; capacity to eat and drink; appearance and quantity of faeces; appearance of urine and presence of alien substances; vomit (appearance, quantity, and so on); sweat; pain and complaints; mental state (delirium, talking nonsense, loss of reason); sleep pattern; spasms; refusal of food and loss of appetitive; appearance of eyes (gaze); sex life; exercise; excretion of the four humours (bleeding, yellow or black bile, phlegm). Observation points on the body were numerous, and the registry of empirical evidence had to be simple, direct and continuous, covering the course of the disease on consecutive days until cure or, more likely, death.

The list hints at the complexity of the approach, with many aspects to be observed. The diversity of elements and body parts represented a significant challenge to keeping comprehensive and accurate records. Diagnostics representing sets of symptoms, without describing each individually, were not available. The orientation thus was fundamentally descriptive rather than attempting to make connections between conditions, causes and treatments. It is understandable that for lack of explanatory models and actual understanding of internal structures of the organs, anatomy and physiology, not much could be done apart from describing the superficially observed signs.

Meanwhile, the number of recorded cases resulted in a sizeable collection of observations made available for those interested in studying the craft. The interested could move between the text and the patients, checking new observations against those previously made by others. The capacity to distinguish cases increased as well as the acknowledgement of other aspects still not recorded. As mentioned, the recording followed the patients throughout their illness, with daily notes.

It is understandable that the sheer mass of such records created conditions for qualitative jumps, whereby new understandings of the human body were devised to accommodate the accumulated evidence. Galen (second century), during the Roman Empire, developed more complex concepts of the functions of the internal organs inside the body. While adopting Hippocrates' texts as his basis, particularly the theory of the four humours, and preserving the prominence of observation, he expanded the sets of elements to be considered.

We can understand that as the recorded complexities of the signs of disease made their recognition easier, and at the same time required dedication to learn how to track the signals, the pressure to develop classificatory and explanatory models increased. The need of *complexity reduction* strategies included the necessity for a consistent and unique vocabulary, with terms that could be maintained and preserved in all the places where medicine was being practised, ensuring permanence and stability of the communicated meanings. Such progress could only be achieved with written texts, which could thus further increase the complexity of what was transmitted to future practitioners by selecting relevant meanings from noise.

We can say that the critical mass of complex information in this period was crucial for: (1) the creation and preservation of a specific semantic with specific prospects for communication about diseases; (2) making undeniable the need for setting up educational structures and processes to reproduce the knowledge and meanings of the new semantics; (3) strengthening the epistemological pressure (*complexity reduction*) for the creation of frameworks and theories to organize, support, explain and schematize the recorded complex observations; (4) making possible and necessary the subsequent emergence of self-referred professional activities to move forward in the direction of establishing formal institutions with self-regulation roles, separating the authentic exercises from the fake ones.

Summary of medicine self-reference: first period

We list below what we believe were the ways in which doctors would self-describe thinking and practices, and would use to identify each other as practitioners of the same craft.

- Medicine was recognized as a rational and non-religious discipline based on how to register observations of the human body and its diseases. Direct empirical observations were conducted without instruments, formalized procedures or methods of testing. Medicine developed its initial self-identity orientated to perform simple observations, describing what was seen without appealing to deities or magical influences.
- Based on accumulation of recorded cases, prognosis of how a disease would unfold could be made. The discipline of collecting reports was predominantly intended to allow identification of patterns of progress of the disease, thus making prognosis of outcomes possible.
- The interpretation of observations could refer to simple causality models: (1) the four humours with notions of balance among them and the disruption caused by excesses; (2) the doctrine of critical days, through which the diseases would follow patterns of improving and worsening, according to the number of days from the start; (3) possible influence of weather, given the seasonality of some conditions, with systematic records of the seasons and weather conditions during the course of the diseases. Galen made more complex models of the functioning of the body, though preserving the Hippocratic theoretical framework of four humours. Such frameworks comprised the main general semantics available to interpret possible causal influences, but the overall approach was mostly orientated towards describing the condition rather than explaining causes. As noted, the focus was on descriptions and prognostics.
- Medical communication was mainly between doctors and patients, and among doctors themselves, essentially those being trained and their masters. Doctors would teach apprentices about the craft, and were paid for that, in the context of their own practices, without

formal regulation of training and certification. Medicine's self-reference therefore was constructed and reproduced through those contacts, whereby doctors would self-observe what they themselves were doing and writing about. No overseeing institution, internal or external to the field of medical practice existed.

- Doctors could practise their craft as other craftsmen would – that is, in marketplaces, at fairs, in the houses of the patients or workplace of the doctor, if they had one. Medicine was thus recognized as a service provided for a fee, like any other business, on an individual basis from doctor to patient.
- Therapeutic use of plants, diets and other resources were common knowledge. The widespread assumption was that nature had the monopoly on healing powers; only it could heal or otherwise, and healers were only supporting actors. There were no controls or regulations over the treatments, or specific settings or sites. Plants were known to everyone and not specifically produced or controlled by any system. Doctors could cultivate medicinal plants in their own gardens and produce the medicines they used in their practices; there was no restriction in that regard either. Therefore, preparation of doctors' own medicines was part of medical practice. Doctors could also acquire plants from regions they visited; travelling doctors with medicinal herbs and other resources carried from places they had previously been was a feature of medicine self-reference – travelling doctors recognized themselves and were recognized and valued as such by the population and authorities.
- The importance of following the code of conduct (Hippocratic Oath and Aphorisms) was introduced as self-regulation of the practices, providing a self-reference base and identity for those engaged in medical activities, independent of legal or religious norms. However, following the oath in practice was left to the discretion of the doctors with possibly only scant mutual observations among those practising the craft together. The oath probably had some effective regulatory role among doctors who knew each other, but it would not prevent a doctor from moving on to practise elsewhere, when caught doing something against the principles.

Second period (from the fifth to the fifteenth century). Medicine of books, universities and hospitals

Characteristics of the second period

What fundamentally distinguishes the second period from the first are the effects of unprecedented accumulation of recorded observations of diseases, which started in the first period and continued throughout the second. The descriptions were compared, analysed and checked, with systematic collation of observations of patients and text descriptions, from the text to the body and back to the text. New descriptions were made. The availability of such rich materials generated complex views of the human body. Built on the Hippocratic Corpus and his own observations, Galen's contributions at the end of the first period were diligently preserved and studied throughout the second period and beyond, for almost two millennia. This period also saw the establishment of institutional innovations of great relevance such as medical schools and hospitals.

The ten centuries from the fifth century AD (after the fall of the Roman Empire) up to end of the Middle Ages (fifteenth century AD), could be divided into two sub-periods: the early medieval period, with no significant progress in the European region, and the late medieval period, with the establishment of medical schools in universities.

In the first five centuries, known as the "Dark Ages", many scientific achievements regressed to previous stages, with strict control of the Church upon the expansion and dissemination of scientific knowledge. Medicine returned to superstitions and the Church's endorsed assumptions that health was God's business and therefore ailments were to be accepted as such. In terms of provision of curative services, during almost the entire period anyone could claim to be a healer and sell services without the Church's consent, let alone endorsement.

In the European Middle Ages, however, medicine did not disappear and continued to be practised as a craft, in the same way as previously, using plants, massages, bed rest, heat, gentle exercises, cupping, purging, bloodletting, and so on. Tolerated by the Church, practitioners were

less concerned with observing and recording cases, in contrast with the procedures started by the Greeks.

However, also characteristic of this time is the appearance of charity "hospitals" – Church-sponsored places to accommodate the destitute in general, including the poor, the sick, pilgrims and those with nowhere else to go, with everybody gathered in the same spaces. These "hospitals" were often attached to monasteries, and assistance was provided by monks with some medical knowledge, without the specific objective of studying and systematically treating diseases. The monasteries had their own medicinal herb gardens for those in need of remedies, thus preserving valuable knowledge. Around the middle of the second period, by the tenth century, more advanced establishments, specifically characterized as places for medical assistance, appeared in the Islamic world.

While medieval Europe lingered on through the dark ages, progress was made in Byzantium and Alexandria (where copies of the original written Greek and Roman documents were kept), and beyond that particularly in areas of expansion of Islamic civilization, with an important increase in medical reflection and production of written works. Baghdad, Damascus and Cairo, centres of the Islamic world and fertile ground for the advancement of medical knowledge, saw writers and historical figures such as Rhazes (with advances in differentiation of diseases), Avicenna (with his *Canon of Medicine*, still used in traditional Islamic medicine) and Averroes (with encyclopaedic texts covering all known medicine at that time). These writers were prolific, producing dozens of written works, mostly strongly influenced by the Greek texts and keeping the prominence of direct observation. Many of the texts did not survive the period though.

Back in Europe, in contrast with the Church authority and restrictions on doctrinaire grounds, the first sub-period is also characterized as "library medicine" by Bynum (2008), highlighting the major contribution to medicine originating from the preservation of copies and translations to Latin of the original Greek, Arabic and Persian medical texts. Such activities were mostly carried out in monasteries by monks, and later on in universities. The reproduction of those materials by hand in the monasteries was crucial for the preservation of the knowledge thus far gath-

ered. Although dissociated from any systematic process of transmission or teaching, the preserved materials became crucial for the later education at medical schools in the universities. The preservation of texts set the basis of a complex field of knowledge and exclusive competences.

In the second half of the period, the recognition of monastery hospitals as spaces for medical studies transformed their outlook and accelerated the process of gathering observations and descriptions of diseases, contributing to the characterization of doctors as having distinct a social role with distinct social spaces for practising.

When in early medieval times (thirteenth century) the Church forbade clerics from getting involved in the practices of medicine, most specifically surgery, due to the uncomfortable contrast with religious teachings about preservation of the integrity and untouchability of the human body, medicine was progressively detached from the Church. Without that barrier, medicine fully entered the scientific domain as knowledge to be learned at universities, creating the possibility of linking the existing monastery "hospitals" to universities' medical courses for treatment and study of diseases. Such processes became crucial for the eventual acceptance of autopsies as a method of learning about the body, the diseases and the causes of death, with the first medico-legal autopsy carried out by Bartolomeo da Varignana in 1302, and first public dissection for anatomic studies performed by Mondino de' Liuzzi (around 1315) Bologna.[2]

The first recognized medical school was established in Salerno (in the eleventh century), followed in the subsequent century by the schools of Montpelier, Padua and Bologna. Salerno gathered a small number of students and teachers, but found a more open terrain for reflections. Two centuries later, schools were established in France and in Spanish cities.

2 Although there are historical references to autopsies conducted in Alexandria in the third century BC by Chalcedon, the references to autopsies in animals and humans by Galen (second century), and the anecdotal story of the post-mortem examination of Julius Cesar, including identification of the fatal stabs, the practice was largely forbidden in the first period.

In the Islamic domain (from the eighth to the fifteenth century) in Cordoba and Toledo, medical manuscripts were translated into Latin.

The medical schools of the period were mostly dedicated to studies of the classics, and educational processes involved discussions, debates, speculations and interpretations of the texts, with little in the way of practical exercises or generation of additional medical knowledge. Nevertheless, the model of medicine-in-universities spread out and several schools in medieval Europe introduced medical courses towards the end of the period. By the twelfth century, medical schools existed in the universities of Paris and Oxford. According to Bynum (2008), by the late fifteenth century there were 50 medical schools in Europe.

Still, training was only for those who could afford to pay and gain access to copies of the texts. Although texts started to become available in Latin and copies were produced in larger numbers, they were not yet printed. However, as the required literacy was only possessed by the privileged few, only a small minority could have access to and read the texts.

The last centuries of the period (fourteenth century) saw huge plague epidemics, which had devastating effects on Europeans, with deaths estimated as ranging from 25 per cent to as much as 50 per cent of the European population. The epidemics nevertheless brought concrete initiatives for controlling and preventing outbreaks, although still without any knowledge of infection and transmission processes. The measures consisted mostly in controlling access to cities and regions by foreigners, disposal of dead bodies and rudimentary quarantine methods, which nevertheless was undoubtedly the beginning of public health. In fact, some of those measures had already been put in practice centuries before during epidemics in the Greek and Roman worlds.

System's features

Similar to the previous period, *interaction* between doctors and patients was the prevailing social system seen during these ten centuries. *Organizations* as social systems were not set up for provision of care or control over healthcare resources; the existing organizations were rather religiously sponsored charity "hospitals", for accommodating all sorts of in-

dividuals, the sick as well as the poor, pilgrims and homeless alike. These "hospitals" were not medical organizations. It was only by the end of the period that hospitals started to become distinctive institutions where healthcare was systematically carried out. Links to university medical courses also started then.

From the social systems perspective, we see a combination of those three major developments: the preservation and translation of medical texts; the opening of university courses; and the recognition of hospitals as sites for caring for the sick. These had determinant effects for the development of conditions for the constitution of health as a social system.

Concretely, these points can be emphasized: the texts further deepened and expanded the exclusive semantics and specialized language; the hospitals made possible the social recognition of practice of addressing diseases in institutionalized settings using professional expertise; and the universities' medical courses initiated the possibility of a clear distinction between those who could and those who couldn't appropriately claim the right to practise the craft, building the pathways for selective membership in the future health system. All three contributed to set the preconditions for the system.

Exclusive texts, specific universities' courses and hospitals moved medicine in the direction of a system with its own codes, and corresponding autonomy for self-reproduction of codes and communications. These innovations brought about reproduction of practices and reproduction of knowledge through organized communications among peers and students.

Medicine started to enjoy an exclusive privileged environment for communicating about the human body and diseases, protected from external interference, and at the same time achieving social recognition and legitimacy for claiming and using such social spaces. In this still unstable but already exclusive domain, medical models and frameworks could start to be discussed, doubted, criticized, revised, tested, retested and approved or dismissed. That was fundamental for the future emergence of health systems.

Other medical currents of thought, such as those developed in the Islamic world, found a place for recognition and further consideration.

In the Salerno school, for instance, although in the area of influence of the Church, there was openness to studying medical texts of diverse origins such as Greek, Arab or Jewish. Constantine the African was a key figure in translating several books into Latin in Salerno. With that we see medicine taking critical steps towards assessing and comparing paradigms in a self-referential movement that could not be carried out outside medicine itself.

Towards the end of the period, an additional step was the establishment of organizations such as the medical guilds, in the same model as the guilds of other crafts. These organizations played the role of self-regulators, preventing strangers from introducing illegitimate codes and meanings from other fields into the realm of medicine, thus averting disruption of the practices commonly agreed among doctors. Such interference by strangers could be detrimental for the recognition of medicine as a valid, valuable and well-defined field of practices and knowledge. The guilds created decisive fencing for the systemic closure, accepting as members only those who had completed university training. By the same token, the guilds represented a significant experiment in self-regulation to subsequently take more legally and institutionally consequential roles.[3]

Until the appearance of medical guilds at the end of this period, the context was of deregulation and laissez-faire, with both charlatans (without medical training and knowledge of the medical texts, mixing healing practices with religion and superstition) and doctors profiting from business opportunities. A health system as a functionally differentiated system could not find a base in such a context of general deregulation and openness. However, although weak, self-reference among doctors who had access and studied the texts was developing into self-recognition of a distinct valuable practice. Such self-references were conducive to subsequent broader social recognition.

The opening of schools of medicine at respected universities a therefore major impact on the self-reference of medical expertise, becoming

3 On the medical guilds, see McLaughlin (1941).

not only crucial for self-identity but also for establishing common recognition of medical knowledge and practices as distinct, unique, scientific and complex. This was of vital importance for the self-reference of the practising "communities of professionals", a condition for identifying themselves as holders of uniquely privileged knowledge. Fundamental elements for a functional health system were thus launched.

Complexities

The numerous descriptions of cases of disease gathered over the previous centuries embodied formidable complexity, exposing and pressuring the limits of the existing interpretative and explanatory frameworks. With details of diseases increasingly described and the new semantics revealing complex observations, the use of the by then extensive vocabulary became progressively more exclusive to a limited number of those who could read and communicate the meanings. As already noted, this process culminated in and took decisive direction with the opening of medical courses at universities, which both constrained access to the specialized language and strengthened its usage among professionals, thus creating conditions for increasing the complexity of medical narratives.

Among the examples of how the accumulations of case descriptions led to more sophisticate explanatory models are the representations established by Galen in comparison with Hippocratic texts. Although Galen still preserved the four humours framework, his texts included more elaborate descriptions of the circulation of blood through the liver, heart and brain, considered the most important organs of the body. The focus of attention therefore started to divert from humours to the organs; that turned out to be a decisive move in the direction of locating diseases in the solid organs instead of volatile fluids.

Medical knowledge was thus striving to incorporate the observed complexities into its descriptions. The frameworks adjusted to the complex observations, opening the possibilities to more complex ones. In a way, the semantic system became prepared for supporting communica-

tion of huge complexities continuously unveiled, once the restrictions over dissection were lifted and looking inside the body was permitted.

As said, the complexities of the bodies could start to be addressed more thoroughly in a consistent and systematic manner as medicine found a place for communicating about complex observations while studying patients and corpses. The universities offered sites for discussions in parallel with the expanded opportunities for diving deeper into the body that the hospitals provided. In short, as a result, medicine was preparing to become more complex in its conceptual and semantic arsenal; complexity was developing in the two sides: the observed bodies and the words assembled to communicate it.

Furthermore, the adoption of the organizational model of guilds (similar to other crafts in urban centres) signalled the potential for reduction of the complexities of a multitude of practitioners performing their craft without any common, consistent orientation and supervision. The delimited gathering of doctors enabled continuous mutual observation and allowed knowledge harmonization. Instead of each one acting as they individually wished, the importance of collective confirmation and acceptance of the procedures and corroboration of what each was doing became clear. Reducing the complexities of uncoordinated and "free-for-all" professional practices, and establishing a sense of collective identity, prevented the chaos that could have prevailed.

Summary of medicine self-reference: second period

- Medicine in monasteries: Initially subordinate to the Church view that health was God's business, medicine in monasteries was medicine for members of the Church, for the privileged few who could afford it, provided in a comfortable setting (with access to the herbs cultivated in the monastery and good accommodation), as well as medicine as a charitable service for the dispossessed, mixing rudimentary treatments with general support for living. That was a rather simplistic notion of medicine.
- Medicine as knowledge of the body as described in copied texts of the previous period and translated from diverse languages and geogra-

phic origins (Indian, Persian and Arabic worlds), strongly influenced by the Hippocratic Corpus with Galen's additions.

- Medicine taught in a few newly established (eleventh century) university courses, to train small numbers of those who could read and write texts in Latin – the medicine of literate and university-trained doctors.
- Medicine in hospitals initially created for wounded soldiers and later on opened as charity for the needy (poor, sick, pilgrims, homeless) and also victims of epidemics. Hospitals not as medical institutions, attached to monasteries, but progressively linked with university courses.
- Towards the end of the period, medicine as science authorized to perform dissection and studies of anatomy on dead bodies in universities and hospital settings (University of Bologna fourteenth century).
- Medicine as an increasingly complex practice of university-acquired competence; gradually circumscribing and limiting access for those willing to practise, setting limited terms of acceptability on the craft and the entitlements and responsibilities for medical knowledge reproduction.
- Medicine of the newly established guilds, bringing together doctors in associations of common interests, intended to protect the craft from external interference and charlatans, adopting the organizational models of guilds of other crafts in urban centres. The guilds contributed to reducing the complexity of a multitude of practitioners performing in isolation without coherent and consistent orientation or knowledge sharing.
- Medicine of publicly recognized and approved medical professionals, differentiated from the barbers/dentists/surgeons category, considered inferior crafts for their close physical contact with customers/patients and their poor training.

Third period (from the sixteenth to the eighteenth century Reformation, Renaissance and Enlightenment). Medicine of the visible inside the body; medicine as science and technique

Characteristics of the period

The third period progressed through a time of scientific knowledge expansion in all domains. The rational roots of the Enlightenment were found in all fields of science. Medicine was part of this overall trend, integrated into the spirit of the times. At first, the Renaissance aimed at resurrecting the values and reflections of the classical period, distancing the endeavour of knowing from the directives of faith, a separation to which the Reformation strongly contributed by bringing about alternative viable faith expressions not subordinated to the Church – in fact confronting the Church's monopoly on the truth about the world. The Enlightenment that subsequently unfolded brought the promise of reason and empowerment as the ultimate way of knowing the world in all its complexities.

This was a period of remarkable advances uncovering the body's structures and functions. In those three centuries, far more was added to medical knowledge than in the previous two millenniums. This was also a period when scientists made their discoveries individually, finding international prominence, setting in motion a virtuous cycle in scientific communities, encouraging new studies and investigations. There were intense and continuous exchanges between centres of studies in the main European cities (and later in America), with scientists often spending time studying or teaching in different countries.

By the seventeenth century, with the vast collection of studies of anatomy, physiology, biochemistry, and such, the systematic classification of disease had become a pressing issue. The concurrent progress in the classification of animals and botanical species offered an example of successful models. Ultimately, diseases could perhaps be classified as distinct metaphoric "species", so some tentatively thought.

The appearances of several nosologies were witness to the great efforts to come to terms with the huge varieties being uncovered. It was necessary to schematize and reduce complexities, "controlling" the increasingly diverse assortment of identified symptoms and signs into manageable classifications. Diseases were placed in the organs rather than on unbalanced humours as in the previous two millennia. Classifications thus were crucial for guiding the approaches to recognizing diseases and communicating about the category of disorders being dealt with as well as the respective treatments.

However, the curative resources did not advance as much as the knowledge of the body. Medicine as therapeutic practice used the same tools for treating the recognized illnesses (diet, bloodletting, laxatives, purging, enemas), although now enriched with new plants and herbs (such as quinine) from newly discovered continents. The radical evolution of treatment methods and resources had to wait for the major breakthroughs to happen in the next period.

At the end of the second period, the invention of printing had an immeasurable impact on the transmission of knowledge from which medicine reaped enormous benefits from this third period onwards. The mass reproduction of medical books and the widespread launching of university courses as standard for medical training contributed decisively to the irreversible consolidation.

Medical semantics were thus systematized and universalized, unifying the field into a worldwide single corpus of meanings. Across the world, medical students would learn from the same texts, written by the same renowned authors. Medicine thus consolidated social recognition as universal knowledge and a professional field.

This period therefore can be characterized as the consolidation of scientific_medicine on three pillars: the growth of university courses and academic medical studies; the mass production of medical books (now translated into European languages); and the expansion of hospital-based treatments with the establishment of hospitals for exclusive treatment of specific diseases.

Medical knowledge acquired more complex explanations about the functioning of the body and the alterations caused by diseases; this was

possible with the expansion of studies of anatomy, physiology, pathology, pharmacology, and so on (see more detail in the section on medical self-reference below). Among the contributors, the names of scientists such as Andreas Vesalius, Paracelsus, Thomas Sydenham, Marcello Malpighi, Herman Boerhaave, Giovanni Morgagni, William Harvey and many others can be highlighted.

With university courses, books and medical hospitals, the medicine of the period became "medicine of the visible inside the body", with a newly developed arsenal of observation tools and techniques, venturing beyond the surface where it had previously been confined. The use of microscopes; the invention of equipment and tools such as thermometers; the practice of dissection becoming routine and making it possible to complement diagnosis with observations after death; the establishment of routine procedures of ward visits by the head of the clinic followed by his students, are just a few examples of numerous developments strengthening the impact of the combination of those three factors: university courses, books and hospitals.

Medical guilds as forms of organization and protection of doctors and medical practices expanded and started functioning in several European countries. Regulations and formalization of processes of authorization for practising medicine brought about the final steps to make the field exclusive to qualified professional practice.

The epidemics of the time (such as the plague, from the fourteenth to seventeenth centuries) started to be addressed using scientific methods with better and more systematic understanding of epidemics and hygiene measures to control them, although still without the crucial knowledge of vectors, transmission and infection. Also on social application of medical knowledge, the first empirical attempts at what would later become mass vaccination took place by fortuitous experiment with self-inoculations by Edward Jenner with smallpox in eighteenth-century England, opening new horizons for the social impact of medicine and the constitution of public health.

System's features

Medicine's relevant systemic features acquired in this third period prepared the finalization of the health system by creating strong self-reference, dedicated social space for intensified exclusive communications, and self-identity for autonomous self-reproduction. More specifically, the important developments for final health system differentiation were:

1) Development of complex exclusive semantics to communicate on matters related to:

    - Understanding of the body's functions and structures, and disruptions provoked by diseases;
    - Links between evidence (signs and symptoms), the corresponding affected bodily structures and functions, and treatments;
    - Design of nosology (disease classifications) based on anatomical and physiological knowledge (while aetiology was not yet developed).

2) Establishment of university courses and social consolidation of the expertise as distinct and socially relevant, making possible the social selection of who was and who was not allowed to practise medicine, based on knowledge and capacity to communicate using a recognizable vocabulary and the possession of valid certificates. The communications thus reproduced represented a solid legitimizing standardized base. Where the conditions for those communications were set up, observing and adopting the same consolidated references, medicine sub-system (and later on health systems) was able to emerge. Medical courses made medical self-reproduction effective; doctors were teaching future doctors on matters that only doctors could communicate about.
3) Spreading over Europe, the organization of medical guilds as protectors of the craft, defining the limits and conditions for medical practice, constituted forms of association and self-regulation, pro-

gressively becoming instrumental for representation and self-identity in interactions with political systems.

4) Confirmation of designated social spaces (hospitals) for medical practices alone. The establishment of hospitals as entities of public interest, with political and financial support, had a profound systemic impact. Doctors would no longer work only as independent individuals at the bedside of those who could afford to pay them. The expertise was intended to benefit the whole society. Hospitals were exceptional sites for the increase in complexities, both of observed and treated diseases and of the exclusive communications by doctors. Hospitals concentrated in the same space a large number of sick people with diverse conditions, thus also facilitating the concentration of professionals. Doctors would have (as they still have) a lot to discuss among themselves about the patients. Daily discussions and education sessions with presentation of cases to large audiences of professionals and students became part of hospital routines. Communication-based social systems largely depend on the continuous intense reproduction of communications.
5) When the health system stabilized as a function system in the following period, it could assimilate internal differentiations in specialities and other professional fields that started to become accepted during the third period. The third period showed progressive incorporation of additional professional practices, which later would become fully differentiated specialities or health professions. That happened with the initial steps in the direction of recognizing surgeons, nurses, dentists, pharmacists, and so on, each with their communications and semantics, as valuable parts of healthcare service provision.

With all these levels of complexity, the embryonic social health system was establishing boundaries and becoming fully in charge of its own reproduction. The system's life was reliant on its self-observation, self-regulation and self-reproduction, all realized and sustained by the increase of communications among doctors.

However, to become a function system it was still necessary to develop external coupling with other differentiated systems. For that matter, the political and legal systems were being activated on matters of health as being of socio-political relevance. The health social system therefore began to emerge. The French Revolution was a landmark in that regard, with public policies issued regarding public health concerns, organizing health service coverage and entitlements of the population (see later). Although the conditions were set, the health system nevertheless only emerged in the subsequent period.

Complexities

To address the complexity issue for this third period we need to briefly recall some theoretical definitions presented in Chapter 1 and add a few more points. Complexity refers to more elements and relations than a system can handle; the system recognizes its inability to observe excessive numbers of elements and relations in its internal and external environment. So here we refer to complexity as reflecting two aspects: the elements and relations medicine in the third period recognized in the body and in diseases; and the composition of medical self-reference in the midst of massive advances in knowledge of the body, or in other words, what medicine would see as its tasks.

Complexity and causes

Up to the end of the period in the eighteenth century, medicine accumulated substantive knowledge about the internal structures and functions of the human body, and also identified transformations associated with diseases. The rational medical approach linked disease symptoms with pathological signs in affected anatomical structures. Descriptions of these structures, and descriptions of observable functioning and visible transformations in the organs, were rich and detailed.

However, basic causal relations remained undetermined or weakly suggested. Where a causal relation would eventually make the phenomenon comprehensively understood, all that was available then

offered only room for speculation. With the knowledge of cells, germs, genes, biochemistry, molecules, hormones, and so on still to be developed, the medical understanding of causality was limited to very simple explanatory models, such as that a single cause would produce all diseases, or that external factors such as the stars, the "miasmas", were the causes of diseases – not very far from the schemes the Greeks had already suggested with the four humours.

There were also some more sophisticated theories suggesting the balance of external stimulations and excitation of internal organs, as well as the tuning of vital energy flows inside the body with tension in the fibres. Furthermore, others indicated the causes of the diseases could be found in the organs and their tissues.

Medicine today has a wide range of causal relations linking pathological signs and symptoms to alterations in organs, tissues, cells, physiological functions, and so on, which can be authoritatively clarified using a large arsenal of tests and examinations. That was not the case then. The causal links between the visible complaints/alterations, and the deeper structures and functions involved were not understood. Crucial measurements and detection instruments had not yet been developed, leaving large gaps in presumed causal links. For example, there were no ways of finding out that the conditions seen in advanced diabetic patients were caused by the increase of sugar in the blood, which would not be recognized.

Multiple-causation was an even more distant consideration, due to the lack of methods and instruments for separating the effects of different factors. In addition, the natural variability of all biological phenomena could not be understood as possible expressions of the same disease, being therefore wrongly interpreted as distinct.

It can be argued that the knowledge of causality has two effects. First, it can reduce complexity, filling gaps and reducing the known unknowns as far as relations are concerned, as it identifies reasons for linking otherwise apparently disconnected observations. Secondly, acknowledging potential causal relations may bring about complex considerations of different factors and different possible connections for the same observed effects. Thus complexity is not ruled out once

plausible causes are considered; it rather acquires a new configuration in correspondence to the variety of causal explanatory models available in the context. Causal links can indeed be highly complex to determine and reconcile with alternative explanatory models.

In the third period, though, we see an open field with multiple possible questions about causes of which there were hardly any satisfactory answers. Explanatory models could not be developed, as many causal factors had not yet been found. In this sense medicine in the third period was still describing diseases, although with a richness of detail that had not been seen before. Each organ was being studied individually and those studies included morphological alterations observed in autopsies, without revealing causal processes.

Self-reference and complexities

Another way of approaching the issue of complexity in the third period is related to the complex articulation of self-references. During this third period, the self-reference of medicine had to incorporate new meanings arising from the knowledge being produced.

The new findings had implication for the way medicine could define itself; the doctors could for instance say that medicine was no longer about the four humours and instead it was about the lesions anatomically described with possible alterations in the respective functions. The new discoveries thus changed the descriptions doctors would make about their knowledge and consequently their job. From the point of view of the time, these implications were rather complex.

In a broad sense, self-reference has to handle the complexities of its own nature with the evolving meanings. We can say that a function system based on communications and meanings is restless in its continuous efforts to maintain its self-reference. It cannot be otherwise. Once a system is set up to refer to itself according to a certain selection of meanings, it creates the need to reaffirm the selections at every instance.

Complexity arises out of the very existence of a multiplicity of possible selections that are constitutive of meanings (Luhmann, 2023).[4] Complexity cannot be avoided or eradicated in face of the multiplicity of meanings available. But it can be temporarily "tamed", so to speak, with meaning-construction operations by which only certain meanings are selected and the other possible ones dismissed. That is how meanings are generated and communications can communicate them.

Complexity is therefore temporarily sent into the background, still nevertheless offering the contrast that makes visible the selected meaning. As there is no self-reference without meanings and selection of meanings, complexity is the very basis on which self-reference is built. This theoretical structure should help us to understand this aspect of the progress of medicine in the third period.

In the history of medicine there are several examples of very important breakthroughs that had nevertheless to wait years before being recognized. These are examples of how obsolete self-references are maintained sometimes even at the sacrifice of relevant innovations that supposedly could bring about severe destabilization of the whole conceptual building thus far assembled. An example of that was Edward Jenner's experiments with self-inoculation of smallpox; the Royal Society declined to publish his paper describing the procedure (Bynum, 2008, p. 74).

The complexity of the self-reference of medicine encompasses all the scientific and technical advances that were happening. Each advance unfolds into questions of what medicine is and how it should be seen. We can consider, for example, that medicine was reformulating its all-inclusive identity, assimilating the advances in scientific fields such as anatomy, physiology, pathology, and the notions of pharmacology, infection, histology and others. The complexity resided in carrying on with the subdivision in different areas while keeping the sense of integration and articulation between all the distinct divisions. The internal divisions

4 In the book *The Making of Meaning*, particularly in the second chapter "How is social order possible?", Luhmann (2023) extensively develops the conceptual architecture of the interconnectedness of the concepts of complexity and meaning.

could in fact enrich the sense of medicine as a unit, as a condition and support for the recognition of increasing internal variety. Paradoxically, the internal diversity did not destroy the sense of unity, but rather strengthened it. We can say the strengthening effect arise from the fact that the new specialized knowledge kept the same fundamental binary distinction (healthy/sick) in place.[5]

Besides that, the description of pathological phenomena, particularly anatomical and physiological features of pathologies, provoked intense reflections on the meanings of the normal and the pathological and where one started and the other ended. The topic was at the heart of the self-definition of medicine as exclusively concerned with diseases (the pathological). Apparently, the solution to the normality problem would allow deciding when an individual would become a patient, and therefore an object for treatment and medical studies.

Although the formal notions of normality and pathology seem to dress the topic in a simplistic outlook, in fact the location of the boundaries and characterization of the two opposite sides and how it was possible to move from one side to the other of the distinction was highly complex. Whether pathology was only an exaggeration of normal physiology, or normal physiology with different intensities, or whether they were totally separate biological conditions, are examples of the kind of questions at the heart of the considerations of "so, actually, what is medicine?" Is normal physiology also something of interest for medicine? Why? (Canguilhem, 1978).

---

5 In Luhmann's terms (2007, p. 43), each partial system (and we may see a medical speciality as a sub-system of medicine and a part of the health system) replicates, co-realizes and reinforces the overall system (the health system) through its own specific system/environment difference (being the system reproduced as speciality or sub-speciality, with the environment being the field of concerns of the speciality in the environment of the system – specific functions and organs in the human body). In other words, medical specialities reinforce the sense of unity of medicine because any speciality reproduces the same binary distinction healthy/sick prerogative of the medical sub-system and is validated as such.

This topic will be treated more extensively in the section on Canguilhem in Chapter 5. For our current discussion our message is that in the third period medicine was asking about its own identity in complex ways. For those potentially disturbing polemics about the normal and the pathological, medicine had already conquered its internal space, and it could afford to discuss its identity at length. Medicine could work out the means of its reproduction within medical schools and fields of practice, even while confronting fundamental questions about itself.

Medicine was increasingly becoming medicine's own business and no one else's, and could face the self-criticism and the internal doubts it was itself creating. The next section gives a more detailed picture of the diversity involved.

Summary of medicine self-reference: third period

Considering the complexity achieved in the third period, it is relevant to add another brief aspect of the theory of self-reference. In Luhmann's theory, self-reference is second-order observation, as it consists in observation of an observer by an observer – in this case, the observer itself. We recall that social systems are self-referential autopoietic systems. Given the complexities of the knowledge gathered in the third period, self-reference was set in continuous and self-reinforcing motion with proliferation of sites where doctors could observe other doctors and through communication select and generalize what all doctors were supposed to know and do. The reproduction of medical meanings required self-observation and communications in the context of the self-reference. This became increasingly systematic as the medical training expanded and covered many years of professional life. We list below the points of self-reference that were established during the third period:

- Medicine's self-referential entitlement to look inside the body: Medicine could recognize itself as in charge of accompanying the patients over the course of diseases, from the bedside to the morgue, establishing clinical-pathological correlation and gathering evidence

of the presumed pathological processes. Autopsies became medicine's business.

- Medicine of the recognized scientist doctors, researchers of an advanced science, with command of scientific methods, observation tools, experiments, working in academic institutions, developing new models of the body and internal functions, expanding knowledge in several directions. Among some of the new fields of studies and most prominent scientists were:
  1) Medicine of anatomy (Andreas Vesalius, sixteenth century, established anatomy and physiology as independent disciplines, developing complete anatomical descriptions still used today);
  2) Medicine of physiology (William Harvey, seventeenth century, described the circulatory function);
  3) Medicine of pathology, physiopathology and pathologic anatomy; (Antonio Benivieni, fifteenth century, verified pathologic evidence with autopsies);
  4) Medicine of contagion and infection (Gerolamo Fracastoro, sixteenth century, made the first proposition of the concept of contagion by invisible particles entering the body after contact with a sick person or contaminated material; Ambroise Paré, sixteenth century, proposed a number of inventions related to antisepsis for surgical procedures);
  5) Medicine of chemistry, toxicology and pharmacology (Paracelsus, sixteenth century, developed proposition of dynamic views of medicine based on chemical elements as vital force of the organism, specifically mercury, sulphur and sodium);
  6) Medicine of microscopy and medicine of histology (Marcello Malpighi, seventeenth century, performed initial studies of microscopic anatomy);
  7) Clinical medicine, linking clinical observation of signs and symptoms and anatomic lesions at the level of tissues during the seventeenth and eighteenth centuries, including: development of nosology (Thomas Sydenham); classification of diseases with expression of symptoms in individuals (Boissier, Cullen); clinical observation of anatomical lesions (Morgag-

ni); histopathology and diseases in the tissues of the organs (Bichat); treatment horizons and the notion of clinical trials (controversies between Pierre-Charles-Alexandre Louis and François Broussais over chances of curing diseases, particularly bloodletting, using rudimentary statistics);

8) Medicine of instruments (Laennec, eighteenth century, invention of the stethoscope, establishing auscultation as clinical method, and the thermometer).

- Medicine self-assessment of its own references: The period saw medicine becoming capable of judgements of what medicine should or should not be concerned with, in a self-coordinated all-encompassing fashion. Doctors became interested in knowing how accurate the identification of symptoms and corresponding diagnosis hypotheses were, by confirming them (or not) through post-mortem examination. Medicine needed to confirm its understandings for itself. For example, self-assessments led to acknowledging that the theory of humours was not enough to explain what was being uncovered in the bodies; better explanatory models were necessary. In this process, medicine was thus observing itself formulating ways of observing; it was observing how it was observing. The new findings brought to light had to be tackled in more complex ways. Paracelsus strongly dismissed Galen's and Avicenna's theories previously admitted as unquestionable; this was a clear example of self-assessment and self-reference dynamics in action.
- Medicine and the semantics to communicate about the body and diseases became admittedly highly complex, not only through the increase in descriptions of organs and diseases affecting them, but also because of the above-noted self-assessments, by which the semantics required preservation and strict rules of presentation and argumentation, subject to continuous reviews. It was not enough to use the correct words but also the conditions and criteria of validity of their use were required in academic settings (this self-reference was new in the history of medicine).
- Medicine of the universities: Medicine could not be practised without formal training in universities or academies of science. Legally

regulated as a profession, self-regulated in its contents, and concerned with the increasing complexity of its field, medicine needed to rely on elected mechanisms to self determine what was or was not valid knowledge and practice to be taught in the courses, and subsequently be verified throughout the professional life.

- Medicine of committee, councils, royal academies and collegiums: While universities were appropriate sites for educational purposes, medicine had to establish itself in institutions for permanent monitoring and preservation of the self-reference beyond the training period. This self-regulatory aim was progressively accomplished with the institutionalization of regulatory bodies. An example is the Committee "des Secours public" during the French Revolution (Foucault, 2003), concerned with public assistance, sanitation of hospitals and prisons, health policies and public hygiene. Another example is the "Regia Sociedad de Medicina y outras Ciencias" in Seville (seventeenth century) (Barona-Vilar, 2023).
- Medicine of complex hospitals: Hospitals became important for preparing the emergence of the health social systems. They provided unique sites for complex self-reference related to both observed and treated diseases and the semantics doctors were using and further developing for their exclusive communications. By concentrating large numbers of sick people with diverse conditions, the communication among professionals increased to levels it would never have reached if the patients were geographically dispersed and doctors had to travel to see them individually in their homes. The number of patients seen in a day became significantly larger along with the number of professionals able to check and discuss the same patients. Doctors met in wards, corridors and examination rooms, and talked about their patients. They met in daily formal education sessions, for the presentation of cases and discussions with large audiences of colleagues and students. Hospitals became crucial for the development of health as a social system, not because of the structures, resources or investment, but most importantly because of the exponential increase in communications among professionals they made possible.

- Medicine of the social, hygiene and epidemiology: At the initial steps of approaches such as prevention, hygiene, water and sanitation, politics of health, vaccination, and so on, these advances were crucial in broadening the concept of medical science as responsible for looking not only at individual patients but also at the health of collectivities. Johann Peter Frank's (eighteenth-century) "System of a Global Medical Policy" is an example of this progress. Edward Jenner (also eighteenth century), although still without the germ theory, tested inoculation for smallpox, opening the door to vaccination and immunization. Public health then began to set foundations for focusing on the health of populations, on the collective risk factors and on social distribution of healthcare services provision.
- Surgery was incorporated into medicine: At early stages in the period, surgery became recognized as knowledge and skills akin to medicine, in contrast with the previous understandings classifying barbers and surgeons in the same professional category. Barbers and surgeons shared the same status over the course of centuries until the third period. With the advances in anatomy and knowledge of the internal structures of the body, surgery was then upgraded to the category of medical practice, leading to important developments of anatomy for surgery and surgical procedures. The establishment, for instance, of the Colleges of Surgeons in Paris (eighteenth century), London (1800) and Cadiz (eighteenth century) marked the change. Such advances brought surgery definitely into the area of medical self-reference.

Fourth period (from the nineteenth to the twenty-first century). Medicine of the invisible made visible, medicine of specialities and medicine of the health systems

Characteristics of the period

The advances in medicine in this last period were extraordinary. The understanding of diseases' pathological processes, causes, curative and preventive interventions, were astonishing. Listing the progress would require an encyclopaedic approach, which is beyond the scope of this book. We therefore highlight here the most consequential advances and discuss their links with the final establishment of health as a social system. In the last section (summary of the self-reference of medicine in the fourth period) there is an attempt to group distinct categories of advances.

Parallel to the continuous progress there has been continuous rethinking about the nature of the diseases, as new sorts of causalities were uncovered in association with large diversity of factors. Clinical trial methods, tests and analysis, with sophisticated measurement devices and powerful statistical and processing techniques, entirely changed empirical approaches. Medical studies and research became large-scale endeavours involving much expertise and resources. The equipment and pharmaceutical industry experienced exponential growth at the forefront of innovation with new advanced products, continuously changing the scenery of therapeutics and availability of treatments.

In the previous period, it was already clear that there was no longer a single cause of all diseases but a multitude of factors contributing at different levels. For instance, the relations between cells and diseases were progressively understood with many advances in microscopic histology in the nineteenth century, considering the variety of cells in the body and cells of external organisms entering the body. With Pasteur's germ theory (nineteenth century) a crucial door was opened for medicine to recognize external causes of the diseases and related physical reactions.

Medicine thus became able to see itself as not concerned with just one or a few explanatory models but instead opened up to a diversity of

elements in complex causal arrangements, contributing to the recognition of disease in an equally complex array of expressions, now revealed in intricate macro and micro interactions. Medicine also opened up to accept any breakthroughs that shook and changed previous assumptions. Breakthroughs became welcome; they were no longer seen as threats; trust of medicine firms' means of verification and validation encouraged openness for innovation.

Health as a function system was fully established in the fourth period. In the current stage of health as a social system it is possible to pinpoint a number of characteristics:

- Huge diversity of diagnostics tools and curative interventions;
- Long periods for training and developing professional competences;
- Differentiation of many specialities (not only in the medical field);
- Complex clinical laboratories;
- Complex development, manufacturing and administration of drugs;
- Complex use of technological tools and resources for diagnostics, treatment and post-treatment monitoring;
- Complex surgical procedures involving several specialists and complex equipment;
- Specificity of the semantics within each speciality;
- Wide-ranging advances in prevention strategies;
- Medicine of teams, covering distinct areas of expertise with specialized professionals working together for the same patients;
- Medicalization of life (from cradle to grave) with diverse range of age-specific preventive and curative interventions;
- Development of public health perspectives and multi-sector interventions addressing a wide range of health determinants;
- Projection of health as a global/planetary phenomenon requiring coordinated action among nations;
- Huge increases in costs and absorption of financial resources in healthcare provision, changing the general pattern from lots of needs/little to be done, characteristic of the preceding periods, to lots of needs/lots to be done/few resources, in correspondence

to what is known and what is actually possible to deliver given generalized resources constraints;
- Uneven distribution of the benefits of medical advances across countries and populations within countries.

System's features

Medicine of the health systems, as we may call the medicine that emerged in the twentieth century, has a comprehensive command of the complexities of the body, the expressions, causes and treatments of diseases. Medicine now exists within health systems that also acquired strong command of the complexities of a plethora of social organizations providing preventive and curative healthcare. The whole assemblage operates as a *social function system*.

We defend the thesis that in this fourth period medicine eventually lent its complexities and closure for the constitution of health as social systems. Obviously "lent" is a metaphorical expression. What we actually mean is that once medicine became exclusively medicine's own business, with its complex specific semantics and universal communication standards, a distinctive social function could be created and differentiated from all others.

With its self-reference established within that social space, medicine could then recognize and interact with several disciplines sharing health concerns, such as dentistry, pharmacology, nursing, physiotherapy, psychotherapy, bioengineering, nutritional science, which, in one way or another, adopted the same fundamental healthy/sick binary distinction. This large ensemble of knowledge and practices progressively developed a comprehensive all-inclusive self-reference as health social system, with wide concerns beyond the strict medical focus on diagnosing and treating individuals' diseases. We briefly recap this development in this section.

Medicine completed its emergence as a specific semantic universe of communications, offering solid anchorage for other disciplines dealing with health. The social recognition of medicine could thus be extended

to the disciplines with which medicine would share the same concerns, each covering distinct areas and complexities (such as dentistry) with which medicine would not be directly involved.

The disciplines dealing with specific health topics developed their own communications, but still remained within the boundaries of the all-inclusive health domain comprehended by the healthy/sick binary code. The progress of medicine was fundamental for that; without it the assemblage of related disciplines could not have evolved and accommodated the array of complex and diverse services that are now available in any health system.

Once operations and communications between doctors and other health professionals became a routine part of the processes of diagnosing and treating, the comprehensive structure (the health system) needed to be correspondingly acknowledged; recognized by itself and its components, as well as recognized by the society's systems at large.

Luhmann's (1998) theory claims that the constitution of a differentiated function system within a society structured according to functional differentiation is based on three systemic orientations: (1) in relation to the society (we add, where health is socially recognized as a distinct function); (2) in relation to other systems in the society for provision of inputs and outputs; and (3) in relation to its self-reflection as specific system.

The societies of the previous period identified medicine as having the purpose of knowing the body and doing something about the diseases. It existed mainly as part of an undeveloped science system, studying the body while providing commonly known treatments of doubtful efficacy. The prominent doctors in the history of medicine in those periods were scientists observing physical ailments. On the service provision side, in spite of the crowded hospitals of Paris in the eighteenth century for instance, medicine still had very little to offer, no different from in the previous periods.

However, the nineteenth century changed everything. Medicine could deploy a number of new therapeutics and therefore medicine's practices were recognized as having a broader scope beyond scientific research endeavours. The fight against charlatans begun in the eighteenth century was the affirmation that qualified doctors could pro-

vide a differentiated service, and that difference became increasingly marked as new therapeutic techniques were progressively developed and adopted, out of reach of charlatans. Service provision became distinctly marked and valued.

To the self-reference of medicine as scientific discipline, developing knowledge of the body and diseases, new competences expanded the social roles for providing healthcare, endowing medicine with a new outlook. The French Revolution brought the social role of healthcare provision to the fore. Although still lacking substantive therapeutic procedures at that time, the broader prospect and benefits of healthcare being incorporated into the politically enacted socio-policies became well understood and desirable. The notion of public health was born. Self-reflection for the health function system expressed not only attention to diseases but also the social distribution of risks of becoming sick and getting (or not) treatment.

The political mobilization of healthcare took place within the strengthened and legitimized space/function in line with the concerns of the political and legal systems. To deal with the new social configurations, meanings had to be created explaining what the new structures were intended to be.[6] The political relevance of having them in place was clear. This is how public health became a reality, first as a set of institutional meanings, politically established, legally recognized, to then operate as a field of communications, according to its own self-reference and self-reproduction.

The political system would not go beyond recognition of the health system, including medicine and public health, and would not run medical or public health businesses. These were exclusive competences of, respectively, medicine and public health. Health as a social system was thus established, with health becoming a function system within societies already structured with differentiated functional systems, such as the political, the legal, the scientific, the educational, the economic.

---

6 See discussion on the changes brought about by the French Revolution in the section on Foucault in Chapter 5.

Within the health social system, public health was the sub-system concerned with the health of populations. Public health was endowed with the corresponding mission to develop and explain the health system to the health system, defining its purpose and reason to exist, reflecting on health as a social phenomenon for which collective representation was needed, not limited to the medical focus on the individual patients. Public health thus became the way of explaining the health system to the political and legal systems, as medicine's language became too specific, within its exclusive focus on individual bodies, for communications outside the health system.

Medicine was not oblivious to the health of the collectives[7] and understood that diseases had common causes and occurred concurrently in many individuals. However, treatments were delivered to individual bodies. On the other hand, public health thinking aimed at what could be done at a social level; the social thus acquired prominence in public health semantics in contrast with its place in medical semantics. The two branches came to constitute the health system inclusive of all the other coexisting professional fields encompassing recognized health practices.

We can say that in this process medicine was strengthened. Within health as a social system, the medical nucleus obtained a recognized secure central position, in charge of identifying and treating diseases, and validating all meanings and understandings related to that. Medical knowledge could be debated, questioned, revisited, revised, modified, reworked and so on, with its achievements preserved.

Despite the fact that, in the very beginning, health systems were strongly reliant on decisions of the political and legal systems, progressively they became "sealed off" from external interference, though internally open for exclusive communications. It is good to remember that when we talk about closure we are talking about meanings, only about meanings. Closure is about communications and understanding

7 Epidemiology was accepted from its onset as part of social medicine, becoming recognized as a modern scientific health discipline with the works of John Snow on the cholera epidemic in nineteenth-century London.

of distinct meanings, within the limits the system considers meaningful for it. In this sense, "closure" is a technical term, a central concept in Luhmann's theory. Every time we use this term it carries the above-explained meanings in correspondence to the Social Systems Theory.

Towards the end of the nineteenth century, the health system had achieved its closure, becoming a function system differentiated from other function systems in the society, with its unquestioned exclusivity over the use of the medical semantics and validated health communications. The semantic territory was precisely marked.

The closure also implied self-regulation, as no other system or professionals could issue valid medical communications or judge the quality or validity of a medical communication. It also became valid for all non-medical professional fields operating within the health system.

To sum up our understanding of system's features in the fourth period, the health system, as the broader semantic space with medicine at its core, developed its self-reference. As a function system, the health system's self-reference narratives were constructed, portraying operations and achievements of the system as a whole.

Besides being the comprehensive domain of all that concerns medical practice, the self-image of a health system also incorporated public health indicators selected for that purpose. The narratives appear in plans, evaluations, policies, investment programmes, as well as epidemiological and demographic measurements, targets, reports and so on (Chapter 7 has an in-depth discussion of indicators).

The narratives reveal a health system addressing increasing complexities. The tension of the trade-offs between providing for the individuals and at the same time maximizing collective access to healthcare is an example of those complexities. The deepening of the distinct roles of the two subsystems (medicine and public health), expanding knowledge and practices in their respective semantic domains, with many internal semantic ramifications, is also a snapshot of the expansion of the system.

The tensions related to the ever-increasingly complex range of technologies and technical resources for diagnosing and treating diseases, and delivering them according to the model of comprehensive perma-

nent mandate by which everyone should be entitled to everything they need (universal health coverage) is in fact overwhelming.

Another feature of the health system's complexities is that it still requires individual doctors, although doctors now rarely work alone. More often they work in teams and require communications among specialists. They also rarely work without laboratory and pharmaceutical support, which have become an integral part of medical care. Consultations, although still seen as a relationship between the patient and their doctor, now requires a coordinated joint performance involving all the information gathered from laboratory tests and examinations as well as expert opinions from a diverse range of specialists. Routinely, in any medical setting, doctors consult their peers for matters within their specialities. This has become health system medicine. As the complexities suggest, healthcare is a multifaceted endeavour and only a system can deal with and deliver it.

While previously, in the former evolutionary stages, the uncertainties the individual doctors had to deal with were their own exclusive concern, now the concern belongs to the system. The system has to make sure that all doubts are dealt with in a systemic way; commissions need to be formed and informed, possible courses of action considered, possible evidence gathered before the system can make a decision, which will be the combined responsibility of the system as well as of the individual doctors.

Complexities

The discussions of systemic features of the period in the previous section have already portrayed the complexities of the fourth period. In this section therefore we have only a few more points to add.

As was already the case at the beginning of the fourth period, the full understanding of medical communications became virtually impossible for outsiders, and to a certain degree even for individual insiders. Hugely complex vocabulary denoting signs, symptoms, diseases, causes, conditions, treatments, and so on became a complex universe beyond the

competence of any individual professional, requiring multiple areas of expertise and connectivity between the semantics of several fields.

We may also say that in this last period the internal complexity of the system has become a crucial feature of the health system itself. While the complexity of the human body, uncovered in a series of successes in the descriptions of diseases and evidence of treatments (with knowledge by which many previously untreatable conditions are now regularly cured), the social organization of provision of healthcare to all who could benefit from the available knowledge was and still is highly problematic. The complexity of service provision thus affected the world of healthcare delivery in a system-wide fashion. Thus we can say that the complexity has now appeared inside the health system.

Obviously health systems' objectives have not been to cure all diseases in all individuals all the time. Human beings have to die at some point, and even in the best treatments, death is and will remain the unavoidable end point. But it has become clear that there are deaths that could have been avoided if the provision of care (including preventive care) was timely and sufficient. These problems mostly exist not because of the complexities and unknown facts about the body and the diseases but rather because the health system has not designed and set up arrangements and structures that could make the avoidance of those deaths possible.

As an example, if somebody dies in a road accident, that cannot be simply dismissed as an unavoidable death (as far as the health system is concerned), because if the person dies while being removed from the accident site in an ambulance not properly equipped with life-saving devices, that was then an avoidable death. What we call avoidable deaths are therefore those occurring because the health system was not optimally prepared to provide the curative or preventive services required (when the life-saving procedures are already known).

We may state that the complexity has become greater inside the health system than in the environment the system has to deal with. The system can show failures because of insufficient qualified professionals, lack of equipment and drugs, lack of organization, lack of correspon-

dence between its design and the needs, lack of resources, information, communications, and so on.

The complexities of the system emerge from the system's self-image as having to provide answers to all health needs, and the resulting dilemmas related to how to crack unsolvable problems of incompatibility of simultaneous solutions of equity, efficiency, responsiveness, efficacy, quality of care, and so on, in the present and the future. The system is therefore submerged in its own complexities, having at the same time to understand its limited horizons and reduce the complexities as feasible, often having to reject the optimal for the benefit of the possible. But the decisions to be taken are far from clear and simple.

Summary of medicine self-reference: fourth period

Self-reference of medicine in the fourth period is an integral part of the self-reference of the health social system. Medicine is identified and self-identifies as being a core element of health systems. It operates in conjunction with other elements comprising the system, making the health system an ensemble communicating with meanings referenced to sickness and treatments.

Nevertheless, during the almost two and a half centuries of the fourth period, medicine developed specific self-identifications, which progressively reflected on and were incorporated into the health system self-reference. These specific self-identity points are explained next.

- Medicine of the invisible – by invisible we mean the minuscule concrete as well as abstract objects or explanatory concepts. This includes cells (the works of Schleiden, Schwann and Virchow), germs (Pasteur, Koch and others, with the establishment of bacteriology as a medical discipline), genes and chromosomes (establishing medical genetics), molecules, the mind (psychiatry, psychoanalysis, psychotherapies), and imagery, radiation, statistical significance. Medicine of the invisible is also medicine of powerful tools such as drugs, radiation, biochemical reactions, equipment, laboratories, artificial intelligence, with new fields of observations, diagnosis and

therapeutics. As part of the medicine of the invisible, biochemistry, the action of drugs, and the close relation between pharmacology and physiology, became a universe of advances in itself. Each of the listed areas of advances has had huge impact in many medical fields; a very simple illustration is the knowledge of germs and the crucial innovations in asepsis and aseptic interventions.

- Medicine of specialities and sub-specialities: With multiple fields developing into areas of expertise with their own specific models and semantics, internal differentiation and categorization of medical fields has been a necessary complexity reduction strategy. Specialists can handle the complexities of their respective fields better than generalists. If kept undifferentiated, the joint complexity of multiple specialist fields becomes overwhelming. The potential for continuous increase and breaking down of new specialities exists, with narrower fields for in-depth observation revealing detailed components of the human body connected to specific diseases.
- Medicine of specialist teams: Compensating for the fragmentation into specialities, the medicine of the end of the fourth period has recognized itself as medicine of multidisciplinary teams working together. The teams bring diverse perspectives together, approaching singular cases with observations bridging specialists' fields. Often a specialist has to rely on teams of experts revising the literature to interpret findings not specifically within their speciality.
- Medicine of "information overdose": The spectacular increase in availability of medical studies through specialized publications (printed as well as online) corresponds to a huge explosion. Such explosion of information does require teams dedicated to searching through the available materials on any topic, making literature review a routine exercise in the practice of advanced medicine. Medical information has become popularized with easy access to publications online; the speed and volume of new works continuously published represents a big challenge to medical professionals willing to remain up to date on their specialities. On the other hand, the gap between non-professionals and professionals has become greater in spite of the ease of access to literature. The terminology used in

medical texts is highly technical and cannot entertain widespread understanding. The frequently defended idea that now patients are or can be as informed as doctors does not take proper account of the expertise required for correct understanding of what is written in the now accessible medical publications.

- Medicine of the collective:[8] The medicine of risk factors, numbers, statistical significance, hygiene, epidemiology, screening, surveys, surveillance, prevention, vaccination, clinical trials, environmental risks, sanitation, social determinants, and so on has its domains comprehensively included within the field of public health. With the notion of social factors, public health brought into its core the unfathomable complexities of more than two centuries of social science theories, with variables and research opening up countless avenues for analysing social determinants of diseases, treatments and deaths. Chadwick and Virchow (nineteenth century) were pioneers in linking poverty and diseases, and vice versa. Previously, although in ancient times the health of soldiers was a matter of concern, and prevention was imposed particularly during the Black Death pandemic in the fourteenth century, preventive measures became a serious technical matter requiring permanent attention in the third period. Thus, we can say that the orientation towards the collective and the social has been, if not latent, incipiently present in medicine since its start. But the advances of the last period in terms of the understanding of causes and methods for monitoring and tackling epidemics brought about major transformations, consolidating public health as crucial for ensuring the health of populations. In line with that, the understanding of the social distribution of health services in correspondence to the social distribution of risks of becoming ill and dying has relevance for the coupling of health systems with political systems. As noted in the previous sections, although originally indistinguishable from each other, public health became a

8 We use the term "medicine of the collective" in line with the orientation we have adopted in this book in correspondence to the emergence of public health as a distinct sub-system of the health system.

realm distinct from medicine, with both public health and medicine becoming sub-systems of the health systems.

- Global medicine: With the rise of international agencies focusing on health and the coordination among multiple international players, medicine and public health are now projected on a planetary scale. Heightened attention has brought diverse international players together to address regionally endemic diseases as well as epidemics and pandemics and access to quality healthcare. The global distribution of the burden of diseases with inventories of factors and consequences of global reach now more than ever includes the environmental impact of human activities.
- System's medicine: Self-reference, self-observation and self-reproduction are key functionalities of medicine and public health as core components of health systems. With increasingly demanding and complex orientations to prevent and treat all diseases, and to tackle all determinants of health, health systems are expected to demonstrate how they are going to go about delivering such ambitious aims. Health systems have fundamental attributes of self-reproduction carried out essentially through self-reference and self-observation; a health system, as constructor of its own references as a social function system, has to decide what it will continue doing, and what it will start anew. All decisions are observed and judged by internal as well as external observers. It is very hard, basically impossible, to satisfy everyone. But the health system, as a social system, above all needs to preserve its reproduction with the means it produces – that is, with its exclusive communications, which only the health system can reproduce.

Chapter 5: Canguilhem and Foucault

The works of Michel Foucault and Georges Canguilhem add a wealth of contributions to the history of medicine, particularly social history and the evolution of medical ideas. In this chapter we do not provide a comprehensive summary of their work, nor do we discuss their views in detail, but they studied what we consider a crucial period in the history of medicine, where we can see the beginning of health as a social function system.

Therefore we take from their works the elements that in our understanding reveal the evolution of health systems from an embryonic to a functional stage. We set out our reading of these authors with conceptual references of the social system theory.

We also benefit from the abundance of historical materials they worked on, with vivid descriptions of the medical thinking of the times. Both addressed the late eighteenth and early nineteenth centuries: Canguilhem, reflecting the intense discussions about notions of the normal and the pathological, and Foucault with a broader reflection, addressing the marked changes in the way medicine approached its subject in correspondence to the transformations in the political and social context of the French Revolution.

From our point of view, the period revealed medicine in the process of expanding its self-references, struggling and redefining its subject in the face of important advances in the understanding of the structures and functions of the human body. At the same time, public health appeared at a confluence of political intentions with medical knowledge and practices, now sufficiently stabilized and complex. In this chapter,

we aim at making these processes visible from the material the two authors explored.

Both authors describe medicine "at pains" (our term), attempting to describe the "nature" of the diseases, which at the same time reveal and hide themselves from observation. Medicine was already able to talk with clarity about structures and functions of the body but was trying to figure out the possible existence of "norms" of how the diseases occurred in the bodies; "norms" that could make intelligible the courses and symptoms of diseases.

In a broader sense, the period shows exciting discoveries and discussions through which medicine was both constituting its new identity as a modern scientific endeavour as well as establishing itself as a society-wide provider of health services with political recognition. We start the chapter with Foucault, then move on to Canguilhem (as opposed to the chronologic order of their work), and then discuss the thesis we defend here of the constitution of health as a social system at the convergence of medicine and public health. The chapter ends with a conclusion.

Foucault

Foucault's book, *The Birth of the Clinic*, published in 1963 (Foucault 2003), was his third book. It reveals Foucault at the early stages of his career when he had not yet addressed the topics of power and history of sexuality, which made his social theory famous. The book is part of a collection presented as *archaeology of the knowledge* from his historical research. The subtitle of the book is *The Archaeology of Medical Perception*.

We focus on three recurrent themes appearing throughout the book: the *Medicine of Species*, the *Gaze* and the emergence of public health in the midst of the French Revolution. Foucault's highly poetical style makes the first two topics challenging. This is due to his excessive use of rhetoric metonyms and metaphors, and imprecise concepts. The third topic, however, reflects the fruits of his historical research; it is more factual and less conjectural.

Medicine of species

Foucault uses the word "species" frequently. He quotes authors referring to diseases as "species". We understand that the metaphorical use of the term "species" in reference to diseases cannot be adequately likened to its denoted original biological meaning. Those uses of the term expressed only an intent to group diseases in some way, rather than revealing any distinct essential characteristic of medical thinking comprehending a disease as an actual "specie" of some kind. We in fact see only a metaphoric expression of classificatory efforts, not the discovery of factual grouping of diseases.

But Foucault deduces that the likening of diseases to species had a deeper effect in the structure of medical thoughts of the time. The move to identify species in biological science could perhaps be transposed into medicine, and medicine likewise could then look for the phenomena compared to the biological models comprehending growth, reproduction and death, as plants and animals do. That was a stretch too far.

The terms "medicine of species" appears in 11 sentences. From our reading of Foucault, what is said about "medicine of species" in fact reveals in poetical language the intention to classify diseases and construct a nosology. Basically, with these terms Foucault talks about the nosology being developed by important figures such as Gilibert, Sauvages, Pinel and others.

Foucault seems to be fascinated by the likening of diseases with "species in a garden". He tries to delve into the classificatory mindsets of the time, following their struggles to link, make analogies, associations, elucidations, and so on, what was then known about the disease, devising a logical grouping that could prove reliable and helpful. To fully reject his emphatic use of the term "species" as well as his colourful descriptions of how thinking about species has influenced the thinking about diseases, one would need to dive again into the texts of the time and their meanings – a daunting, time-consuming task.

For us, however, the metaphoric use of the term "species", although being one among several possible inspirations for classification of dis-

eases, in fact obscures the differences between conceptualization of natural science (classifying animals and plants into orders according to stable morphological and functional criteria) and its possible use in medicine.

Certainly the idea of "species" had some appeal for the elegance and simplicity of the biological rationality. We can only conjecture whether the medical authors of the time understood the limitations of approaching diseases with species in mind. We believe they may have done so just because it would seem an obvious scientific approach at that time; in any case, the notion of medicine of species did not find its way into the history of medicine, and became irrelevant.

Foucault also talks about other types of thinking, which did not include "species", such as when medicine assumes a "statistical structure" (Foucault, 2003, p. 125), and its perceptual field becomes a "domain of events", no longer a "garden of species". And he acknowledges that "it is no longer a pathological species inserting itself into the body [...] it is the body itself that has become ill" (p. 167). Anatomical locations, "organic spatiality of the body" (p. 224), connections, proximities, were principles of nosological analysis used even previously in the seventeenth century. "Species", so to speak, was not the only or the main game in town.

Foucault says:

> „Until the end of the eighteenth century the gaze of the nosographers was a gardener's gaze; one had to recognize the specific essence in the variety of appearances. At the beginning of the nineteenth century another model emerged: that of the chemical operation – "*Instead of following the example of the botanists, should not the nosologists have, rather, taken as their model the systems of the chemist-mineralogists, that is, be content to classify the elements of diseases and their more frequent combinations?*"' (Foucault, 2003, p. 146)

He quotes Demorcy-Delettre, an author writing in Paris in 1818. Here we see a significant turn of events where he acknowledges the idea of medicine and the garden of species being replaced by other styles of classification for directing the nosological efforts.

For our analysis, the advantages of the classification of diseases was obvious considering how classes of diseases could facilitate the work of doctors, as prognosis and treatments would follow once the class was identified. This would simplify the tasks and reduce the complexities to be communicated. The class of the disease would say a lot about it and would make irrelevant the detailed description of signs and symptoms. So the analogy with "species" could not be more than a metaphorical attempt but the search for classificatory schemes was indeed very relevant.

To reflect on how little the idea of species could help the classificatory effort, we can make some considerations, which could indubitably be equally made by the nosographers of that time as well. An individual member of a species is a living autonomous unit in its own right; its life has a great degree of independence from the life of other members of the same species. While the independence of a disease in a body in relation to the same disease in another body holds a partial fulfilment of the conceptualization of species as individuals, diseases obviously could not be successfully recognized as autonomous living units. Diseases were studied as disruption in autonomous living human bodies, which were the true autonomous units in these cases.

While a member of a biological (animal or vegetal) species has the autopoietic orientation of reproducing itself, the same could not be said of diseases, as the anatomopathological alteration of a body, the actual disease, is not intended to reproduce its pathological form in another body. Even where infectious agents (not known at that time) aim at self-reproduction, the disorders they produce in the human body are not part of their "project", "intent" or self-reproductive aims; they are rather "undesirable and unintended upshot" that may even lead to the destruction of the invading agents together with the body hosting them.

Furthermore, as a logical construct, species is only the end point in a classificatory scheme. So, the effort to produce a scheme for the classification of diseases may have borrowed the term "species" (possibly a fashionable one at the time) where it could have used other terms such as "classes", "categories", "orders", "types", "kinds", "sets", "groups", "clusters", "families", and so on. Foucault's fixation on the term "species" may be of his own choosing or may be a reliable reading of the original texts;

nevertheless, that does not help to illuminate the main issue at stake, which is complexity reduction.

It is necessary to say that the use of metaphors carries the peril of diverting attention from the main objective, loading the topic with undertones and hints of an unrelated nature. The misuse of vocabulary does not help to reveal the fundamental feature of medical evolution of that time; it may have misled the audiences in Demorcy-Delettre's time as much as it did Foucault's audiences.

From a social systems perspective, there is nothing essential in the use of the term "species" that could reveal a stage in the evolution of medicine as a system; the classification efforts have been present since earlier stages of medicine as a scientific discipline, as those studying it were confronted with an overwhelming mass of descriptions of all sorts of disorders, hugely diverse in their forms and locations in the body. This was in effect the underlying motivation driving the classification aims and processes for organizing a complex picture.

Where could the effort of classification start? Classifications could be anchored anatomically as well as to observable symptoms and signs, as known etiological causes then offered a poor scheme to comprehensively assign diseases from different anatomical regions to the same causal categories. The visible signs and evidence detected on the surface of the living body or inside the dead body still relied on direct observations and descriptions, while causes remained unknown.

So we can say that there has never been a "medicine of species" as Foucault labels it. Diseases may have been conceptually characterized and understood as an entity, a being superseding and imposing itself onto the otherwise healthy body, causing the observable disruptions, but such ontological dimensions were too close to the superstitions and magical phenomena that scientific medicine struggled to remove from its semantics, for the sake of maintaining the tenaciously won and treasured objectivity, in words immersed in religious fervour and restrictions to think outside the religious dogmatic frames.

As the range of known possible causes of diseases was too narrow and unsatisfactory to speak about, language certainly offered shortcuts and convenient ways of quickly communicating what was otherwise a

difficult subject. The history of medicine is also the history of its struggle to communicate with precision and clarity exactly what was going on in the body, even if the causes were not fully known.

The use of the word "species" would perhaps provide concise communication for classificatory purposes, which could however make nosologies less convincing. It would take no longer for a medical scientist to say, "no, the nosology is not a scheme of classification of 'species', it is rather a scheme of anatomical and functional grouping of disorders seen in unhealthy bodies". Nevertheless, I cannot say whether that was the case and Foucault preferred his favoured metaphors.

The point is that we cannot settle with the idea of a "medicine of species" because over the course of medical evolution, if at some point there was a nosography that could describe its classification method as criteria for classifying disease species, that was not enough to say the period between the late eighteenth and early nineteenth centuries was dominated by such "medicine of species". Nosographies did not characterize medicine; they characterized only the type of complexity reduction strategy deployed with specific classificatory objectives.

Furthermore, species as a classification model implies ranking. In the biological realm, it is accepted that species belong to genus, which belongs to family, which belongs to order, and then class, phylum, kingdom and finally domain. In the eighteenth century, Linnaeus' taxonomy, the first recognized and adopted, had five ranks. Surely, it would be a formidable challenge to find the ranking or hierarchy that could explain a disease as a specie belonging to its respective genus, order, class and kingdom, as Linnaeus had proposed for living creatures. There would not be enough points to link the ranks and relatedness. Therefore, no one would be really serious about the endeavour of calling diseases "species in a garden". They probably knew they were simply making poetry.

In conclusion, clearing the field by removing distorting metaphors is part of the process of evolution of any science. The complexity reduction efforts are already arduous, and the metaphors obstruct the job by directing attention to unproductive tasks. Although the successful classification of species in biological realms may have inspired the use of similar criteria, the crucial task for medicine was the classificatory exercise

itself. Doctors knew they were dealing with different sorts of beings. This perhaps is in the end just an example of where Foucault's metaphors may not correspond to the historical facts or help us understand the mindset of the time.

The gaze and the clinical

Foucault is one of the first proponents of the term "gaze", which took on a life of its own and is now found with different connotations in postmodern thinking in fields as diverse as cultural studies, feminism, psychology, critical theory. The term "gaze" appears 209 times in the book. To trace all connotations, qualifications, forms, characterization, contradictions and relations the term has in Foucault's text would be an enormous hermeneutic task. There isn't a unique conceptualization that covers all uses of the term. We can say that the semantic meaning of gaze as looking, staring or observing is obviously part of the meaning wherever the term is used, but still these understandings do not at all exhaust the meanings indicated in Foucault's prose (maybe I should say poetry).

It is important to clarify two points before we start our discussion. First, we approach Foucault's utilization of the term gaze from the point of view of science, where we expect precision. A scientific "gaze", or the gaze of a scientific discipline, has to be seen differently from, for example, an artistic "gaze"; the gaze of the artist (or those appreciating a work of art) is surely different from the "gaze" of a scientist or those using scientific knowledge, in our case the medical "gaze". We cannot address here the debates about the different types of "gaze"; this is not our purpose. But the interested reader will easily find the corresponding literature.

The second point is a brief warning to the reader about the lengthy quotations of Foucault's texts we make at the beginning of this chapter. The reason for that is to convey to readers, particularly those not accustomed to Foucault's style and rhetoric, the feeling of navigating Foucault's narratives. The discussion of Foucault's statements about "gaze" and "glance" (see next paragraphs) may sound chaotic and in need of detailed explanation. But we do not explain each usage of the term; that

would be a tiring and time-consuming job with no valuable outcome. We just want the reader to feel how Foucault presented his views, at least at that early stage in his career.

The gaze comes with many attributes, possibilities and forms as the following short random sample of statements copied from the book demonstrates:

> "it is not faithful to truth, nor subjected to it" (Foucault, 2003, p. 45); "the gaze that sees is a gaze that dominates" (p. 45); "the unimpeded empire of the gaze" (p. 46); "a purified purifying gaze" (p. 61); "the unaided brightness of the gaze" (p. 61); "the gaze that traverses a sick body attains the truth that it seeks" (p. 72); "it is not the gaze itself that has the power of analysis and synthesis, but the synthetic truth of language" (p. 72); "for a language without words, possessing an entirely new syntax, to be formed: a language that did not owe its truth to speech but to the gaze alone" (p. 113); "gaze has the paradoxical ability to hear a language as soon as it perceives a spectacle" (p. 132); "this regular alteration of speech and gaze, the disease gradually declares its truth, a truth that it offers to the eye and ear" (p. 137); "A hearing gaze and a speaking gaze: clinical experience represents a moment of balance between speech and spectacle" (p. 142); "the gaze of the nosographers was a gardener's gaze; one had to recognize the specific essence in the variety of appearances" (p. 147); "the clinician's gaze becomes the functional equivalent of fire in chemical combustion; it is through it that the essential purity of phenomena can emerge: it is the separating agent of truths" (p. 147); "clinical gaze is a gaze that burns things to their furthest truth" (p. 147); "the clinical gaze is not that of an intellectual eye that is able to perceive the unalterable purity of essences beneath phenomena. It is a gaze that travels from body to body, and whose trajectory is situated in the space of sensible manifestation. For clinic, all truth is sensible truth" (p. 148); "Bichat's gaze is not a surface gaze in the sense in which clinical experience was a surface gaze" (p. 158); "The gaze plunges into the space that it has given itself the task of traversing" (p. 166); "the medical gaze embraces more than is said by the word 'gaze' alone. It contains within a single structure different sensorial fields" (p. 202); "The medical gaze is now endowed with a pluri-sensorial structure. A gaze that touches, hears

> and, moreover, not by essence or necessity, sees" (p. 202); "absolutely integrating gaze that dominates and founds all perceptual experiences" (p. 203); "Alone, the gaze dominates the entire field of possible knowledge" (p. 206); "in time, medical gestures, words, gazes took on a philosophical density that had formerly belonged only to mathematical thought" (p. 245).

To make matters even more complicated, on page 149 Foucault introduces the term "glance", and contrasts it with the term "gaze" as saying different things. Glance appears 16 times but only in chapter 7 with a specific denotation compared to gaze. Before or after that point, the term does not appear. We added below an unfortunately long quotation of his writings about glance; again, our justification is that it will give the reader not familiar with Foucault an experience of what we mean, and will also facilitate the comments we make afterwards.

> „The sensible truth is now open, not so much to the senses themselves, as to a fine sensibility. The whole complex structure of the clinic is summarized and fulfilled in the prestigious rapidity of an art: "Since everything, or nearly everything, in medicine is dependent on a *glance* or a happy instinct, certainties are to be found in the sensations of the artist himself rather than in the principles of the art". The technical armature of the medical gaze is transformed into advice about prudence, taste, skill: what is required is "great sagacity", "great attention", "great precision", "great skill", "great patience".
> At this level, all structures are dissolved, or, rather, those that constituted the essence of the clinical gaze are gradually, and in apparent disorder, replaced by those that are to constitute the glance. And they are very different. In fact, the gaze implies an open field, and its essential activity is of the successive order of reading; it records and totalizes; it gradually reconstitutes immanent organizations; it spreads out over a world that is already the world of language, and that is why it is spontaneously related to hearing and speech; it forms, as it were, the privileged articulation of two fundamental aspects of saying (what is said and what one says). The glance, on the other hand, does not scan a field: it strikes at one point, which is central or decisive; the gaze is endlessly modulated, the glance goes straight to its object. The glan-

> ce chooses a line that instantly distinguishes the essential; it therefore goes beyond what it sees; it is not misled by the immediate forms of the sensible, for it knows how to traverse them; it is essentially demystifying. If it strikes in its violent rectitude, it is in order to shatter, to lift, to release appearance. It is not burdened with all the abuses of language. The glance is silent, like a finger pointing, denouncing. There is no statement in this denunciation. The glance is of the nonverbal order of contact, a purely ideal contact perhaps, but in fact a more striking contact, since it traverses more easily, and goes further beneath things. The clinical eye discovers a kinship with a new sense that prescribes its norm and epistemological structure; this is no longer the ear straining to catch a language, but the index finger palpating the depths. Hence that metaphor of "touch" (*le tact*) by which doctors will ceaselessly define their glance. And by that very fact, clinical experience sees a new space opening up before it: the tangible space of the body, which at the same time is that opaque mass in which secrets, invisible lesions, and the very mystery of origins lie hidden. The medicine of symptoms will gradually recede, until it finally disappears before the medicine of organs, sites, causes, before a clinic wholly ordered in accordance with pathological anatomy." (Foucault, 2003, pp. 148–150)

We confess that after reading these paragraphs innumerable times our impulse was to beg for help. If gaze was already a slippery concept to grasp, on which we have been basically left to do the conceptualization work ourselves by grabbing bits and pieces we find along numerous mentions of the term throughout Foucault's book, glance sounds even more imprecise and exoteric. Each sentence is pregnant with poetical indulgence and the choices of words lead to countless questions. So we leave the "glance" here. We only mention it for the sake of offering the reader a hint of our puzzlement; we concede though that those used to Foucault's texts may feel comfortable with it.

As we said, the use of the term "gaze" occurs throughout the book, with new connotations constantly implied. It is hard to avoid certain dizziness and surprise at the verbal licence in using such poetic language. One could be tempted to dismiss the term with the justification

that is nothing other than a bunch of nonsense put together to confuse rather than clarify. Was it not for Foucault's status as a major thinker of the twentieth century, one would set the book aside and forget about it; as Annemarie Mol (2002 a, p. 61) says, "Foucault has been abandoned".

In many instances we do not know whether Foucault is talking about his views on how the gaze were performed then, or whether he is talking about what he thinks were the states of minds, or perceptions, or understandings, or intuitions of the doctors of the time; or whether he is trying to reveal cognitive structures (episteme) acting out at that time, or his understanding of how those structures could have operated then. We do not grasp his intention of whether he writes about the cognizance of eighteenth- and nineteenth-century doctors, or whether about his figurative representation of what the mind-set looked like for him.

The gaze is not a factual procedural thing; we may understand the gaze as a scheme of perception with selective functions and purposes; but we also see that it is a way of typifying how medics approached patients; but in this case, it is a mere artifice of communication that Foucault has invented to suggest a characteristic epistemological posture he believes existed at that time.

We may see the gaze as a caricature of observational approaches; one destitute of contours, uncommitted to an observational strategy for distinguishing one thing from the other based on formalized distinctions; a free approach for looking at anything – and let the "truth" speaks for itself; a naïve characterization of an unworkable observational technique, sustained by an affirmative will, supported from who knows where.

The word observation instead carries a sense of determination and clarity of limits to which gaze is not committed. We may argue that observation was in fact what the doctors and scientists did; they did it before that century, during that century, and have been doing it since – observation informed by distinctions that lead to the differentiations and selections of what was being observed.

From our point of view, the idea of gaze carries a problematic connotation of agency. In Foucault's discourse, it seems, the doctor is not the owner of the gaze. The gaze realizes itself through the doctors; doctors are vehicles, so to speak, and the gaze comprehends broader domains,

beyond the doctor's view. Doctors just channel expressions of that impersonal gaze. From our perspective, Foucault inverted the terms of the relationship. We consider that the doctor observes, is the agent of the observation; on the contrary, in Foucault's terms the gaze is the agent and the doctors its vehicles.

Foucault's metaphors seem to say that the gaze can make many forms appear as the eyes of the doctor scan the disease. The gaze as such is ungraspable. It has an overarching formless attribute that nevertheless can only take concrete forms through doctors' observations. The gaze does not have shape or temporality. The forms it generates are made corporeal and temporalized through observations; the forms are only traces of the gaze itself as the gaze evaporates once the observations are established and communicated. The observations survive in their spoken or written forms while the gaze vanishes. However, the metaphysical nature of such a gaze renders it useless as a concept. Observation is a more valuable concept. When we speak of observations, the agent and the act are clear, as opposed to talking about gaze, where neither is clear

But we return to Foucault's texts and keep the motivation to find in his work a portrayal of medicine of the time, describing its huge efforts to deal with the growing revealed complexities, as the bodies of patients were scrutinized when admitted in hospitals, as well as thoroughly explored in dissections and experiments. This was a time when universities offered institutional sites not only for teaching but also for examinations, reflections and investigations.

Inspired by Foucault's poetical imagery we can figure out the difficulties for the classificatory minds of the time in their attempts to find adequate keys and pathways for building a categorization of diseases. Although the number of identified diseases was far smaller than today, the consideration of the now 100-year-old International Code of Diseases (ICD) is still instructive of the dimension of the challenges. The ICD currently offers a "neat" version with more than 60,000 codes in 22 chapters according to groups of diseases and disorders (ICD 10 from 1990, Wikipedia). In simple terms, any classification effort would have to make decisions, privileging one of the several interconnected dimensions, such as the topographic (by region of the body); the anatomical (by

tissues or organs); the physiological (by function and effect); the pathological (by processes of diseases); the etiological (by causal factors); the legal aspect (by attribution of responsibility); the epidemiological and statistical aspects (by occurrences in collectivities). It is hugely complex, and was certainly already sizeable 100 years ago, at the first publication of ICD.

We thus conclude that, if we want to preserve the term gaze and find some sense in its use, we may consider it as denoting orientation, as indicating an intention to observe although not fully equipped with the determination of the elements to be scrutinized and hopefully found. The observation directionality is not fully explained in the gaze, which only intends to suggest the configuration of the observation while not making explicit the exact points to be observed.

The gaze thus establishes that there is a doctor in front of a patient; both understand something is not going well with the patient and should be investigated by the doctor. The gaze sets the context and configuration where the attention is to take place, while the observation sets the directionality of the attention to the objects (signs and symptoms a patient may show or complain about), examining them in their minutiae. The gaze is prompted by the intention to set in motion a relation between doctor and patient, while the observation is the concrete act of looking at the body and searching for diseases. The gaze is sociological, so to speak, while the observation is epistemological. The gaze wants to be recognized as such, while the observation wants to recognize what it observes. The gaze is accepted and recognized by both sides with each performing its role in the tacit or explicit agreement, while the observation is the doctor's prerogative, with the doctor choosing where and how the observation must be carried out, without requiring the understanding of the patient.

The gaze counts with patients who want to be looked at and expose themselves to the attentive eyes of the doctor, who on the other side wants to look at the bodily expressions of the complaints of the patient. Observation, on the other hand, may be carried out independently of specific requests by the patient, who may be unconscious, sleeping, anaesthetized or even revealed in laboratory results and images dis-

cussed in clinical meetings, far away from the actual body. In autopsies there is no gaze, only observations. Gaze is relational; it is a relationship. Medical observation is unilateral. Gaze is a compromise between looking and being looked at, where both sides participate and more or less accept the engagement.

Doctors talking to each other would never express themselves in terms of "I gazed at the patient and my gaze identified consistency of such and such structures seen in the Magnetic Resonance (MIR) of such pathology". They would laugh at each other. But surely they will go deeper if one says: "I observed such and such alterations in this MIR". This example (which illustrates the agency problem discussed above) helps to clarify why the term gaze has a peculiar meaning, which does not correspond to routine communications among health professionals. Likewise, the word glance would not be helpful; it would convey an idea of superficiality, haste, and perhaps neglectful browsing.

If this is correct, the gaze is as old as medicine, instead of a characteristic of the medical development stage of the eighteenth and nineteenth centuries. On the other hand, medical observation, although also as old as medicine, has evolved along with it, with new categories, distinctions and procedures, characterizing each evolutionary stage's observation capabilities.

Inspired by Foucault, this is the best and most poetical way I could find to explain my understanding of the gaze. However, I should say that I do not claim that my understanding of gaze to any extent corresponds to Foucault's meanings; I cannot claim that I understood him correctly and that my comprehension is fully based on all he has said about gaze. I do not know how close I am to him in his attempt to write a history of the ideas or where I might have "abandoned" him. However, I do not feel the need to clarify that any further.

The French Revolution and its health concerns

Chapter 5 of Foucault's book (Foucault 2003), *The Lessons of the Hospitals* brings a rich description of the revolutionary processes related to the training and practices of medical professionals. The authorities then be-

came invested in regulating entitlements and provision of care. In the revolutionary spirit of curbing the benefits and beneficiaries of the previous regime, healthcare and health professions became objects of attention both to eliminate spurious or suspicious practitioners and practices of the old regime, together with the corporative organization of the medical profession of the time. The French Revolution reasserted the right of the masses to healthcare. Foucault describes these political processes in detail.

In the context of the Revolution, the newly installed powers seeking to end the privileges of the monarchy, including by the closure of universities, brought attention to the masses of destitute and the obligation of the Revolution to address their needs. The reflections on this development perhaps had a decisive influence on Foucault, linking political power and medicine as strategic alliance to strengthen the power of the state.

According to Foucault, the elected Assembly of the Revolutionary Government (National Convention) received requests to intervene in the exercise of medicine by charlatans and incompetent doctors as an urgent matter for the protection of the citizens. Foucault mentions the "Comité de Salut Public", taking a decisive role in drafting bills disciplining the medical profession. He also mentions that: "from all sides demands flowed in for proper control and supervision" (Foucault, 2003, p. 80).

The political processes unfolded over the course of about a decade, with several conflicting interests struggling around regulatory matters of control over professional licensing, the role of hospitals and the organization of care within them linked to training activities.

The subjects and content of training courses were then revised and defined for each year of study in medical school. Doctors' visits to the patients accompanied by students and subsequent discussions were instituted as routine processes. Clinical practice thus defined was balanced against theoretical studies of the texts, which became less prominent than before.

The initiatives developed in the context of the French Revolution show health becoming a matter to be dealt with by the political system;

the political system became aware and was pressured to provide answers to the social expectations and political voices that saw health attention as a matter requiring supervision, organization and orientation by the state.

This implied an irreversible transformation of medicine from a job for the elites (with charity as an aside) to an obligatory topic of social policies. If medical doctors lost some privileges granted by the previous regime, medicine on the other hand became solidly entrenched into political projects as the revolutionary process unfolded:

> „The enlightened classes, the intellectual circles, who had returned to power or obtained it at last, wished to restore to knowledge the privileges that would be able to protect both the social order and individual lives. In several cities, the administrations, "affrighted by the ills that they had witnessed" and "afflicted by the silence of the law", did not wait until the legislature had made its decisions: they decided to establish their own control over those who claimed to practice medicine; they set up commissions composed of doctors of the Ancien Régime, who would pass judgment on the qualifications, knowledge, and experience of all newcomers." (Foucault, 2003, p. 81)

In this description we see that medical doctors arrived at a position of self-regulation, now with a broader scope. In that process we can say that public health was established as the articulation of policies and healthcare services, with medicine providing care within an overarching policy-defined framework.

We nevertheless understand that was not the birth of the clinic as the title of Foucault's book indicates, as the clinic had already existed for centuries, as taught medical practice combining established knowledge with the still rudimentary treatments available. Instead, we understand it was the birth of the health system, as a differentiated system coupled with the political and legal systems.

The health system as a composite of medical sub-system and public health sub-system was thus established, with the drafted regulatory framework defining who was and wasn't entitled to practise medicine,

who could receive medical attention in the country, and how the medical profession could control and deliver training programmes (only doctors could train students to become doctors). The following passages copied from Foucault (2003, p.90) gives a picture of the on going spirit:

> "Article 356 of the Directoire Constitution declared that 'the law supervises those professions concerned with the health of citizens'; it was on the strength of this article, which seemed to promise control, limitations, and guarantees, that all the polemics were conducted."

> "[…] the controversy was centered mainly around the question as to whether one should first reorganize the system of teaching, then draw up the conditions for the practice of medicine, or, on the contrary, first purge the medical body, define the norms of practice, and only then decide what form medical studies should take. Between these two theses, the political division was clear-cut; those least removed from conventional tradition, such as Daunou or Prieur de la Côte-d'Or, wanted to reintegrate the officers of health and all the amateur practitioners of medicine by providing a very open system of teaching; the others, around Cabanis and Pastoret, wanted to hasten the reconstitution of an enclosed medical body. At the beginning of the Directoire, it was the first group that had most support."

The political push was crucial to create the conditions for self-reference of the health system, now comprising all necessary features. Although Foucault does not speak of social systems, self-reproduction or self-reference, his rich description of the historical processes allows us to see the genesis of health systems from the perspective of the social system theory.

The French Revolution therefore created the context for definition of the social role of medicine. The political system was acting, and doctors interacting with it, to achieve objectives as the opportunities appeared in the awakening of revolutionary "politicization" of all topics concerning the population and the state, without limits to where the state should stop. Everything was up for grabs in the sense that it was in need of or

vulnerable to interventions and state regulation. The old order collapsed and a new order was being built from scratch.

Health was not yet a system, so it did not have autonomy and the political system therefore intervened in everything. At early stages in the process, doctors perceived the chances and the need to take some control. The revolutionaries accepted it because medicine was too complicated a matter for them to decide upon. So the political system took on the role of regulator and the doctors performed acts of self-regulation. That was a key contribution of the Revolution. The political process unfolded while other evolutionary processes were also unfolding in the field of scientific medicine, medical practice and training.

Foucault describes medical professionals managing to acquire voice and influence over the course of the decisions the revolutionary government put in place. The main concerns were with the training of doctors, control or elimination of charlatans, mobilization of professionals to meet the needs of the military fighting a war, as well as hospital institutions for the treatment of wounded soldiers and the population lacking resources.[1]

From the Social System Theory point of view, what is essential in the French Revolution and the genesis of health as a social system is the political constructions (with doctors' engagement) defining how medicine should be provided, to whom (the population to be covered) and how doctors should be trained and allowed to practise. These constructions were carried out through regulatory instruments, projecting the image of the system, through which the system could create its self-identity. Professionals knew they became members of a system that regulated, organized and supervised their activities. In that construction, they themselves were designers and executioners of the acts of systemic order. The identity thus created was self-identity, in the way that it set clear social

1 In "The *Officiers de Santé* of the French Revolution: A Case Study in the Changing Language of Medicine", Maurice Crosland (2004), presents a detailed description of the processes unfolding during the Revolution concerning the definitions and regulations of health professionals.

distinctions and boundaries for those who could see themselves as being part of the order and those who could not.

Identity was not only a tautological matter but also distinguished precisely what society would get – how the system would be constructed, comprised of distinct parts such as recipients of medical attention and people with entitlement to it (or not). Self-reference, and its opposite side, hetero-reference, thus emerged together with the health system (hetero-reference being the necessary other side of the distinction).

Medicine had reached the stage (given the semantics and complexities that it developed) where only medicine could define medical values. This is the constitutional basis of the health system. The political system could create the political space for it, recognizing medicine's own semantic closure and complexity, but the closure itself had to be created and reproduced by medicine itself. The doctors had to be fully on-board with the decisions the Revolution was making.

In our understanding, as explained in other sections in this book, medicine appears as both a sub-system in the scientific social system and a sub-system in the health social system. As a sub-system, medicine has the systemic self-reference attribute to designate what are medical matters – that is, matters that medicine takes to itself as being of its exclusive concern. Medicine thus defines itself by taking on itself the notions of what it is about. Only medicine has the authority and legitimacy to say what it is and what it is not. All that came fully into being with broader political and legal confirmation in the French Revolution.

Canguilhem

Georges Canguilhem's book *The Normal and the Pathological* takes us to the debates of the late eighteenth and early nineteenth centuries and the difficulties of the time in relation to the characterization of normality in medical matters. There was a quest to distinguish the signs and symptoms of diseases, the thresholds for moving from one to the other, and how to treat them.

Canguilhem shows us the struggles in medical thinking to find the order, the schemes, structures, to clarify the nature of normality and of its opposite, pathology. The questions of where the normal stands, where and how pathology settles, and how normality and pathological phenomena interlink with each other, perhaps sharing the same essence, evolving from one to the other, or inhabiting the same body in discontinuous, segmented and mutually influential interferences, succumbing or fighting one another off; these were all puzzling questions.

The advances in anatomy, physiology and pathology, with detailed descriptions and large numbers of recorded cases, made those questions more challenging than they had ever been. If, for example, since the Greeks, the four humours framework had settled enquiries, however unsatisfactory the explanations may have looked, the abundance of new materials and knowledge sharpened the senses and led to the need for more reliable understandings.

In a context where other sciences were showing outstanding advances and the logic of scientific rationality was celebrated as the way of knowing and answering any question about the world (including diseases), the disquiet led to thorough reflections on the nature of the phenomena being studied: health and diseases.

The context was no longer amenable to superficial explanations and rudimentary models. The complexities revealed inside the bodies demanded equally complex and comprehensive explanatory models. Canguilhem discusses the reflections of the most prominent thinkers of the time.

We can here just briefly mention some of the names and the points that arose in Canguilhem's book. Our purpose is not to write a review of the book; we aim at considering the struggles discussed by Canguilhem from the point of view of the Social Systems Theory. He described the evolution of the health system at its birth, where the semantic binary opposition healthy/sick took the form of normal versus pathological.

Canguilhem (1978, p. 20) discusses how at that time the theme of "the real identity of the normal vital phenomena and the pathological" was the object of investigations to be settled in quantitative terms, despite

the apparent radical differences. This theme and related ideas were at the centre of the controversies of the period.

August Comte (building on François Joseph Victor Broussais's work) and Claude Bernard are the prominent pioneer thinkers in Canguilhem's narrative; Comte departed from the pathological to explain the normal while Bernard proceeded in the opposite direction, although both maintained the same point of view that the normal and the pathological were phenomena of the same nature.

"The diseases are only the effects of mere changes of intensity in the action of the indispensable stimulation for the preservation of health" (Canguilhem, 1978, p. 26) was a vision inaugurated by Broussais (eighteenth century) in opposition to the then current understanding of the ontological distinctiveness of health and sickness as two separate entities. According to Broussais, and maintained by Comte, pathology is the normal physiology under the stress of over or under excitement. The pathological can be traced to quantitative variations on what a normal body handles – as Canguilhem (1978, p. 74) notes, the "real homogeneity and continuity of the normal and the pathological".

Bernard considered medicine the science of diseases and physiology the science of life. Rational therapy can only find its basis in a science of pathology, which nevertheless has to be based on the science of physiology. However, physiology and pathology cannot be separated and are essentially the same thing. Bernard opposed the idea, popular among many at that time, that diseases have extra-physiological elements.

The discussion of the unit of the normal and the pathological appeared animated by the advances in physiology's aspiration to find quantitative explanations of the differences between the normal and the pathological. In contrast, the Hippocratic tradition did not delve deep into such concerns as its interests were focused on describing external observations of sick bodies. Diseases had an implied ontological nature, which persisted until the conception of identity and quantitative differentiation was formulated in the period covered by Canguilhem.

Furthermore, René Leriche (nineteenth century) another author to whom Canguilhem dedicated close attention, is quoted saying: "health is the life in the silence of the organs"; "diseases is what disturbs the men in

the normal exercise of their life and occupation, and above all what make them suffer" (Canguilhem, 1978, p. 63). According to Leriche, the body has more physiological capacities than is usually known, and diseases are necessary to reveal them; in other words, the body has more capacity than required to keep living.

Medicine seemed to have changed from the idea of the confrontation of opposites to the idea of a unit, with variations of continuity and discontinuity, leading to the notion of health and sickness as a continuum. But the complexities were not amenable to easy and simple explanations in terms of such variations.

The question whether the pathological state was or was not just another expression of the normal physiology was apparently settled but other thinkers approached the topic differently. Xavier Bichat, also mentioned by Canguilhem (1978, p. 93), was a prominent figure in the period and contributed to strengthening the distinctions: "in the phenomena of life there are two things: 1 the health state and 2 the sickness state: from that appear two distinct sciences, the physiology, that works with the first, and the pathology, that works with the second".

However, if the distinction between normal and pathological could instruct on how to approach patients and diseases, the arsenal for identifying etiology and causes was insufficient. Although the concept of etiology had existed since medieval Islam, medicine was observing complex phenomena, but not so much their causes.

Canguilhem also referenced Adolphe Quêtelet and his anthropometry. Quêtelet worked on the quantification of what is normal or not in human health. The true mean offers ontological distinctiveness, constituting a norm. Discussions were then opened in relation to the "median man" and the variety of factors (among them biological, social, environmental) influencing human anthropometric characteristics and the suitability of such measurements for the definition of normality and abnormality.

Surely such close attention and intense discussions of the nature of the normal and the pathological had to be revisited in face of the discoveries that followed. The cells, the tissues, the germs, the hormones, the

genes and so on – all new discoveries placed questions of the normal and the pathological on new ground.

Etiological findings changed the conception about the nature of health and disease. When Louis Pasteur entered the scene, for instance, his germ theory redesigned the horizons separating the normal and the pathological. The reaction of an infected individual could not be classified as abnormal as opposed to normal, as a normal body was reacting normally to an invasion of external organisms. Abnormal would not be an appropriate word in such a case. The physiology of a patient with an infection is not essentially different from that of a non-infected individual.

For another condition, but also putting the concepts of normal and pathological under stress test, Canguilhem (1978, p. 104) says: "All the functions of an hemophilic are carried out as in healthy individuals. The only difference is that the hemorrhages are endless". This is another example of how fluid the terrain was when the concepts of normality, abnormality, health, pathology, diseases, and so on were trying to settle.

From our point of view, Karl Jaspers, also quoted by Canguilhem, seemed to move the controversies in the right direction, away from the rather complicated dialogues about where the normal stops and abnormality starts. Jaspers said: "the doctor is who less enquires about the meanings of the words health and sickness. They are concerned with the vital phenomena". "The 'sick' is a general concept without value that comprehends all negative values possible" (Canguilhem, 1978, p. 88). According to Jaspers, what is desired in terms of health from a physiological point of view is evident, and this adds to the notion of stable recognition of physical infirmity.

These examples, in fact only a small sample of the controversies addressed in Canguilhem's book, are only intended to illustrate the evolutionary stage of medicine, when it became able to discuss crucial notions of health and sickness, and to keep the topic open to further considerations as the knowledge advanced. Medicine's internal environment allowed it to have such key discussions and not be discredited by such critical processes. From the social health system perspective, this is evidence of medicine becoming skilled in addressing its most difficult and

contentious issues, while preserving its coherence and integrity in the midst of radical controversies. Our thesis is that medicine had therefore reached its "adulthood" and the health system was ready to be born after public health entered the scene.

Health system

In this section we discuss Foucault's and Canguilhem's points from the Social Systems Theory perspective. In line with the theory, the health system as a function system is based on a binary distinction, in this case, healthy/sick. The discussion about the normal and the pathological back in the nineteenth century and earlier was therefore a constitutional basic polemic, so to speak. It was necessary to make that distinction, bringing it to reasonable operational stability and consistency, although this did not require settling rigorous and precise definitions that could characterize the limits between the two sides of the distinction.

The systems theory states that a distinction should be established without a third term: one could be either healthy or sick (likewise, in the legal system, a judgment should be whether an act is legal or illegal; in the political system, the political statement should be of the government or the opposition; in the system of art, a piece is recognized as art or not; and so on); a social system would not be able to handle a third term. Even if the movement from health to sickness (or vice versa) cannot be precisely traced and the crossing point indicated, the crucial issue was (still the critical one) to assign the individual to either of the categories, so the health system could operate (communicate) accordingly – that is, on the side of sickness. There should not be any in-betweens. Canguilhem makes visible this tension in his historical analysis, although he did not have the Social Systems Theory among his references and was not trying to argue in that regard.

In relation to the fundamental binary distinction that is the basis of social systems, a long discussion can be held on the meaning of normal (in accordance with norms) and their opposite. Both Canguilhem and Foucault are interested in the genesis of that notion and its implica-

tions and ramifications. Canguilhem is more concerned with the medical theme of normality and Foucault, in his subsequent works, more interested in power effects of norms and how medicine became part of strategies of disciplinary coercion applied through different methods in society. An in-depth discussion of these themes would take us far away from our main focus. However, a few additional points can be made.

Social Systems Theory does not delve into the theme of normality, which is a slippery concept. There is a huge diversity of meanings associated with the word normal. It may appear in the medical discourse, for instance, in reference to the pulse of a patient and results of tests and examinations, but apart from strictly observed and measured parameters, normality are not where medical attention should focus.

Etymological studies tell us that the Latin word "normalis", the origin of the word normal, had its meaning in carpentry, denoting the fitting of a pattern of angles, without connection with the meaning of norm (juridical form), which then added the connotation and correspondence to a value system. From the matching with the notion of value, normal was introduced into the medical vocabulary, meaning agreement with a given pattern, referring to bodily measurements such as height, weight, pulse) – without prescribing normative connotations but rather simply identification of deviations.

However, the actual diversity of meanings of norms can be illustrated with the ten different meanings for normality in mathematics alone. In general, Wikipedia would tell us, there are eight fields of science and technologies using the word normal, each with specific meanings. In the common language there are around fifteen different synonyms. Normality, normalization, norms, normative, anomaly, normal, abnormal, all these terms can be used to convey specific technical references in specific fields of science or technology.

The distinction of normal and its opposite, abnormal, does not seem therefore to suit the operations of a health system. In fact, the health system is not based in such binary distinctions, and this cannot replace the healthy/sick distinction. Although the term normal is often used in any health system, it always refers to a standard (often quantitative), which indicates thresholds and limits of the two sides of the distinction. The

blood test or blood pressure is normal, we may hear health professionals saying, meaning it is within the expected parameters. This conforms to the general use of the term normal in whatever technical field it is used – denoting conformity with established adopted limits and parameters.

As a disease affects the body, it would be an inappropriate stretching of the term normal to apply it beyond the technical parameters it refers to. Abnormality, on the other hand, would suggest something different – perhaps that the whole body does not correspond to human characteristics, like an anatomical monstrosity (Canguilhem, 1978, p. 98). Let's say therefore that abnormality would denote a very severe occurrence, rare and strange, not necessarily recognizable as a classifiable disease.

On the other hand, the discussions of that time seemed to have settled with the understanding that the pathological could not be a synonym of abnormality or something in clear contrast with the norms. Pathology was thus recognized as possibly being another facet of normality, being part of the repertoire of the body's physiology. While something seems not to be healthy in a particular body, many other observed variables might be perfectly within the expected parameters; yet the person is undoubtedly ill.

For all of that, and for the "heavy" denotations and connotations of the term "norm", a health system would be on more solid ground in avoiding the normal/abnormal distinctions. Normality and norms would also open the semantic space to a lot of confusion, as both words are widely used in other fields. The word "normal" would be loaded with connotations while the binary opposition health/sickness is not "disputed" by any other field of knowledge.

A final word on the notion of "norm" as Canguilhem uses it is necessary. He discusses the theme ad nauseam in his book, covering a lot of ground. However, we think he left something unaddressed. As we said previously, norm appears in many areas of knowledge. A "juridical norm" about legal processes and legal normativity is an arbitrary convention in the sense that it is the result of a collective decision without any essential foundation in the world. Therefore it is contingent and changeable. The same can be said about any prescriptive, conventional, technical norms.

In this sense, a "therapeutic norm" is also the fruit of decisions; surely it should have a foundation in empirical observations and should be justified as much as possible without resorting to arbitrary convention or judgement. Therapeutic norms (such as the evidence-based clinical guidelines) are intended to facilitate medical decision-making where a doctor's knowledge may need support. Sometimes the norms are issued for economic, cost-containment reasons, but still must have strong medical rationale, carefully weighting the causal relations at stake. Therapeutic norms may also be contingent, but surely less so than legal norms.

Canguilhem calls therapeutics "normative activity", addressing notions of the physiological functions as norms of the living body, which takes him far away from the sense that normativity is the business of human beings deciding which norms they want to put in place. Biological organisms do not follow "norms". Contrary to that he says, "medicine uses results obtained from all other sciences to serve the norms of life" (Canguilhem, 1978, p. 176). We would question this: there is no such a thing as "norms of life", unless one is talking about enacted regulations.

To be acknowledged and followed, norms need to be written and communicated. The denotation of convention has been carried by the word "norm" since its oldest etymological use. In Luhmann's (2008b, p. 54) terms, "norms stabilizes expectations"; norms are enacted for that end. Biological blind determinism does not obey norms as such; norms always carry the idea of observability (compliance can be checked) and contingency (they can be established differently). So, we would say that the health systems did a very good job for themselves by avoiding the conundrum of the concepts of norm, normal and normality, building a semantic architecture using the exclusive binary distinction healthy/sick, however imprecise it may sometimes look.

Conclusions of the chapter

These conclusions bring the Social Systems Theory to the foreground; the section is organized in three parts: in the first we revisit Foucault's ref-

erences to the Greeks; the second focuses on complexity and meanings at the beginnings of the health system; and the third talks about the consolidation of the social health system.

Foucault and the Greeks

"[F]ifth-century Greek medicine would seem to be no more than the codification of this universal, yet immediate, clinical medicine; it formed the first total consciousness of this clinical medicine" (Foucault, 2003, p. 66). Foucault is clearly admitting that clinical medicine existed from Greek times. Maybe the title of his book could therefore have been "The birth of the gaze", as he tried to characterize the way clinical medicine started to be practised in the late eighteenth and early nineteenth centuries. Yet in the same paragraph on the advent of clinical medicine he says:

> "but insofar as it organizes it into a systematic corpus in order to facilitate and shorten the study of it a new dimension is introduced into medical experience: that of a corpus of knowledge that can be said to be, quite literally, blind, since it has no gaze".

If we understand him, he meant to say that the systematic recording of cases in a written "corpus of knowledge" (which, according to historians of Greek medicine was to serve as a reference for subsequent observations and recording) was "blindfolded", requiring continuous reading of descriptions in the texts. In consequence "it did not have a gaze". If we understand Foucault correctly, the clinic that started in those ancient times was "clinical" but "without gaze". How could that be possible? According to historians of the Greek period, the Hippocratic Corpus had rigorous recommendations for observing and describing the manifestations of diseases in the body. How could the doctors observe without "gaze"?

We have already spoken a lot about "gaze", so we only made this point here to restate our conclusion that what really appeared in the eighteenth and nineteenth centuries in France was the systematic assembly of medicine as an established distinctive university-based

medical knowledge/profession, and the political orientation to regulate training, practice and distribution of medical services; this orientation constituted a nascent public health function. This was the beginning of health as social system.

Despite Foucault's praise of the period, the systematic and careful attention to diseases and descriptions of what was observed obviously did not appear for the first time in the eighteenth and nineteenth centuries in France. Scientific medicine in Hippocrates' Greece emerged with the attentive observation of external expressions of the diseases followed by detailed descriptions of what was observed. Although the period Foucault studied benefited from routine autopsies and studies of corpses after death, which allowed the observation to venture deeply inside the body in search of connections between clinical and post-mortem findings, observation of the surface of the body was part of medicine from its very beginnings. Thus Foucault exaggerates the historical occurrence, and the "birth of the clinic" actually happened many centuries before. Looking at sick bodies as far as the rules of the time allowed, the scientific medicine started by the Greeks had to comply with observation constraints existing from its very start. By the same token, the doctor/patient frame of interaction always coexisted within the feasible limits of observations and practices.

Foucault in fact does not talk much about the Greeks, and only makes few references to Hippocrates. He compares the Hippocrates observation with the "pure gaze" and the new observation style of the period he studied as "at the same time" "pure gaze" and "gaze equipped with […] logical armature", following simple "empiricism":

> „"Hippocrates applied himself only to observation and despised all systems. It is only by following in his footsteps that medicine can be perfected". But the privileges that the clinic had recently recognized in observation were much more numerous than the prestige accorded it by tradition and of a quite different nature. They were at the same time the privileges of a pure gaze, prior to all intervention and faithful to the immediate, which it took up without modifying it, and those of a gaze equipped with a whole logical armature, which exorcised from

> the outset the naivety of an unprepared empiricism." (Foucault, 2003, p. 131)

And he continues: "In the clinician's catalogue, the purity of the gaze is bound up with a certain silence that enables him to listen. The prolix discourses of systems must be interrupted: 'All theory is always silent or vanishes at the patient's bedside'".

Although we can see in these sentences his recognition of the predominance of empirical observation in Greek medicine, he does not admit that there were no other options then; the first empirical systematic observations had to be on the surface of the body and could not be anywhere else. However, his remarks remain problematic. This combination of having a "catalogue" and being silent at the "patient's bedside" makes a poetical construction with which doctors most certainly would not identify.

Theoretical and rational frames of observation, in any scientific endeavour, or any technical undertaking, require a continuous reflexive and recurrent process of going repeatedly back and forth, looking at the evidence in signs and symptoms and corroborating those observations with the organized textual categories of knowledge. The silent gaze goes nowhere if its steps are not diligently and continuously constructed, followed, revised, communicated and repeated by the observing doctors, often resorting to the texts they may have at hand; this could not have been different, and never was, unless there were no texts and no peers to dialogue with, and consequently no scientific medicine.

Complexity and meaning surplus

From the seventeenth century onwards, medicine used very rich descriptions. From the social systems perspective, the available semantics comprehended an inventory of anatomical descriptions as well as identified signs and symptoms with which doctors could communicate their observations. To illustrate the context, Andreas Vesalius' book *The Fabric of the Human Body*, published in 1543 and 1555, held detailed descriptions with hundreds of illustrated pages covering all anatomical features of the hu-

man body. The use of such wealth of descriptive terminologies depended on the doctor's abilities and training and also on the contexts where the communications took place among peers, trainers and trainees. From a social system perspective we can say that in the eighteenth and nineteenth centuries medicine had gathered, so to speak, an excess of *potential descriptors* of a disease being observed.

The extensive vocabulary describing structures and functions of the body could be articulated in many ways, producing interpretations that, however tentative, as they could not easily be empirically confirmed, constituted a formidable requirement for consistent medical communications. That surplus of signifiers (anatomical, physiological and pathological descriptions) available to doctors nevertheless still had fragile notions of causal links and relations.

Besides that, treatments were scarce and of low efficacy, with the same procedures (such as bloodletting for example) indicated for countless different types of disease. We can therefore see that there was a peculiar configuration with, on one side, a rich vocabulary and repertoire of descriptive terms, and on the other side, very poor means for verification of diagnosis hypotheses and corresponding treatments.

At that time, there were in Europe many examples of successful advances in scientific fields. That encouraged scientist doctors to experiment and investigate. The quantification of physiological phenomena, for example, had many followers, although qualities such as dryness, ardour, debility, discharge, excitation and tiredness, and subjective expressions of disease (Foucault, 2003, p. 14) were challenging for efforts at quantification.

We could say that the context reveals medicine as a hesitant science in the process of development, trying to find possible ways of moving forward. Tentative semantics could articulate communicable sentences, meaningful and well-understood constructions to be further discussed, articulated, accepted, replaced or dismissed. This was in fact happening in the universities and hospitals on a daily basis, wherever doctors were working. But the conclusions in terms of causality and the results in terms of treating patients remained poor.

Foucault does not employ notions of complexity or complexity reduction, concepts often appearing in Luhmann's texts. However, complexity is a key concept for understanding the period. Complexity reduction has to be obtained with selections of what seems relevant in a given patient, and also selections of what is characteristic of diseases already known, as described in the texts. At that juncture, there is a medicine on one side with a mass of findings related to the human body (anatomy, physiology, pathology, and so on) and on the other side a mass of evidence and descriptions of countless patients admitted into hospitals. The selections to be made were becoming increasingly demanding.

The experience of the hospitals increased the turbulence of the seas medicine was navigating. The quantity of patients in close contact, perhaps affecting the diseases the patients already had, was an additional source of complexity to be considered in what was observed. This complex context could be overwhelming, demanding doctors' attention to too many possible signs and symptoms of disease, and also factors of the environment and the procedures.

In short, anatomical, physiological and pathological studies progressively ventured into highly complex inner worlds, accurately describing them. The deepening of the observations produced correspondingly highly complex descriptions of everything being observed. Medical semantics became increasingly differentiated, sophisticated and demanding for all who aspired to be part of the profession and be recognized as such.

These two processes – the recognized complexity of the body and its diseases and the corresponding complexity of the medical language – called for systemic solutions. The new corpus of knowledge needed to be protected and reproduced with guarantees of validity for their deployment. The complexities needed to be reduced to ensure consistency, stability and continuity – and at the same time openness to what was still to be investigated and uncovered. All of that could only be achieved through mechanisms of what in Luhmann's terms we call progressive systemic operational closure, with self-observation and self-regulations.

Closure of the system

Operational closure suggests a system that has reached the end of its embryonic development. From that point on the system can select the communications that are identified as having a legitimate place in it. This includes all of a system's validated exchanges between professionals, and also includes selecting out communication attempts not capable of using the acceptable semantics.

By establishing such closure, the system is already pursuing its autopoiesis – performing its self-reproduction with the means it creates itself. The system became capable of self-observing and thus distinguishing self and hetero observations, separating meanings belonging to the system or not, according to the recognizable semantics.

Concurrently, the sophisticated and complex language also becomes exclusivity of the system. From then on, no other system would be able to communicate using it. Medical terminology became fully intelligible only for members. In a single stroke, medicine became able to recognize itself (self-reference), and also able to make itself recognizable as external (hetero-reference) from the perspectives of the other function systems.

Illustrating the process, the anatomical-pathological findings did not impose themselves by their own arguments. Rather, they encountered fertile ground where scientific observation was becoming trusted and relevant. A medical profession looking for ways of affirming its identity and at the same time searching for substantive answers to countless questions derived from the problems encountered at patients' bedsides, welcomed the new world of reflections. Medicine understood the value of the new ways of communicating about diseases and treatments.

It can be said that at that time, in the eighteenth and nineteenth centuries, the possibility of descriptive communication about the diseases (by following the patient all the way through to autopsy) was recognized as having a value of its own, detached from the frequent and rather frustrating attempts to treat and cure.

Few treatments were available in the crowded Parisian hospitals, and they had little or no efficacy. Medicine was nevertheless being built on

the communication about the findings being made, rather than on the successes being achieved. If the health system had depended exclusively on successful treatments, then it would never have appeared.

The closure of the system made acceptable the internal communications about disappointment with treatments, because understanding failure was as important as understanding successes. While there was acceptance, tolerance or even encouragement of internal controversies within the protection of the system, it could also protect from external controversies that could bring destructive impacts from outside the system. In Luhmann's (2013) terms any incursion of the environment into a system has destructive effects on the system. The system therefore could not remain open (without closure) and had to ensure the exclusive internal validation of its communications.

The medical science was trying to precisely define its domain, but what it managed to set was the grounds where controversies were accepted and the self-reflection of medicine could take all polemics on board, as nobody else could in fact interfere. In short, the closure created a safe environment for medical self-reflection.

Closure and power: a few words

In line with the Social System Theory we understand the operational closure of medicine was not aiming at the effects of normalization for social domination, as perhaps a Foucauldian tradition would have it. In general, medicine's driving concern is with the marked side of the healthy/sick distinction: sickness. It is interested in the diseases: where they appear, how they progress, how they can be contained, cured or not – normality does not tell us how or why a disease will or has appeared. There is no normality in etiology; normality is what is left behind.

Many power relations in society had already existed for centuries, before the strengthened roles of physicians in relation to their patients. Doctor/patient was not a new power relation. The health system would not achieve its closure with the intent to provide society with a new

model of how to exert power inspired by the "domination" of the patient by their doctor.

The health system was too busy trying to ensure its effective closure and the sustained preservation of the flow and content of its internal communications; its selections of semantics were still being stabilized and secured. Worrying about the benefits other social systems would gain from adopting its alleged "power strategies" would be far removed from medicine's concerns.[2] The survival of the system was the chief concern. The confirmation and permanence of the new semantics was the absolute priority. Everything that could help that objective was welcome. Coupling with political, legal, economic, religious and other systems would be pursued with the intent of getting those assurances.

Power over patients was less of an existential necessity than the need to internally communicate using the common semantics being developed. What was done with patients was primarily justified in the eyes of the peers; the patients themselves would hardly be equipped to understand the communications. The distinctiveness of the medical language was not primarily intended to alienate patients or keep them unaware about what would be done with their bodies. The complexities of the medical justifications, as well as, for instance, the complexities of legal reasoning argued between judges and lawyers, are features of semantic expansion within the confines of systemic closure, in the process reinforcing and reproducing it.

With Social Systems Theory a particular conception of society appears. A society organized in function systems cannot be assumed to follow a single plan of power and domination. Each system strives to survive according to each one's drives; autopoiesis of each function system is pursued independently and if the coupling between systems is possible, and often happens by increasing the likelihood of individual systems' autopoiesis, the coupling does not follow a macro "narrative" intended

2 In a recommendable discussion of Foucault's views on use of medicine as strategic power, Mol (2002a) raises critical points. For example, she says: "And however influential it may be, (medical) science does not have the power to impose its order" (p. 61). Mol adopts Bruno Latour's perspectives in her reflections.

to foster a specific world view and social hegemony and dominance. Any individual system's priority is its survival. The system survives and associates with others; there is no macro script coordinating the arrangements between all systems aiming at achieving a selected end. This is a fundamental feature of Luhmann's (2014) grand theory, which often contrasts with other ways of seeing social order and social evolution.

Final remarks

The works of these two authors give us elements to corroborate our thesis that Social Health Systems emerged in the late eighteenth and early nineteenth centuries, with the initial convergence of clinical medicine and public health during the French Revolution, representing the first political exercise to serve a population, regulating a self-referential medical discipline while formally recognizing medicine's self-regulation prerogative; the self-reference of the health system was thus inaugurated. As Luhmann could have said, a self-referential system organizes the observation of its observations and the descriptions of its descriptions in the second order of ordination (Luhmann 2013) – that is, second-order observation being self-observation as observer of itself. Medicine thus had fully achieved the structures and functions to become the core of the health system.

PART III

Chapter 6: "Anatomy" of Public Health: Indicators and Self-Reference

Introduction

We may say that public health "feelings and intuitions", understood as an orientation towards the health of populations, can be traced back to the Greeks and even earlier. We can find ancient views that seasons, miasmas, emanations from swamps, stagnant waters, human waste not properly disposed of, for example, could have negative effects on people's health. The environment was thus long ago recognized as important for the health of populations.

Habits of hygiene, such as regular bathing in clean water and good cleaning practices, were important for healthy lives. Fresh and good-quality food was also seen as important. The links between health, behavioural and environmental factors have thus been imprinted in the minds of many cultures for centuries. In one way or another, all societies, even the most primitive ones, adopted and valued principles for obtaining and keeping good health. Good health is of high value. We are not saying anything new here, are we?

Fast-forwarding, by the second half of the eighteenth century, with, if not abundant, the already well-established presence of doctors in European societies, the relevance of environmental and behavioural factors (including hygiene) for the health of individuals and populations were seen in a different light. Doctors now could study and identify the possible problems related to those factors. Medicine was gaining the position of having voice on those matters – a voice the governments of the time

listened to, and perhaps accepted the advice. The convergence of medical attention and the interest of those in power were amplified as societies faced devastating epidemics. Besides that, armies had to be kept in good health and fighting order, and cities needed to be places of reasonable living conditions.

Meanwhile, however, medicine and its rather "enigmatic" language of organs, tissues, humours, normality, abnormality, and so on, became increasingly distant from the conversations politicians and city managers were having. They could nevertheless learn more about health risks and how to protect populations. In this case, the actions to be carried out were mostly outside the realm of medical acts concerned with sick individuals. It was easier for politicians to recognize the value of improvements in cities' dwellings, for instance.

Furthermore, even appreciating the medical profession as of public interest, and the need for regulating their training and licensing, the benefits medical professionals could bring for the political powers depended on the numbers and distribution of professionals in the territory, regardless of the effectiveness of treatment those professionals were delivering (which we now know was low).

Medicine thus entered into public decision-makers "Seeing like a State".[1] But with medicine developing so that only doctors could understand what doctors were talking about, and public powers unable to make decisions on matters of the body – constitution, anatomy, physiology and other bodily functions – medicine remained excluded from the "Seeing like a State" endeavour, confined within the space of meanings it had developed for itself.

The potential social fruits of medical practices could be incorporated into the thinking of political systems, as long as it could be incorporated

1 *Seeing Like a State* is a book written by the anthropologist James Scott (1999). He was not influenced by the Social System Theory and does not refer to it in his studies. However, his book, critical of public initiatives' lack of consideration of local diversity, sheds interesting light on the processes, development and outcomes of public sector initiatives. His work does not focus on health topics though.

into the schemes of public administration. So they did. The political systems recognized the need for politicians to have closer links to medical practitioners so that they could discuss and reach agreements. Public health was therefore born out of: (1) the necessity for medical professionals to find a way of dealing with politics, to protect themselves from nasty surprises and threats from that direction; (2) the necessity for political systems to use, appropriate and capitalize on the prestige and social recognition of doctors (even if the outcome of interventions were still not convincing), and in the process implement intentions to improve life conditions for everyone in the society. The picture can thus be summarized as follows: politicians and doctors had a lot to talk about, but they needed intermediaries who could translate their different languages into something both could recognize. Public health found its place then; where it still is.

Fast-forwarding again, we can now move to the present context where public health is established in key departments in any Ministry of Health, any Faculty of Medicine, and any university around the globe. It also is at the heart of major international players like the World Health Organization (WHO), the United Nations Children's Fund (UNICEF), the United Nations Population Fund (UNFPA), Joint United Nations Programme on HIV and AIDS (UNAIDS). It conquered spaces in development agencies and multilateral banks providing support throughout the developing world. Public health has also found a solid position across non-governmental organizations orientated towards health and related topics.

Public health has acquired a characteristic language that can be understood all over the world. That language, we can say, has become universal. Any public health professional from anywhere in the world would be able to talk to any other and reach a good deal of understanding based on the terms they can deploy in those conversations. In this language a particular vocabulary is at the centre of communications; we are talking about health indicators.

This chapter is dedicated to discussing and reflecting on the health indicators and addressing them from the Social System Theory perspective, according to which we can say that indicators are instrumental for

the creation of public health self-reference as well as the self-reference of the health systems where the health indicators are communicated. In the sections to come we will dive into the health indicators and their utilization. We call this chapter "'Anatomy' of Public Health" having in mind that indicators indeed show the "spine", "muscles" and "tissues" of the health of the public.

The form of problem

The Health System Performance Assessment (HSPA) framework, published by WHO in 2022, advocates the notion that health systems have four functions: resource generation, governance, service delivery and financing. The service delivery function is further divided into three broad sub-functions: public health, primary care and specialist care. In particular, public health is acknowledged for having diverse definitions and complex conceptualizations due to the possibility of including (or excluding) a large number of activities and fields of concern and expertise.

We adopt instead the brief definition as described in previous chapters where public health is essentially defined as working with the binary distinction of *at risk/not at risk*, attributable to any given collectivity in relation to health and health services. Risk is measurable and attributed to an identified collectivity, which is denominated as *exposed* to (or affected by, or subject to) health risk factors of some type or recognizable determinants of ill health or constraints in the provision of healthcare. Obviously public health communications find connectivity on the *at-risk* side of the distinction, whereby all communications are concerned with prevention and preventive measures. Populations deemed *not at risk* are not of much interest for public health interventions, except in few cases where the aim is to keep the collective in the *not-at-risk* position. In this sense, and we will address this in more depth later in this chapter, all public health indicators become instrumental and orientated to distinguish the *at-risk* segments in the population. Particularly in the concluding remarks of this chapter we have an important discussion about the concept of *risk*.

In the conceptualization presented in the introduction of the book, as opposed to public health and its collective focus, health service delivery concentrates on the individual patient, the object of attention (observed, diagnosed and treated) about which the health service delivery sub-system – that is, medicine – in the broad sense we adopted, deploys its basic binary distinction *healthy/sick* and all range of unfolding codes applied to the individual considered sick (the connective side of the distinction).

Recalling what was previously said; we call medicine "health service delivery", seeing medicine as the broader category of healthcare for individuals, even when there are concerns encompassing other individuals or communities regarding the disease factor faced by the individuals. In this sense, we can say that vaccination of a child is a medical concern when the general health status of the individual child is assessed and the condition for receiving the vaccine or not is evaluated together with possible risks of reactions and side effects. However, as a programme, vaccination is clearly a public health intervention aiming at reducing the risks of specific disease in a given population.

We acknowledge that we are taking a certain degree of liberty in using the term medicine with such connotations. It does carry some potential for controversies and perhaps immediate rejection of the arguments before careful reflection. But it can be accepted that any profession now involved in taking care of individual patients (nursing, physiotherapy, psychotherapy, dentistry, and so on), despite the fact that in most cases doctors would not perform the specific tasks carried out by the respective professionals, exists within the macro frame of medical endorsement and acknowledgement of the technical value and appropriateness of the delivered treatments. In this sense, they all inhabit the semantic space created by the healthy/sick binary code, which historically developed and became the prerogative of medicine. Those distinct professional fields are clearly identified with their participation in the all-inclusive realm of healthcare deliverers of therapeutic value in the health system's sub-system providing healthcare to individual patients.

In consequence of that, the two sub-systems of a health system are distinguished based on the dichotomy *individual/collective*, where the

healthy/sick distinction applies to the *individual* side of the distinction (sub-system of medicine), while the distinction *at risk/not at risk* applies to the *collective* side (sub-system of public health). With this conceptual framework it is possible to classify most actions and communications taking place within a health system as public health or healthcare service delivery (medicine, in the broader sense we defined here).

In this conceptualization, while the service delivery sub-system is intensely and continuously involved in addressing the health *problems* of individual patients, the public health system is tackling *problems* and figuring out what the system as a whole is doing and whether it is achieving or not the intended reduction of health risks faced by the population or collectivities it is responsible for.

We can say that both sub-systems deal with *problems*, identified and tackled differently by each sub-system, independently from the interfaces and connections they have with each other. Accepting that *problems* are at the heart of the concerns of these two sub-systems, their respective undertaking can be seen in a new light. Medicine (service delivery subsystem) facing *problems* of evaluating, diagnosing and treating patients, and public health addressing the *problems* of population health, assessing the distribution of risks and healthcare services to the collectivities.

We can ask a general question in relation to public health in the following terms: "why and how does a health system elect some issues/topics and problematize them, giving them a form of problem to be tackled, rather than others?" A first tentative approach to answering this question says that *problems* are formulated with attention to a variety of issues including:

- Preserving what has already been achieved – a continuous source of problems – formulated as how to guarantee the provision of the same sets of services, with the same quality, with the same technology, with the same coverage, and so on.
- Answering concerns and interests continuously voiced by external stakeholders (communities, media, politicians, donors, international organizations, religious groups and suchlike), requiring services and improvements in the health of the society.

- Addressing evaluations of needs and conditions for service delivery, raised internally in the system or by external stakeholders.
- Answering unexpected global, regional or local health threats or emergencies.
- Aiming at the expansion of services with the incorporation of new technologies and procedures not yet implemented.

In the process of answering questions from these areas of concern, the system creates internal divisions specialized in finding and formulating the *problems*[2] and indicating the solutions. These divisions are staffed with qualified professionals.

The public health sub-system has a number of advantages in operating with what we call the *form* of *problems*, specifying what needs to be solved and how. The "form of problem" is a linguistic "device" that allows far-reaching standardized communications of what has been identified as undesirable and/or as something that needs to be addressed and improved.[3] It can then be possible to communicate: the object of concerns and why; how it can be identified and tackled; and which interventions and actions are expected to bring about favourable changes and how they can be implemented. An identified problem therefore already points at, or at least suggests, how it can be solved. The advantages of working with the *form* of *problems* are thus:

- It strategically reduces the complexities of the full range of issues and options the system would otherwise have to consider, therefore focusing the attention on priorities and specific targets and courses of actions.

2 We may say that "choosing a problem" is rarely problematized, except in small technical groups making decisions about which problem to tackle. For the system's narratives to internal and external audiences, the hesitations about the problem to be addressed are kept as an internal matter of the small group of experts.

3 "How Is Social Order Possible", Chapter 2 of Luhmann's (2023) book *The Making of Meaning*, has a valuable explanation on the form of problems in the context of social science.

- It allows the formulation of solutions drawing from already known strategies, narrowing down the concerns.
- It facilitates justifications and connectivity of resources and operations.
- It facilitates the visualization of the mission and aims of the system, and therefore contributes to the system's self-reference and identity.
- It facilitates the flow of communications, establishing validated semantics and relevant channels.
- It offers justification to exclude, leave unaddressed or ignore concerns that cannot be related to the identified problems.
- It stimulates considerations and communications within the premises and limits created with the specified problem.
- It removes attention from the question "what is the problem?" to an already chosen problem; by doing that, instead of facing the indeterminacy of unlimited problematic issues, the attention is directed to focal points.
- It suggests a politically convenient impression that a problem's causes have been identified and are already consigned to the past as the system moves towards solving the problem in the near future using the adopted strategies.

In the public health sub-system the formulation of problems requires the use of indicators, as indicators translate the elements considered into specific and standardized formats. Indicators can also be associated with potential interventions, and consequently the assessment of the results of the interventions. This chapter specifically pays attention to the public health sub-system's problems and the role of indicators in the construction of those problems, by which the segments of the population considered to be *at risk* are identified and targeted.

Meanings and complexities: what the theory tells us

Before getting into discussion of the indicators themselves, we need to highlight a theoretical point already briefly mentioned in Chapter 2, but

still in need of complementary explanations. For the formulation and use of indicators, a social system based on communications, like the health system, needs to select meanings by which it can build indicators. Indicators consist of paradigmatic synthetic meanings, expressing standard relations between selected elements, which can then be applied to any context, allowing comparisons across contexts and time. Indicators are developed using methodological prescriptions to be strictly followed in order to guarantee reliability, acceptability and comparability. Most, but not all indicators are numerical and require calculations.

The Social Systems Theory tells us that "meanings are complexities". To be clear, as has already been explained in other chapters of this book, any meaning is the result of a process of selection among many possible alternative meanings, of which only one is adopted and the others are therefore discarded (Luhmann, 2022). There is no meaning without the "shadow" background of countless possibilities, which eventually have to retreat to the background to allow the visibility of the chosen meaning at the forefront. Meaning is the form distinguishable from the backdrop of noise of other possibilities. The complexity we refer to is therefore the set of possibilities around the selected one.

Emerging from the contextual complexity, the formulation and communication of a meaning requires the often-difficult tasks of addressing countless connotations, misrepresentation and nuances that the used meanings bring about or can be connected with. Such "cleaning" operations have to be carried out communicatively as both sides of the communication link need to be certain what they are talking about.

However, the "cleaning" cannot be exhaustive or call for full awareness of what has been excluded, because the complexity involved is vast and may never be apprehended in its entirety. In the meaning-construction process, the selected out meanings have to remain unacknowledged (latent) to make the communication possible, even if not perfect. Such difficulties are unavoidable and are inherent to any communication-based system operating with meanings.

Furthermore, any communication-based system observing its environment needs to reduce the complexities of the variables involved by focusing on elements and relations it can work with and their respective

meanings. The system creates simplified models of what it observes in its environment in order to be able to meaningfully approach it. Meanings are the "raw material" of models.

So, we have the complexities inherent to meaning construction, as explained above, and the complexities of the environment, with its surplus of possible meaningful elements and relations. To be meaningful, something needs to be separated and distinguished from the complexity it is immersed in. The complexities from where the meaning has to distinctly emerge are thus reduced. The complexities of elements and relations, which are abundant in the environment, are also reduced with the focus brought about by the selected meaning.

These theoretical references should help us in the discussions to be held next.

Indicators: anatomical features of public health

In this section we present the most common public health indicators in the literature and official documents of international institutions such as the WHO, development banks, aid donors and countries' Ministries of Health or other authorities.

We briefly present the *technical structures* of each indicator as explained in textbooks and the literature on the topic (for instance WHO, 2010b). We also discuss how each of them is predominantly orientated to inform health systems' *internal* or *external* narratives. We discuss their *inherent complexities* and how the systems usually deal with them. We also discuss the form of problems and solutions the indicators help to elaborate and suggest.

Not all definitions presented here are exempted from controversies; however, it is not the purpose of this chapter to address those polemics or propose final solutions for classification or applications of indicators. Despite possible disputes about accuracy, for the purpose of discussing the functional systemic meaning of indicators, uncontroversial definitions are not paramount; we are negotiating a conceptual territory with many contingencies. We believe this becomes clear in the chapter. Rather

than precise textbook definitions and classifications of indicators, for our objective of illustrating the use of indicators as building blocks of self-references the presentations we make here are sufficient.

The fact that public health indicators are orientated according to the fundamental key binary distinction of public health – *at risk*/*not at risk* – is important for grasping our message. The usefulness of the indicators consists basically in allowing the identification of populations that can be considered *at risk* and therefore targeted to benefit from risk reduction interventions. Those segments become the targets and the justifications for any intervention based on the respective indicators. We make brief comments related to the issues of *risk* in each indicator in the next section.

However we can anticipate that each indicator has peculiarities in the way it approaches risks; they are not the same in the sense that each indicator deals with different forms of observing risks. For instance, briefly, an indicator concerned with reduction of risks of specific diseases is useful for moving people from the at-risk category to the not-at-risk one; meanwhile, aggregated measurements of the burden of diseases, without detailed specifications, comprehend the whole society as being at risk (more on this later). We hope this will also become clear in the subsequent sections. To facilitate the follow-up of the risk topic, we separate the corresponding paragraph within the discussion of each respective indicator.

Finally, for presentation purpose, we group the indicators in three categories: (A) indicators directly linked to population's health needs and risks factors as the health system's object of concern; (B) indicators linked to means for achieving the necessary conditions for healthcare provision; (C) indicators predominantly of the interests of observers outside the health system (only circumstantially linked to system's internal decisions).

(A) Indicators directly concerned with population's health needs and risks

Efficacy: This expresses the causal link between specific action and the attainment of the intended specific objective. It can be used, for instance, to indicate the rate of success of a specific drug, technique, method, strategy or intervention (including protocols, clinical guidelines, standard procedures) for achieving specific outcomes. In general terms, the more specific is the action and the intended effect, the stronger the indicator. With that, this indicator can support managerial decisions for procurement, resource allocation, and so on. This indicator can also be relevant at the micro level of treatment decisions, including establishment of protocols and training programmes. An example of the kind of problem formulated with this indicator is: "Which antibiotics have better efficacy for this type of infection?" The indicator tries to reduce the complexities related to uncertainty regarding treatment results, and the contingencies of the circumstances surrounding the treatment. Some complexities may be ignored but not eliminated; Organisms' diverse reactions and uncertain effects of possible interactions with undetermined or overlooked factors (except in rigorous controlled trials) may remain latent. However, such shortcomings do not prevent the use of the indicator for clinical decision-making and establishment of standards of service delivery. Fundamentally, this indicator is of great interest for internal operations of the system but less so for external audiences.

Risk: In relation to populations *at risk*, typically this indicator can be instrumental for identifying populations to receive benefits from the strategies according to their verified efficacy. Efficacy is therefore a valid argument demonstrating the number of potential beneficiaries among those currently *at risk*. Beneficiaries might then move from the *at-risk* to the *not-at-risk* category for the specific disease for which the efficacy of the treatment/intervention has been determined.

Performance: In simple terms, performance is about actions and achievements of predefined goals measured along scales of quantitative and/or qualitative values, observed in a broader (for the whole system) or narrow (for sub-systems or system components) sense. Performance can also be analysed considering relations between means and goals both often broadly defined. Performance indicators tell us about how far the system got towards the targets, outcomes and aims it set for itself, and help to formulate problems about the technical composition and missing elements. Performance also may indicate the system's responsiveness to identified challenges, combined with measurements of outcomes achieved; in this regard, it is also used with the intention to report and justify to external audiences what the system does. Other key terms for complementary assessment of performance are adherence and compliance with standards, and suitable combinations of means and goals. It can have very broad and comprehensive inclusions of other indicators, as the WHO's (2022) recently published HSPA shows, using for example epidemiological indicators as measurement of outcomes in relation to the actions implemented.

Risk: Public health validation of performance indicators requires estimations of benefits obtained, possibly expressed in terms of reduction of health risks by: (a) reducing the chances of those at risk of getting sick; (b) increasing the chances of moving from the at-risk to the not-at-risk category; or (c) achieving assurances that those not at risk will not in the future lapse into the at-risk category. Although difficult to precisely translate performance measurements into those categories, given the number of variables involved, the judgement of performance may involve comparisons with interventions and outcomes from elsewhere, showing increases and improvements in performances (measured as number of activities and/or quality), usually acceptable as a *proxy* indication that health risks seem to have been addressed and reduced.

Efficiency: This is measurement of productivity, comparing outputs (such as services provided) with the inputs used (sometimes expressed as costs) to obtain them. The causality relation between inputs and

outputs does not need to be strict or exclusive but does need to be plausible. It is standard for comparisons across different undertakings with similar objectives. Usually, the assessments are conducted with internally generated data, even when carried out by external agencies, and aim at making the system more productive with the same inputs (more efficient). This indicator has inherent uncertainties related to the estimation of inputs (and their cost expressions) and the decisions on what to include or leave out of the estimations. Besides that, there are uncertainties about the relations between the factors and the probabilities of obtaining the desired results; the causality links may not have more than suggestive plausibility. For instance, the comparison between facilities of similar structure, delivering the same standardized package of healthcare: "which one of those facilities, having the same human resources, presents the lowest ratio of services delivered by cost?" While narrowing down the observed variables may make the assessment easier, complexities are only left out but not eliminated or neutralized. The problems this indicator tries to solve, particularly those in relation to how macro decisions should be taken, are often occurring in policy contexts where conflicting agendas may force the inclusion or exclusion of variables and measurement details (for instance, services produced regardless of quality). Almost always, results can be questioned due to the contingencies of the choices made while formulating the indicator, as the complexities involved can hardly be fully apprehended. Efficiency is measured in accordance with the model employed to summarize the variables involved; but model construction requires reduction of complexities by eliminating or ignoring some variables. Nevertheless, this indicator is useful when optimization is high on the agenda of a system's concerns.

Risk: In itself, the efficiency indicator has economic rationale. To say that, for instance, a system is operating more efficiently because it delivers the "same quantity of services" while spending less doesn't carry public health meanings. It is necessary to clarify whether the "same quantity of services" corresponds to actual needs and, even better, the actual services represent reduction of "risks". Similarly to the

performance indicator, it is hard to gather evidence of such reductions; therefore the indicator may work as a "surrogate", suggesting that risks may have gone down. Being more efficient implies saving resources that can then be used to tackle health risks not addressed previously.

Effectiveness: This indicator tells us whether specific deployed strategies reach expected standardized comparable results in real contexts, without ruling out interference of known and unknown as well as controlled and uncontrolled variables. The results of the strategies may be the same outputs/outcomes, or may be different, but they should be measured in comparable units such as saved lives, averted deaths, prevented cases, speed of recovery, disability adjusted life years (DALYs) averted, quality-adjusted life years (QUALYs) gained. The comparison between strategies can be expressed in terms of their respective costs to obtain a unit of the desired comparable result. Usually the construction of such indicators requires trials in real settings and research projects conducted by external agents (such as academic institutions, drug manufacturers). An example is: "What is the cost-effectiveness of a fifth COVID-19 vaccination dose?" In other words, these indicators are concerned with specific results achieved, possibly considering the resources consumed in the process. A particular case of these indicators is cost–benefit, in which monetary values are on both sides of the equation – that is, the cost of the strategies and the benefits obtained where benefits, however they may be defined, are expressed in monetary terms. In these indicators, the complexities left out include the indeterminate factors that may affect trials and causal links, even after cautious randomizations of variables, and all that concerns the plausibility of assumed causalities, and the comparison with estimations made in different settings. They are often relevant for internal decision-making related to selecting alternative courses of action to obtain a well-defined result. External audiences may also have economic, political and other interests in these indicators.

Risk: Effectiveness is perhaps the indicator par excellence for selection between strategies aiming at risk reduction. It draws clarity and precision from the controlled-trial nature of the procedures to estimate effec-

tiveness. It offers a valuable indication of what risk reduction can be expected by adopting the admittedly most cost-effective of the compared strategies. We may say, though, despite being the public health golden standard for estimation of risk reduction (Jamison et al., 2006), effectiveness has a demanding format, where validity depends on fulfilling the strict requirements of the measurements. The indicator indeed measures risk reduction in one of those previously mentioned possible formats: (a) reduction of chances to move from at risk to sickness; (b) increase in chances of moving from at risk to not at risk; or (c) reduction of chances of moving from not at risk to at risk. However, effectiveness refers only to the considered strategies; therefore, the use of this indicator as "proxy" or "surrogate" for broader multiple health targets and multiple operations for risk reduction, as the previously mentioned indicators (performance and efficiency) address, is not acceptable. Gains in specificity unfortunately correspond to losses in generality; this indicator wins when specificity (targeted intervention) is more relevant than generality (multiple objectives).

Equity: This indicator compares the distribution of availability or results of healthcare services and goods with the distribution of needs and other characteristics of population groups. More specifically, it assesses the distribution of opportunities to access and reap benefits from services across populations divided into defined categories. Population categories can be constructed according to many variables: for example, ethnicity, income, education level, gender, age group, residence, occupation. The indicator can also assess the existence of barriers preventing some population categories from having access to and benefiting from certain services. They can be estimated with multiple different aims in mind, such as equity of distribution of: life expectancies; access to specific or general care; delays to access healthcare; access to quality care; outcomes of treatments; survival rates after treatments. They are often estimated by external agencies with specific political purpose. Different estimations may reveal different types of equity/inequity, even considering the same population groups – equitable distribution of basic services together with inequitable outcomes, for example. To com-

plicate matters, scarcity of resources often means that access for some is at the cost of access for others. Complexities arise from the multiplicity of perspectives and from the inclusion of time variations. Setting aside unaddressed complexities, equity indicators may be amenable to formulations shedding a favourable or unfavourable light according to the intention of the studies. But still, these indicators are of interest both for those praising the system's achievements and for those criticizing them in arguing about fairness and distributive justice.

Risk: The validity of an equity indicator derives from the clues it can reasonably give. Ideally, equity would be estimations of fairness in the distribution of risks and risk reduction interventions across populations seen to be *at risk*. However, these estimations of fairness for diseases, treatments, prevention measures and so on are hampered by enormous difficulties. If, for example, access to healthcare, quality of treatments, quality of outcomes, are unevenly distributed across populations, following clearly demarked differences according to income, ethnicity, education, the indicator provides enough suggestive evidence that some populations are disadvantaged in comparison to others, and therefore some are likely to be more *at risk* than others. The corrective measures should consequently be justifiable. In assessing equity, considering risk and risk reduction effects, these possible formats are also relevant: (a) reducing chances to move from at risk to sickness; (b) increasing chances of moving from at risk to not at risk; or (c) reducing the chances of moving from not at risk to at risk. For that matter, the more specific the interventions and respective measurements, and the distinctiveness of the populations being considered, the greater the validity of the equity judgement.

Coverage: This indicator is intended to point to how key inputs and service provision capacities (such as human resources, healthcare facilities, laboratories, complex healthcare services) are distributed throughout the country and located according to population or population needs (often in association with demographic and actual services production). This indicator can take into account a number of features of the struc-

tures of healthcare provision, such as the existence of organized referral networks, the integration and continuity of services from the lowest to the highest complexity levels and the ease/difficulty of accessing the facilities/services. This indicator may complement but may also clash with the equity ones, when certain uneven distribution of services may be justified based on concentration of cases and scarcity of resources, leaving parts of the population with disproportionally lower coverage (of healthcare services, for instance). These indicators are often of internal as well as external interest. They can be used internally for planning and investment decisions, and also provides inputs for equity estimations. They carry uncertainties related to the assumed equivalence of the officially recorded or directly observed distribution of inputs, assuming also uniformity across inputs (presuming same efficacy, same effectiveness, same efficiency, and so on). Besides that, they may be estimated in ways that make possibly large variations in quality of care invisible. Of many possible descriptions of coverage, the intended narratives have to select those suitable for the concerned audiences. Consequently, the complexities inherent in having different possible descriptions are left out of consideration. Although of high political relevance, these indicators carry vulnerabilities, open to contestations given the difficulties of solving trade-offs between high/low coverage of expensive/cheap services in contrast with presumed or measured high/low need or utilization.

Risk: Furthermore, the translation of this indicator in terms of *risk* reduction – that is, aiming at or assessing whether coverage reduces risks, is highly complex and difficult. Coverage means the services are available and reachable at the point of need (or when prevention is necessary). However, given the unpredictability of needs, blindly pursuing coverage availability would imply maintaining services regardless of actual needs, in order to ensure the service will be there if it is needed. This obviously is not feasible in a context of scarce resources, which implies that sometimes some services may not be available at all. The differences between coverage aims and actual availability at the point of need create insurmountable practical problems. Coverage, in its broader general sense, does not strictly discriminate between those *at risk* and those *not at risk*

in the population, given the unpredictability of determining and distinguishing one from the other at all times. Therefore, coverage (particularly universal coverage) often becomes coverage for those not at risk as well, because it is uncertain when someone will cross the line of the not-at-risk/at-risk distinction. In this sense, coverage becomes conceptually resistant to approaches trying to distinguish between the necessary risk-reducing coverage from the wasteful coverage with no risk reduction effect. More than resistant, coverage may create an unsolvable problem; it will always fail, either for not being enough or for being wasteful. Anyway, the logic of scarcity of resources eventually leads to a practical solution for resource allocation, which can't be precisely optimized in terms of risk reduction, but certainly and justifiably reduces risks. And coverage may be still valid when the respective services are clearly specified, like coverage – of vaccination, of infectious disease control programmes, of antenatal care, and so on – when *at risk* can be precisely specified as *at risk* of, respectively, not getting the vaccines, the disease not being detected, not accessing antenatal care. In addressing general needs, coverage loses strength as a guiding indicator.

Quality: It is clearly understood that healthcare services can be delivered in accordance with additional attributes that are not included in the technical combination of inputs deployed. The same services can be judged according to quality standards of provision that not only consider the valuation of the outcomes of the treatments from the point of view of curing the disease, but also considering the level of satisfaction and comfort of the patients during and after treatment, measured for instance by quality of life years after treatment. As part of health system ethos, attention is given to assuring that services are made available and operate according to established standards and requirements, including perceptions of users. Professionals are often mobilized with the objective of establishing the standards of care for their respective professional categories, which often consider patients' perceptions. There are motivating factors pushing the election and compliance with standards originating internally in the system as well as externally (insurance companies, government, and so on). Quality of care can become

a contentious issue, as the perspectives of the players are not always aligned. Professionals may be motivated to create protection against the consequences of potential mistakes, unfortunate results, and to avoid subjective criticisms. The scene therefore involves a wide range of complex aspects selectively elected with exclusion of some possibilities that could otherwise create controversies about the indicators. Nevertheless, they are useful for decision-making on specific matters, where they are considered as valid arguments either in favour or against certain solutions, like treatment procedures that are less painful or uncomfortable for the patient even if they have lower efficacy.

Risk: From the point of view of risk reduction, the indicator has mixed outlooks, as the issues of risk reduction and quality of care do not neatly map out one another, and may project different orientations towards the outcome of the services provided. Low quality may still be risk reducing and high quality not necessarily so, particularly when quality emphasizes perception, presence of amenities and elements that do not affect the reduction of risks. In many circumstances when quality of care refers to the skilful and correct deployment of healthcare procedures, risk reducing acquires the connotation of, for instance, avoiding: sequels, unnecessary or excessive use of drugs, side effects and errors. So, while discussing and deciding on how to approach quality, risk reduction effects can be quite relevant. Surely decreasing medical errors is something public health pays attention to as inherent to both quality of healthcare and risk reduction.

Health profile: This is not a distinct indicator per se, but rather a collection of measurements, predominantly of epidemiological nature (such as prevalence, incidence, morbidity, mortality, fatality) portraying the health status of a given population, with particular emphasis on measurements such as: infant mortality, maternal mortality, under five mortality, morbidity due to infectious diseases or non-communicable diseases, and so on. Indicators such as life expectancy at birth, general mortality rate among others reveal the overall health status of a society, reflecting all possible causes, where health services may be only one among

many other determinant or contributing factors. These indicators are usually processed from data collected by civil registration institutions in charge of producing vital statistics, often independently from the health system. The health system cannot alone take on the task of improving general societal indicators such as life expectancy. However, these are often used to indicate progress, and the existence of problems on which the health system may need to focus. Specific components of the health profile can constitute clear targets closely related to access to health services (preventive and curative alike), such as reduction of infant mortality and maternal mortality.

Risk: We may say that the elaboration of health profiles is definitely an important process for risk reduction estimations. Profiling is in fact part of the approaches and methods for verifying risk reduction and therefore setting targets for additional reductions. A specific indicator in health profiles is the burden of diseases. Given its relevance, we dedicate a few lines to it here. Together with effectiveness, burden of disease is a particularly valuable public health indicator. Let's take the DALY (Global Burden of Disease (GBD) estimations by Institute for Health Metrics and Evaluation), representing the loss of healthy life years due to illness (calculated for sets of diseases or as an overall figure for a given country or groups of countries) – avoided DALYs over time can be interpreted as indicating risk reduction. If DALYs is used in a cost-effectiveness study (for comparing estimations of outcomes of two strategies for tackling neonatal and maternal diseases for example), it should allow reliable conclusions regarding the strategy to be adopted. It offers the reduction of the burden expected from the winning strategy, which can also be read as aggregated reduction of risks made possible by that strategy. In this sense, reduction of the burden of disease and reduction of risks tells the same story, offering strong arguments for adopting the winning strategy as the one that is expected to yield lower risks in the future, following implementation. The same is not the case when the burden of diseases is compressively estimated as an aggregation for a country as a whole for large sets of diseases, including diverse types of interventions; in such case, links between strategies and risk reduction cannot be established.

In this sense, we can say that in estimations of general burden of diseases, the whole society can be alleged to be at risk, and the burden of diseases cannot be used for estimation (or proxy) of risk or risk reduction. This is because of the three possibilities of risk reduction – (a) reducing the chances to move from at risk to sickness; (b) increasing the chances of moving from at risk to not at risk; or (c) reducing the chances of moving from not at risk to at risk – estimations of the burden of disease only mention the possibility (a) (moving from at-risk to sickness) because the risk in this case is expressed as registered diseases, while (b) and (c) do not include the appearance of actual diseases. In short, public health is not only concerned with diseases and population *at risk*; it is also pays attention to populations *not at risk*.

(B) Indicators linked to means for achieving the necessary conditions for healthcare provision

Financing: These indicators obviously deal with money matters. They may have diverse focus and concerns: for instance, on sources and amounts of funds being brought into the health organization; allocation of funds across services, regions and populations (checking optimization in terms of benefits for example); on the distribution of the funding burden in alignment with needs and utilization; on financial protection and guarantees to the disadvantaged; on cost compatibility with financing profile and justifiable comparison standards; on payment and incentives for healthcare services providers and their efficiency. These indicators may gather information from several sources, depending on how broad the interests are, as addressed for example in the National Health Accounts methodology (Maina and Mwai, 2019). The indicators can be developed internally or externally depending on the objectives and concerned players. Uncertainties may arise where information is not made available by the interested parties or may not be properly accounted for (as it is often the case with informal payments and unrecorded out-of-pocket expenditures). They are often employed with political and business intentions, leaving out the complexities

inherent to the variables and estimations considered unworkable or inappropriate for the purposes in mind.

Risk: Financing can only tentatively suggest risk reduction, when financial resources increase and expectations of healthcare services improvements can be reasonably expected. The assumption that positive changes in financing lead to risk reduction, as comparisons across different financing arrangements suggest, can only be tentative until empirical confirmation is possible. The sets of variables are indeed huge. Financing as a means for acquiring and paying for inputs does not involve direct relationships with the effectiveness, performance or efficiency of those inputs. Productivity can vary immensely for countless factors. While financing, as a macro indicator remains relevant (even vital) for public health self-assessments and narratives about the health system, they may have a "loose relationship" with the reduction of risks. The case of low- and middle-income countries such as Cuba, Costa Rica, Kerala state in India and others showing better health indicators than high-income countries is an example (Balabanova et al., 2013). On the other hand, we can consider that the three possibilities of (a) reducing the chances to move from at risk to sickness; (b) increasing the chances of moving from at risk to not at risk; or (c) reducing the chances of moving from not at risk to at risk, can be influenced by financing at a micro level. Financing can operate as a risk-protection "device", when certain financial arrangements such as health insurance make a patient's access to healthcare possible. In such cases, they may enable the patient to move from sickness to an at-risk or optimally not-at-risk position, becoming countable in the sphere of public health concerns and interventions. Indeed, financial protection may also open doors to preventive care, allowing people to access services, and moving from the at-risk to the not-at-risk position, which could not perhaps be possible without the financial protection. Nevertheless, precise estimations of reduction of risks in relation to changes in financing require sizable work.

Sustainability: These indicators aim to tell us whether the resources generation are sufficient to keep services going and structures in place. Assuredly, depleted hardware should be replaced and manpower preserved and reproduced with self-generated resources. As such indicators rely on predictions for the future, they carry with them all the corresponding time-related uncertainties. The conditions are fluid and the relations between production factors and results have the inherent unpredictability of development of technologies, availability of funds, availability of human resources, timely decision-making, favourable political context, and so on. The level of uncertainty can be so high that any serious requirement for evidence of sustainability as a condition for proceeding with an endeavour might be equivalent to giving up the initiative altogether. Only circumstantially might such indicators become of interest, particularly when funding agencies want assurances about political commitment to the continuity of the investment. In this sense, rather than balancing a business-like equation of revenue-raising potential against cost potential, it becomes a balance between preservation of interests in keeping the initiatives going and the feasibility of covering costs. As we can see, there are a lot of "potentials", and therefore uncertainties. Competing agendas are not in short supply and the emergence of unforeseen threats may make all forecasts rapidly outdated. Sustainability became a must-be-included word in international health narratives, but the carefully circumvented or disguised uncertainties are more like "we do not need to talk about the elephant in the room". In this fashion, complexities have to be left unaddressed and the projects might go ahead based on good faith and hope. As mentioned, sustainability indicators have to deal with questions inherent to the dimension of time. Lives are supposed to be saved or improved today, not later when the initiative will prove to be sustainable. All initiatives can only claim to be sustainable in the future; until then, sustainability is anyone's guess and hope, informed or otherwise. However, the need to improve and save lives should not be deferred. Decisions are therefore taken today regardless of the reliability of predictions of future results. Assurances of provision of inputs such as human resources, equipment, maintenance and finance can never rule out unpredictability, and the streams

of benefits accrued from future services can only be good guesses at best. Unpredictability and complexities cannot be ruled out.

Risk: Sustainability is a criterion in its own terms; sustainability maintains what has been achieved, even without implying risk reduction. The criterion is thus satisfied; in that sense this indicator has a weak public health meaning, however relevant it is in business decisions. Furthermore, the sustainability of a health project and the sustainability of the risk reduction effects of the project are two different questions. A successful project, generating valuable health impacts, which nevertheless is extremely dependent on the funds provided by external donors, obviously does not have a good prospect in terms of future benefits (expected risk reduction) once the flow of funds stops. In this case, the merits of the project should not be judged according to its sustainability rather than the benefits it actually generates. Conversely, a sustainable project (funded by the state, for instance) may not operate at optimum levels in terms of risk reduction. Nevertheless, risk reduction must be the priority criterion, while sustainability thus may remain a less relevant one. It is often the case, however difficult to admit, that a project's merit is judged in terms of the risk reduction it brings about instead of whether it is sustainable or not.

Organization: These sets of indicators reflect characteristics of the organization of the system, considering aspects such as management structures, centralization/decentralization, hierarchies, authority, partnerships (including public–private partnerships – PPP), autonomy, strategic decision-making, integration of services, facilities network, planning, logistics, monitoring and evaluation. These indicators may describe organizational structures and how management is exercised at different operational levels in the system. Portraying the organizations, these indicators also point to how they may or may not change, with the respective links between institutions and agencies (including multisector collaborations). Of relevance for macro planning and large investments, they also concern operations and functionalities of programmes, projects and initiatives articulated at different levels in the system. The

narratives about organizations explain, for internal guidance of the system, how it is organized, but also provide topics for external communication. The presentation of the diversity of organizational features needs to be as precise and specific as possible, therefore complexities should be addressed; organization structures and functionalities need be portrayed as consistent and coherent or otherwise adjustable. Emphasis may be placed on critical matters related to, for instance, how far decentralization is real; how far informal networks inside the organizations are controlling it; how provisional relations of trust, partnerships and leadership are effective; how membership in the organizations is recognized as a source of influence; how far autonomy is respected. These organizational and managerial matters are often described in narratives explaining the structure and functioning of the organizations operating in the system. These indicators are mostly descriptive and often employ social science paradigms and frameworks explaining organizational aspects in terms of networks, power stakeholders, leadership, feedback loops, game theory, and so on. As many organizational structures and functions are of a contingent nature, these indicators address complexities in cross-sectional snapshots of time, which may not reveal the complexities of the dynamics driving the system towards changes. Such changes might be unpredictable, however comprehensive the provisional descriptions of the system may be. In a context of the presence of many different organizations of different types, and diverse public and private providers and funders, the portrayal of the system tends to be limited and incomplete.

Risk: Optimization of organizations is in pursuit of a diversity of aims and objectives. Individually, each organization, such as providers of hospital care, tries to minimize risks for patients in all that concerns the hospital environment and treatments deployed. In terms of our discussions in this chapter, patients are those who crossed the line from at risk to sickness – that is, they crossed the *healthy/sick* distinction and entered the realm of the medical sub-system. As individuals they are not of specific interest for the public health sub-system, although, we should highlight, the collective of patients may again be treated as

a *collective at risk*, because of what can happen to them – for instance, hospital infections, where they are at-risk as far as those infections are concerned, independent of their original ailments. Public health orientation will try to devise strategies to bring them from the *at-risk* position to the *not-at-risk* position while they are in the hospital – a difficult task in some cases. In this sense, public health orientation is not exclusive to Departments of Public Health. The orientation spreads throughout the organizations and structures of the system, wherever approaching *collective risks* is appropriate and in fact necessary. In that, the indicators of organizations may narrate links of organizational attributes/initiatives and risk reduction. However, Departments of Public Health (or similar) have the responsibility to deal with risk reduction in society at large, wherever people *at risk* may be, including those collectives under organizations' initiatives but also all those risks and risk factors present in society that are not within the remit of any specific organization. As mentioned in Chapter 2, the Social Systems Theory sees organizations as one of the three types of social system based on communications. Perhaps the reader can refresh their understanding of organizations by checking the glossary at the end of the book. Organizations, including any involved with healthcare services provision, are primarily orientated to their survival and the reproduction of their distinct communications among their recognized members. Tackling risks may become part of the survival strategy of the organizations, but in these cases the focus is narrowed down to the scope of actions of the organization (it does not go beyond its boundaries). Health organizations aim at reducing risks for those receiving the services the organizations provide; the operations of the organizations are exclusively concerned with the service they provide and those directly receiving them. Nevertheless, while being organizations themselves, the assignment and mission (self-reference and identity) of the Departments of Public Health include identifying and tackling all sorts of health risks in the society and reducing them. These departments perhaps are the only organizations for which reproduction and survival are dependent on effectively reducing society's risks; no other organization in the system has such broad scope. The self-reference narratives of the health systems should therefore include

accounts of how, when and whether their Public Health Departments are fulfilling their risk assessment and risk reduction mission. This mission surely includes accessing whether the organizations operating inside the health system are adopting organizational structures that optimize their risk reduction possibilities.

Health workforce: We have already mentioned human resources in relation to indicators of coverage, denoting the availability of health professionals according to population size and needs. The workforce theme is given prominence in the literature and international organization as evidently expressed in WHO frameworks, where the workforce is a fundamental pillar of health systems. Besides coverage, we understand that workforce can also represent the aggregate composition of the human resources operating the health system and how they are distributed throughout the structures comprising it, including public and private organizations and institutions. Health workforce assessments also indicate possible gaps and possible excesses and the capacity of the system to generate and absorb more human resources wherever they may originate. It may highlight serious limitations and bottlenecks of the system, and this means that the complexities that can possibly be addressed while scrutinizing workforce structure, composition and dynamics are vast and need to be narrowed down. For that matter, the portrayal of workforce problems should be in line with anticipated solutions. Complaining about insufficient neurosurgeons in a country that cannot cover the needs of general practitioners in primary healthcare facilities would seem unacceptable.

Risk: Risk reduction has close links with workforce; the workforce makes risk reduction possible. Therefore, the understanding of workforce should be part of public health assessment of risks and risk factors reduction. Health professionals involved in prevention and preventive care, in curative service delivery and public health programmes, should understand the scope and amplitude of risk reduction objectives, incorporating it into the professional ethos. The public health narratives can therefore explicitly link workforce composition and distribution

with risk reduction aims, without resorting to specifically and precisely quantifying the magnitude of the reduction achieved or expected.

(C) Indicators mainly of interest to observers outside the health system

Governance: These indicators bring to the fore whether the provision of care and the functioning of the health system correspond to normative principles of legally established rules and rights, as well as to judgements of standards of public services operations, including ethics. The indicators carry strong political interests as they closely relate to political objectives and may be used as a demonstration and justification of politically motivated actions. In its elusiveness, governance resembles more a medium than a form. In the medium of governance – a semantic medium, we can say – in the last decades, a plethora of forms have taken shape, appearing in the academic literature and reports of international institutions such as the WHO. Very much a characteristic of a medium, the term governance is always presented with the addition of the form-shaping adjectives and specifications giving visibility to it. As an example of forms used to explain governance, we can list rule of law, accountability, strategic policy formulation and vision, effective oversight, coalition-building, correspondence with legislation and regulation, effective implementation of regulation and legislation, hierarchies, networks, stakeholders' voice, protection of the vulnerable, information and intelligence, institutional design and efficiency, patient centredness, political economy, transparency, multi-sector collaboration, social participation, awareness raising, responsiveness, institutional memory. The list is indeed very long and still growing. These forms often overlap but still each brings specific nuances. Obviously, selections are made according to the preferences of the context as the full range of meanings related to governance become impossible to address in empirical terms. Such a collection of terminologies and meanings, from our point of view, corroborates the understanding of governance as a medium with many potential forms, instead of a specific tangible, singularly recognizable object. The term and its forms are often of external interest for substan-

tiating political judgements, political platforms of actions or critical monitoring of the system. Besides that, these indicators may reflect the overall structuring and functioning of the government across the public sector, closely reflecting the type of political regime in place, by which all sectors are also affected. The concepts are rarely of the exclusive internal interests of the system. The terms are mostly of particular interest for those who want to portray a certain view of the health system in focus. Given the diversity of views the concepts introduce, the complexities are conveniently left outside the debates, remaining unsolved. It is always possible to pick up the forms with connotations to fit the purposes, while the meanings left out might be activated later on, if there is the intention. The assumption of clarity and enforceability of governance indicators in practice is overburdened by the complexities that have been added to it in the last decade. It is possible to say that the more medium-like an indicator is, the more prone it is to attract proposals of new candidate forms.

Risk: The links of the governance indicator with risk reduction are complex and difficult to trace. Indicators situated at the convergence of management, political and legal science have rationales that do not require precise links with reduction of risks and risk factors. The justifications for such indicators are raised for advantages seen from the point of view of the science where they originate. Their relevance in terms of risk reduction are thus inferred and assumed without precise verification or validation. Public health literature shows an "invasion" of such frameworks (of which system science is in a peculiar way also a part) without revealing whether or not risks are effectively being reduced. While agreeing to distance itself from its core business by becoming involved in such debates, public health may weaken the references to what it is really about: risk reduction. Social Systems Theory, however, by bringing public health self-reflection and self-descriptions to the fore, also brings the signalling roles by which public health should recognize its differences from public management in general. The *at-risk*/*not-at-risk* binary code must be kept as the core identification of the public health sub-system and its fundamental reference for self-reflection.

Voice and accountability: Usually also under the wide umbrella of indicators of good governance, these indicators more specifically try to observe whether policy-making, policy implementation and management allow for participation and voice, particularly of the disadvantaged supposed to benefit from the system's activities. These indicators are often external and used for formulating the degree of attention the system gives to those who are served or supposed to be served by it, and how integrated they are in evaluations and decision-making. They are of political interest and require independent assessments, which often can be criticized in terms of bias and distortions, perceived therefore as of limited scope and intended for specific audiences. Where they are of general attribution (reporting whether or not the system as a whole is accountable in its institutional mechanisms), they may be imprecise and used for expressing overall criticism with little practical implications, given the high complexity of health systems operations and the closure of technical communications. As is usually the case for many indicators, complexities are left out but not eliminated. The indicators have to deal with the indeterminacy of motivators and the attribution of links between factors/effects and responsibilities. The term accountability, however, can also refer to the internal processes of communications where the system and its components are internally accountable in daily operations not objects of external attention or scrutiny.

Risk: Where voice and accountability are criteria for judging a system's performance in terms of expected risk reduction, communications between the operators of the system and those receiving or entitled to receive healthcare services is complex. Expectations on the receiving end may not be fully in line with the plausibility and order of priorities as seen from the provision side. Risk, risk reduction, risk factors and the causalities involved are often expressed in technical terms which recipients of services may find difficult to grasp. The evolution of the health systems, as discussed throughout this book, required communications fencing the semantic space and therefore setting rigorous limits for participation. Medical language, for instance, became too highly complex for anybody other than doctors to use and understand. Therefore, ac-

countability may end up happening in accordance with the meanings the system itself develops and adopts for that particular purpose. As said in this book, public health has the ability to explain the system both internally and externally, to the other systems and organizations. The narratives thus generated need to address issues of risk reduction, though how far the incorporation of service users' and potential users' voices and the accountability procedures involved have any effect on risk reduction is an empirical question to be further determined. Having to be accountable and explain its actions to society make the public health sub-system more engaged in producing such narratives, participating in external communications, but such purposes and efforts are not directly linked to risk reduction effects. Public health engages in risk reduction activities not because it needs to explain what it does; essentially, risk reduction is the self-identity, the reason for existing and reproducing, and thus an essential self-reference of public health.

Environmental impact: These indicators scrutinize whether the interventions and services provided by the system have negative consequences, becoming health risk factors (such as pollution, contamination of the environment with hazardous substances or organic materials, destruction of natural resources), damaging the physical environment where the system operates, or contributing to the occurrence of disease. These are mostly external indicators maintained by environmental agencies in connection with politically established rules with environment protection objectives. Among the complexities omitted but not eliminated that can be mentioned are the imprecision in determining environmental effects and their consequences and the time scales required, with problems of consistency over long-term follow-up or retrospective scrutiny. Where rules are well defined, these indicators consist basically in observing whether the system and its organizations correctly adopt the necessary procedures to comply with rules of environmental protection.

Risk: As an evaluation criterion on its own, environmental impact is concerned with avoidance of new risks. In this sense, it seeks assurances that the populations in the environment of the healthcare operations, who

are considered *not at risk*, will remain as such, instead of moving to the *at-risk* category or, worse, to the *sick* category, due to negative impacts of health facilities in their physical environment.

Gender awareness: These indicators are concerned with assessing whether the system observes and acknowledges needs and services according to gender considerations. External stakeholders with specific agendas on gender issues, and capacity for making surveys and disseminating information, are the ones often investing in them. Internally the system deploys them usually as reaction, directing internal attention when the system observes pressures coming from its social and political environment. Among the complexities which this indicator may try to avoid without eliminating them, it is worth mentioning cultural factors in which the system itself is immersed, influencing gendered perceptions, observations and communications. The formulation and use of these indicators reflect the overall interest and attention of society, and particularly vocal key stakeholders on gender issues.

Risk: These indicators have become relevant, as awareness of gender issues has been demonstrated to have influence over decision-making concerning many aspects of service provision affecting risk reduction aims. As a type of indicator within the typical effectiveness measurement, focusing on specific genders and respective needs, the indicator can provide clear orientation for risk reduction. In more practical terms, awareness of the needs of privacy and due respect for pregnant women during delivery or antenatal care examinations, for example, are relevant not only because of quality of care but also effectiveness, as better-standard services are more likely to achieve patients' adherence and compliance.

Discussion

Before we start the discussion, it is important the reader realizes that this is not a textbook of indicators or a manual for fieldwork where indi-

cators need to be measured. This is a reflection on health systems' self-reference. Self-reference is self-observation – observation of how the system constitutes and deploys the meanings that are its prerogatives. We also indicate to the reader that in the concluding remarks section in the end of this chapter there is a brief subsection on risk and risk reduction, summarizing the topic.

Broadly speaking, in line with the Social Systems Theory, indicators are standardized narratives made and used by health systems for self-observation and self-description. Self-description includes all sorts of plans, projections, evaluations, reports, statistics, and analysis and so on about the system. The system needs the internally made narratives as decisive tools for its self-reference and reproduction. However, externally elaborated indicators are also relevant and are acknowledged as such, particularly where the system couples with other systems elaborating and making use of those indicators. But, in short, indicators are schemas for communications of condensed representations of a system's features, operations and attributes of interest.

Excellent examples of such narratives can be seen in the publications produced by the European Observatory of Health Systems in Transition. The publications have thorough assessments for each country case, with consistent interpretations of possible system problems to be addressed. The countries' health systems can find rich descriptions about themselves and use them in the construction of their own narratives. As these publications are elaborated with academic rigour, the countries in focus may need to select the themes and matters they can actually address and solve. The complexities portrayed in those publications constitute a potential hindrance for the use of the studies by countries with limited public health skills and resources. Countries do need to construct their own narratives about their respective systems, and those narratives often have to be deployed in political arenas where comprehensiveness needs to be reduced and complexity "tamed" (reduced).

When a system elaborates its own narratives, it is also concerned with communicating how other systems should see it. A health system orientated to exchanges with science systems may construct its own interpretations of data and complexities of the environment where it op-

erates. The system of politics may use indicators for political purposes and the economic system (for example, insurance organizations) may use health indicators for their business decisions. However, the health system decides whether or not and how far the indicators of external interests are relevant for it.

In general, indicators are constructed using combined efforts to reproduce meanings, explaining and justifying what the system is about and the results it achieves. They refer to the internal and external environments of the system, where its actions are expected to attain certain effects. In this fashion, indicators give form to the *problems* the health system tackles or intends to address (more on this later).

Indicators are employed to measure/describe conditions and selected types of activities, outcomes and levels of performances, in reference to those defined aims of the system. The measurements are related to actions and action conditions as well as effects and their factors and consequences. In this sense, indicators have time dimensions, with measurements of the recent past becoming a reference for future measurements, and are also supposed to indicate the "way ahead", clarifying the range of options the system (at different levels inside it) may consider and decide upon.

Essentially, indicators are based on distinctions, which specify, select and limit the fields of observation, and the elements to be taken into account or otherwise ignored. They become meaningful in their articulation, with self-reference constructions made by the system and therefore taken as guidance for possible actions. Outside these connections, indicators become useless and are forgotten by the system, even when some external stakeholders carry on valuing them.

The indicators therefore perform complexity reduction that is vital for the system. This is one of the important messages of this chapter. Indicators are pre-selected forms that are intended to capture the attention of the audience and make links to subsequent observations. Indicators narrow down the fields of observation to meaningful elements to focus on, and by doing so they reduce the complexities the system has to deal with.

In that way, indicators also signal the boundaries of the system, understood as selected meanings the system considers relevant. Beyond that, everything should remain unspecified. Therefore, indicators stabilize the structures and symbolic boundaries of the system.

At the same time, indicators project the image of the system as a unit, upon and about which the indicators bring to light a unifying vision. For the observing eyes (internal or external), the system to a great extent becomes what the indicators tell about it. We can say that indicators create the profile of the system, making it visible.

Once articulated, with targets and objectives the system sets for itself, indicators become guiding tools, creating expectations and acquiring normative force. The system recursively employs its selected indicators to conduct self-observations and make the corresponding narratives, reporting (or omitting) evaluations of successes or failures of plans and programmes designed with the guidance of the indicators.

Regardless, indicators can be interpreted or reinterpreted (or forgotten) by the system according to the identity the system wants to project. As mentioned above, the system also considers whether or not to pay attention and incorporate in its communications indicators produced by other systems (the science system for instance).

Indicators are thus designed, structured, calculated, used and updated in defined time scales. The system creates the time scale of the usefulness of the information contained in the indicators, the reliability of the observations and the validity of possible temporal comparisons.

There are multiple indicators, as we saw above. As may have become clear, it is not possible to reconcile all indicators, using them simultaneously, making all of them point in the same direction. There are trade-offs, as progress in an indicator may be achieved at the expense of progress in others, and there is great diversity of scope, with indicators focusing on specific operations, while on the other extreme, indicators address comprehensive features of the system. In the overall set of indicators, multiple values are at stake without necessarily having complementary or synergic mutual effects in any given situation or context.

Furthermore, and perhaps more importantly for our discussion, all indicators carry degrees of indeterminacy and *paradoxes*. When used for prescribing solutions at system level, the indeterminacy has to be somehow tackled or omitted. Meaningful indicators require selections of elements to be considered while discarding others. In that, at the same time they highlight certain aspects, and omit or hide others, which however may still return to be considered at later stages by other stakeholders.

The appearance of precision of indicators in fact disguises the indeterminacy. Indeterminacy arises in consequence of several factors: lack of data; imprecision in the data; complexity of the causal relations between the variables involved; gross simplification of factors and effects when comparing indicators across different contexts; pressures to make positive (or negative) reports, and so on. Often a lot remains unsaid or dismissed in order to attempt an effective use of indicators for some specific purpose of system representation.

For instance, indicators may be omitted in order to strengthen a certain specific intended message about the system. The inclusion of the selected out indicators would make the picture less clear. The deployment of complex arguments instead of achieving the desired reduction of complexity makes the messages difficult to understand. Such selections are therefore largely unavoidable. Indicators always have intentional purposes in line with the way the system portrays itself and expects to be recognized internally and externally. The system deals with expectations of how and whether the combined meanings of indicators are going to be scrutinized.

In short, the system needs to select which indicators it will pay attention to, as consideration of the full set of indicators would create complexities beyond the capacity of the system to handle. As already mentioned, the system may also pay intermittent attention to some indicators and concentrate only on a few for planning and policy-making purposes. Other indicators may be disregarded or remain outside the focus, although still valued and used by organizations outside the system, participating for instance in the science or the political systems.

Moreover, contingencies are relevant. Indicators remove contingencies from view, suggesting that no contingencies are at stake or even ex-

ist, and the indicators are expressing the reality of objective facts, pertinent therefore for the orientation of the system. But the choice of indicators is contingent (and can be different) and the measurements and methods can also be different, if other factors are included in the calculations or other perspectives are taken into account. However, the system assures itself that it is using the appropriate sets of indicators, adequately observed, measured and calculated.

A final general point concerns the links between indicators and the decision and actions based on them. Such links are not always straightforward, as is suggested in justifications for policies, plans and programmes. The objectives may need to be consistently pursued over the course of years until results are achieved. Health resources have a great deal of inertia and their allocations cannot be changed easily. Unfulfilled expectations based on indicators may require discrediting the original indicator. Causalities can be reinterpreted in line with new observations and techniques. Indicators therefore may carry the "seeds of doubt" into the future. This gives them a provisional and tentative nature and only temporary validity.

We can say that indicators hide ambiguities, and that is also one of their main roles, because no indicator has a definite answer about what the system is or has to do. Indicators indicate in different directions and it is always possible to find an indicator to recommend another course of action, even the opposite one. The system therefore has to make selections of indicators and keep the other possibilities out of sight. There are contingencies all around; contingencies, a term that often appears in Luhmann's texts, and we have mentioned a number of times in this book, points to the fact that something contingent is neither necessary nor impossible; therefore it can be different.

Concluding remarks

Let's say an international development donor has XX million dollars to donate and/or lend to country Y's health sector. The money is already scheduled for disbursements. What is missing is the formulation of the

problem and activities to solve it. For that, teams of consultants are mobilized to develop preparatory analysis and documents, and among them narratives spelling out in detail the actual problem to be tackled. Then, indicators play a useful role.

Indicators draw the frames and the sketches by which the problems become visible. There are countless potential candidate problems; we can say that there is always an excess of possible problems. But a selection needs to be made, as resources are limited. The selected problem has to fit with previous decisions and policies, and somehow connect to what has been done previously. The indicators must show the way.

The donor will rely on consultants and the public health sub-system of the recipient country to establish the narratives with indicators, spelling out the operations. They know that throughout implementation, indicators will need to be constantly calculated and perhaps re-evaluated and reformulated.

Newer indicators may appear, as public health incorporates concepts and notions from other fields of knowledge. The implementers may see that the complexity increases, as the translation of plans into actions causes unforeseen factors and variables to emerge. Nevertheless, meaningful use of indicators carries recommendations for preserving the originally adopted guidance. The public health sub-system is well aware of the need to keep indicators narrowed down to those the system can handle and will remain relevant for communications with other organizations, particularly in this case the donor.

The public health sub-system is also well aware of the impossibility of combining all indicators; the excessive complexity would overwhelm the endeavour. Indicators hide ambiguities, as we discussed previously. No single indicator has a definite answer about what the system has to do, and they often indicate different directions. It is always possible to find or explore indicators recommending a different course of action. The system therefore selects the indicators it will work with, keeping the others out of sight. As we can see, finding and formulating problems is a problem in itself, and the indicators are part of the problem as well as the solution.

The academic world does not feel such pressure as it operates within the boundaries of the science system. Researchers do not feel the pressure to narrow down or decrease the formulation of new indicators because the academic world has a different approach to complexity. While it is relevant for academic papers to "embrace complexities" and bring new features and their indicators into the descriptions of the realities, the health system implementing programmes, facing constraints of all sorts, goes in the opposite direction; it wants to keep the narrowed-down range of indicators as they have already been incorporated along with the meanings deemed useful.

Furthermore, for the implementer, indicators should, as much as possible, not leave "empty connections" – that is, areas open to controversy. While controversies are welcome in the academic world, for the wealth of possibilities they bring about, they are not well received in the world of business and politics, especially after decisions have been made.

Indeed indicators are indispensable tools for framing problems, constructing solutions and following up implementation. But we need to understand that to put indicators to good use, the system has to reduce complexity, being careful in the selection it makes and the indicators it can rely on. The public health sub-system has to deal with the unavoidable uncertainties of indicators, the ambiguities of their joint indications and the difficulties with external observers to understand the indicators in light of the use made by the system, and not in accordance with their own agendas.

The use of indicators is supposed to be framing and solving problems, as well as eliminating or reducing chances of controversies about what the indicators themselves are. In the political arena, indicators, deemed reliable and truthful, are used in contested topics to assert positions defended by one side against the other, and may steer controversies in political disputes that are welcomed by one side but not much by the other. By choosing indicators, the managers of the health system select those that have highest chance to portray the system in a good light or make it easier for the system to progress, or blame others for

lack of progress. There is nothing in the indicators themselves that can prevent them from being used in such a manner.

International organization such as the WHO put on the table an overwhelming list of indicators to be considered and possibly adopted in the health systems of developing countries. The HSPA framework is an example of that. Indicators from the academic literature and the WHO are gathered. That hardly reflects the selection challenges the countries' health systems face. In the academic world, the formulation of indicators is very often disconnected from the task of concretely using them to formulate problems and monitor progress on the ground. Those who deal with the task of implementing activities over the years, making real measurements and incorporating them into objective problem-solving narratives, clearly see the differences between the two tasks.

Risk

With the more empirically referenced discussions on risks related to the indicators in the previous sections we can now reflect on the concept of risk with theoretical intentions.

Public health departments in Ministries of Health as well as in academic, research institutions and international health organizations maintain under the umbrella of public health numerous approaches and concerns that we can say belong to knowledge and semantic fields that are preserve of the public health domain and orientation towards collective health risks.

On the other hand, we have also highlighted in this chapter that public health has no "property" over conceptual constructs and methodological practices belonging to scientific and technical fields on which public health cannot properly reflect, criticize and change, because it does not have comprehensive command of those fields and their particular semantics and conceptual structures. Nevertheless, imported concepts are often used and become part of public health conversation, sometimes to the detriment of its core business of finding, defining, addressing, observing, measuring and acting on health risks and health-risk factors.

We therefore advocate that public health needs to continuously go back to its "soul", which is to say, go back to the business of dealing with collective health risks and health-risk factors. Such self-reflection is needed and has to be kept as part of the self-reference of the public health sub-system.

With these introductory notes, we can now talk about the notion of *risk* and its specificity. First we talk about Luhmann's conceptual works on *risk* and then we conclude with our "take-home messages".

Luhmann's approach to risk

In his book *Risk: A Sociological Theory* (2008b), Luhmann approaches risk with a different orientation in comparison with our view of risk as chances of negative health events, mostly involuntary, unrelated and independent from any purposeful risk-taking decisions by individuals and communities, eventually finding themselves suffering or under threat of suffering health losses.

From his sociological perspective, Luhmann studied risk as the consequence of and embedded in actors' decisions, where risks are weighted and considered in antecedent decision-making – when for instance enterprises deal with the potential profits or losses of their decisions. On the other hand, we have looked at risk as mainly determined by natural events – that is, the organic development of diseases related to an universe of factors whereby decision plays only indirect, unintended, involuntary or no role at all.

No one gets sick by their own choice, even allowing for self-harm, where subsequent bodily disorders follow their own organic, often unknown causality paths, unsubordinated by and/or indifferent to the original intentions of self-harm. Similarly, some philosophical schools of thought indicate that no one commits suicide knowing what the after death will be like.

Of crucial relevance in his text, Luhmann draws a distinction between *risk* and *danger*, where *risk* implies decisions and the consequences of what actors decided, and in contrast, *danger* comprehends the probabilities of negative events brought about by the environment

(nature especially), over which the systems and individuals have no determination or even awareness.

We could say that Luhmann's concept of danger does not sound appropriate for our concerns with the health thematic. The connotations of the term danger are loaded with the sense of peril and threat linked to meanings of severity of consequences (such as in earthquakes and natural disasters) that are not always the case in healthcare. Danger implies life-threatening conditions. This does make the term unsuitable for deploying in health systems communication to replace the usual employment of the term risk, which tolerates a diversity of degrees and specifications. No one will treat a mild fever in an otherwise healthy teenager as a sign of danger. Health risk terms have already been incorporated for many decades in the public health and epidemiological jargon, and signify anything that may justify getting the attention of a health professional.

In this sense, public health professionals easily accept and understand that being *at risk* is not always a danger nor a matter of choice for the patient or those exposed to risk factors; therefore, it is not at all similar to an entrepreneur who makes "risky" investment decisions with the possibility of negative returns on the invested money.

We can further argue that even a chain smoker, who despite all warnings about the risks of developing lung disease still carries on smoking, thus staying in the *at-risk* position, cannot be compared to a reckless adventurous entrepreneur. The chain smoker can claim to have seen many of his non-smokers friends dying of lung diseases while he is perfectly fine. Although uncertainties exist on the horizons of the entrepreneurs making conscious decisions of risky investments, uncertainties for the smoker are much greater. The smoker faces a lottery with a diverse range of undesirable outcomes, which cannot be selected, predicted or protected against. On the other hand, the entrepreneur faces only a range of possible losses (still with optional protective hedging mechanisms) but also the prospect of the desired gains, within clearer time frames. So the structures of the gambles are very different. They are so different that we can defend our decision not to use the term "risk" with the same meaning as Luhmann does. We believe Luhmann would not be offended as he

acknowledges in his book that the word "*risk*" has many denotations and connotations in different fields of science.

In the health field, as public health works have demonstrated, health risk structures are consequences of many possible factors, hierarchies and temporal sequences of related factors. These include genetic predisposition; temporary vulnerability; circumstantial weakness of the body; lack of prevention; exposure to harmful substances; contact with pathogens (be it from other individuals or vectors); unavailability of diagnosis and treatments; low-quality treatments; lack of response to treatments; comorbidities; mistakes. To these factors, basically involving organic-level causalities, there are several sets of influencing factors with no biological determination but strong influence and conditioning power, making possible and often triggering the organic causalities; this is where social determinants of health appear. The complexities of the risk structures are clearly reflected in the variety of public health indicators we discussed in this chapter, with each indicator aiming at risk and risk reduction possibilities with different horizons and approaches, some addressing organic causalities and others orientated to social determinants.

Luhmann reflects on the position of the observers: the first-order observer as the one who makes the risky decisions, and the second-order observer the one who observes the first one deciding. In relation to health, say the second-order observer picks the distinction *at risk*/*not at risk* for observing, possibly seeing retrospectively how those who developed conditions became at some point *at risk* without acknowledging or choosing it.

From our perspective, also in relation to health, one would prefer to remain *not at risk*, and may have even made efforts with that purpose, but often the negative outcome comes unannounced and by its own causal determinations. Decision-making influences in these processes are often only marginal. The observer, on the other hand, considers whether or not something is a risk, exposure to risk, or risk factor; the observed (at risk or not at risk), on the other hand, frequently does not know whether they have options to pick up from. We will return to this point later.

In spite of these nuances of conceptual divergences, Luhmann's reflections on risk bring highly valuable insights to our discussion. The prominence he gives to the observer and point of observation of risks is highly relevant for us. In his view, the concept of risk does not indicate a fact independent from the observer. Whether the observer is the one making decisions on risky matters (the entrepreneurs), or the ones suffering as victims of unpredictable infections (the patients or collectives), risk requires an observer that defines it as such. In that we are on the same page.

Luhmann calls attention to the attributions that accompany the observation of risk; attributions of presence and strengths of causal factors; and the conditions where decisions are made. Although decision does not play a significant role in our approach to health risks, we can still recognize that the attribution exercise requires decisions by the second-order observer, who tracks the conditions of the identified risks and constructs the explanations and preventive interventions.

In fact, defining risks of undesirable health events is expected to come together with, however imprecise, explanations about the causal factors and possible solutions. To be clear, the second-order observer in our case is the health professional dealing with the risks of those who are sick or *at risk* of becoming sick (or their sickness worsening).

Pointedly, Luhmann also talks about prevention and how prevention is expected to affect risk and risk assessments, and the risk of having preventive actions with desirable effects, or not. These considerations fit well the discussions about health risks and preventive care, where the notion of "avoidance" becomes part of the characterization of risk, or more precisely the constitution of the notion of avoidable or preventable illnesses. This is relevant for the public health conception of risk, where the priority is to search for avoidance strategies as close as possible to the organic causality level (vaccination, prophylaxis, disinfection, elimination of vectors, education about healthy lifestyle, prompt access to healthcare services, and so on).

Furthermore, Luhmann highlights the importance of the time dimension in the construction of ideas of risk and related expectations and prevention. His discussions about time are of fundamental importance

for grasping key features of public health. We can exemplify time as a relevant dimension in the following senses: as the frame wherein risk can be traced from the onset of trends; as the identifier of limits; as the interval between the start of prevention and its expected effects; as the estimated transition to and from *at-risk* and to and from *not-at-risk* positions, and so on. We may say that public health exists because individuals can move from *at-risk* to *not-at-risk* positions, or be prevented from moving in the opposite direction. However tentative the predictions and the estimations of these passages might be, unavoidably such moves require time.

The meaning of risk therefore implies a time dimension where the actual risk is an index pointing at the future where the foreseen risk factors are expected to produce the undesirable negative events. We could say that risk is the present tense of a verb to be conjugated in the future tense (the disease might happen, the treatments might not be effective, the patients might die, and so on). Addressing something as a risk is an attempt to reduce the uncertainties about the future with indications from the past. Interventions aiming at reducing risks of future occurrences are carried out over the current forms of the identified factors, as they were or are observed in the recent past and/or in the present (which is already rapidly joining the past). The approach takes for granted the permanence of the factors, and the permanence of the factors' power and influence.

Perhaps more accurately, we can say: (a) we estimate in the present the risks of the past and assume our estimates are still valid in the present and will remain valid in the future; (b) we estimate the risk of the future based on the projections made with expectations of results of interventions and/or planned or otherwise fortuitous changes in risk factors. We do not know for sure the actual risk in this very present (we only have an echo coming from the past), and even more so the risks of the future.

Still on the time dimension, because risk is the risk of being a "victim" of what has not yet happened, although based on or understood from what has indeed happened earlier, by tackling risk, the public health subsystem is aiming at tackling what hasn't yet happened. Dealing with uncertainties in the time dimension, public health faces difficulties in pre-

cisely knowing whether it succeeds or fails in consequence of its actions or inactions, and its adopted or not conceptions of causality. Threatening paradoxes may carry on hovering above.

For a concrete example, let's consider mathematical simulations forecasting numbers of malaria cases in a given population, considering factors currently known and their respective relations and strengths. The simulated interventions may then be put in practice according to the constructed scenarios. The future will still remain uncertain. Then, when the future becomes present, it will be possible to compare the predictions with the facts. Most likely, there will be surprises. For one thing, among the possible variations, the intervention might largely differ from the planned one. Interventions often do not fully comply with plans, or results with expectations. So, for public health there is always a good deal of "business risk" in believing health risks predictions.

In this sense, dealing with and responding to expectations related to risk puts one in a tricky position. No surprise then that many works considered as public health in fact apply notions from other fields of knowledge where risk remains unaddressed. In this sense, risk is an unforgiving topic. We say again that risk is not observable; it is only suggested, contingent, inferred, because when risk is identified using the available data, it is already in the past, referring to the past, and therefore no longer observable as a current on-going "object". In contrast with an actual disease precisely detected in an actual body (which medicine concretely observes), collective risks are only assumed (within "helpful" confidence intervals).

Luhmann (2008b, p 42) notes: "There is no longer an objective vantage point for correct evaluation", and, we add, for risk. By correct, we understand he meant precise and objective evaluation. A few more useful sentences:

> "With hindsight, we evaluate risk in terms of whether a loss has occurred or not." (p. 42)

> "And from the future another present stares us in the face in which we will in retrospect certainly come to a different appraisal of the risk situation we are experiencing in the present." (p. 42)

> "It has to reckon with too many possible systems states." (p. 43)

In short, decisions have to be made about how to define and construct the risks of what will happen in the future. We have then to wait until then to see whether our predictions were correct or not, nevertheless remaining aware of the fact that the actions we currently take in line with our predictions may (or may not) have significant effects on the chances of the future occurrences.

We cannot do justice here to all the points Luhmann addressed in his book. It would require at least a whole chapter, or even perhaps a book. Before finishing though, we can mention his reflections on the late appearance of the concepts of risk, which from a historical point of view certainly has had an important influence on the emergence of public health, as a discipline and practice essentially based on the notion of observing, tackling and preventing risks within limited time frames. Risk is risk only in a time frame. Currently, the time dimensions of risk are part of the notion of health risks, but risk remains a complex concept, for not being self-evident and for its constructivist nature. Public health will certainly progress by devising forms and approaches to tackle risk and its time dimensions.

Take-home messages

Finally, we summarize our understanding of indicators as reflected in this chapter. Indicators are schemas of observation operating with distinctions. Those distinctions make forms appear in the medium of society (a semantic medium, we should keep in mind), ultimately constructing forms in terms of risks of becoming ill and dying. Even when the indicators seem apparently not to be talking specifically about risk, as for instance the indicators about organization, voice and accountability, and

others do, they indirectly are still addressing problems of health risk or intending to frame risk reduction possibilities, even if only tentatively.

Those examples (communities with *voice* demanding *accountability*, and *organizations* of services that can offer larger coverage and efficient health services) are to be judged for the benefits they bring for the health of the population; or, in other words, the decrease of risks of getting sick and/or dying. In the end any judgement about any indicator is to be made according to their contribution to understanding better the risks for the population and how they can be reduced. In all senses, the public health self-definition is about being in charge of the understanding and tackling of health risks.

Public health must not lose track of its domains: risk and risk factors. If something is recognized as a health risk, public health needs to say and do something about it. Furthermore, being the interface with the other systems, the political system included, public health will explain how "risky" the risk is and what can be done about it. Still "troubled" with uncertainties, public health will need to present the problems related to risks to relevant concerned individuals, systems and organizations, which may not yet have understood properly the unavoidability of uncertainties.

Chapter 7: The Construction of the Self-Reference of Social Health Systems

Having presented the theory, a short history of medicine and a discussion on public health as a sub-system of the health system, we can now go back to the three questions formulated at the introduction and try to answer them:

1) How does the history of medicine show the development of self-reference?
2) How did medicine self-reference become the foundation of the self-reference of health systems?
3) How did the coupling of the self-references of medicine and public health create health systems?

However, before addressing these three questions we revisit some of the themes discussed in previous chapters, particularly the topics of self-reference and complexity. The separation of the four periods of the history of medicine was intended to make recognizable when major transformations of medical knowledge and practice happened, and how the periods could be understood from the perspective of the development of social differentiation for the construction of health as a social system, as proposed by the Social Systems Theory.

While the marked differences of medicine development can be traced and described in those four periods, the evolution of society according to the differentiation thesis went through two historical forms: differentiation based on *stratification* and *functional* differentiation (see

Chapter 2 or Glossary for explanations of these concepts). We can only talk about fully developed functional differentiation in relation to the fourth period. The first three periods in the story of medicine unfolded while the societies had internal differentiation based on stratification.

To refresh the point, in the fourth period medicine became embedded in societies structured according to functional differentiation. With functional differentiation societies are organized in several function systems, each with their specific functions, codes, semantics and communications. Health became one social system among the others.

Self-reference and complexity

These two themes have appeared in a number of previous sections in this book. The reader may already be familiar with them, if not perhaps a bit weary of reading about them. However, as we approach our concluding chapter we still need to note a few aspects of these concepts, which are of crucial relevance for the thesis we advocate.

According to the Social Systems Theory, self-reference happens in the web of communications maintained in the system. We try to illustrate this process here. Doctors will sometimes talk with each other about what medicine is or isn't. Regularly in fact, they talk about what would be the correct medical procedure for treating a case, including what should be done or should be avoided. In that, we can say that medicine is what the doctors agree about it. That is, what they say reflects and is reflected in what they do and observe others doing.

It is not only the formal university training and post-graduate courses that reproduce medicine, but also, more importantly, daily professional communications recognized as carrying valid medical meanings. Obviously we are not talking about a single doctor or a small group of doctors redefining the conceptions of medicine; there are many organizations and institutions involved, all using medical terminology on a daily basis.

There is widespread social recognition of medicine and related social expectations about medical services and their effects. However, it

takes doctors to validate, endorse or otherwise disqualify any statement as medical or not. A medical statement cannot be validated outside the medical world; so, in this sense, medicine is what doctors say it is.

These communications reproduce and therefore maintain what medicine is. They communicate selected meanings and leave aside what is excluded. The history of medicine is the history of its self-reference because, at any time, only the selected meanings incorporated into the universe of communicable topics are registered as part of medicine and therefore part of its history. What was rejected was forgotten; what was adopted was preserved and reproduced.

A brief paragraph here on the conceptualization of self-reference in complement to what was presented in Chapter 2 should help us to move forward with the discussion about self-reference of health social systems. Luhmann (1990, p. 114) says that a system's self-reference is not only identity (the narratives the system constructs about itself), or only self-organization (structural and operational dispositions the system develops reflecting those narratives). He adds that self-reference refers basically to the constitution of system's elements by the system itself; elements that are produced and reproduced as self-referential units – that is, units that refer back to themselves as elements of the system that created them.

To step back a bit from the formal abstractions into an illustrative example, we can think of a medical speciality – let's say orthopaedics – created as an element of the medical assemblage of self-referred constituent specialities, referring to itself as a component of that assemblage. This apparently convoluted reasoning nevertheless has relevant implications for the analysis of self-reference. Luhmann calls attention to the fact that the admission of self-reference has relevant epistemological consequences for scientific studies of systems, and self-referential systems are indeed empirical objects without any transcendental status.

Continuing our journey, we acknowledge that self-reference is historically self-reproduced. We can say that, through communication, medicine selects valid statements that could be made about itself. The reproduction of these selections is repeated and communicated continuously, confirming and maintaining medicine's recognition of itself.

This denotes the very mechanism of identity creation, made possible by the self-referential functionality of a communication-based system. Still, history shows us that over the centuries medicine has developed diverse and increasingly more complex meaningful constructions to be included or excluded in its matters of concerns. Medicine as scientific endeavour or as daily undertaking to improve patients' lives is how it has been defining itself. We are repeating ourselves here, aren't we?

Moving on to the topic of complexity, if we accept the thesis that the development of societies and their structures goes in the direction of increasing levels of internal complexity,[1] the history of medicine can therefore be described as an evolutionary process of increasing complexity in a number of areas, namely:

- More complex models of the human body's structures, functions and diseases affecting it;
- More complex processes of observation, elaboration of prognosis and diagnosis;
- More complex procedures, tools and technical apparatus for treatments, with complex matters related to use, accuracy and reliability;
- More complex regulations and self-regulations of professional fields of practice;
- More complex training and professional development pathways;

1 It is important to say here that the unavoidable historical trend of increasing complexity is also continuously adapted, with strategies and tactics for reducing complexities where needed and possible, avoiding overwhelming the components of the system with complexities beyond their capacities. For instance, the creation and recognition of a medical speciality strategically reduce complexity, as the matters of concern for the speciality become its exclusive domain and the other specialities can resort to consulting the respective specialist when necessary. The overall system becomes more complex with additional specialized knowledge, but the additions are dealt with – i.e. reduced – by creating adequate communication channels. The theory calls attention to the continuous adjustments of these two trends of increasing and reducing complexities.

- More complex public health observations and actions for surveillance, prevention, monitoring and management of health services;
- More complex sets of organizational and institutional arrangements for healthcare provision and management;
- More complex coupling with other systems (legal, political, educational, scientific, religious, media, and so on).

These fields of complexities correspond to semantic realms sustaining communications in increasing levels of sophistication and differentiation of expertise. The self-reference of the health system has to "account for" all these complex domains – it isn't a simple task.

In a snapshot, medicine evolved from individual craftsmanship at patients' bedsides to the diversified practices in institutional and organizational settings, with specialist doctors in the currently complex integrated field of interrelated expertise, working in coordinated teams inside organizations' structures of health systems. The evolution has thus been from individual provider to systems provision; this reflects the parallel evolution of the medical knowledge uncovering the complexities of the human body and the diseases afflicting it, and, more recently, public health comprehensively addressing the health of populations.

We need to add here a few words in relation to the differentiation of the systems of science and health. The disciplines in the system of science have their own self-reference – as being a scientific field. In the domain of physics, for instance, communications are about the subject of physics and the validity of methods of observation. The same can be said about chemistry, biology, sociology, psychology and so on, each with its own self-reference and criteria to judge what is meaningful for them.

To be clear, as we have said in previous chapters, disciplines are not social function systems; science is the function system. Therefore, medicine is also a discipline in the system of science, as medicine is not strictly confined within the biology domain. Although human bodies are biological organisms, medicine approaches them with the specific objective to identify the health and more importantly the pathological structures and functionalities, intending to find ways of recovering health, if possible. Medicine also emphasizes perceptions, experiences

and descriptions the patients make of their conditions; biology does not have such a universe of possibilities and nuances. Therefore, it goes without saying that, despite being in the same scientific realm, medicine and biology are distinct disciplines, however controversial this statement may still be.

On the other hand, as also said, medicine is applied knowledge, dealing with concrete patients in the processes of diagnosing and treating them. Its interest in any particular case is distinct from the scientific intention of finding generalized or generalizable patterns common to human diseases. Medicine in this case is not just applied science; it operates in the empirical world, even when the science has not yet found the answers to the problem at hand. In this sense, medicine is an undertaking inside the domains of health as a social function system. We can distinguish between the science system with its interest in finding general truths (regardless of the use made of the formulated generalized knowledge), and the health system with its interest in dealing with the problem of any patient captured in the net of attention of the system, whether or not the scientific knowledge and means at hand are effective.

Medicine self-reference as a scientific endeavour is projected in its assessment and revision of concepts and paradigms it considers as constituting its field of concern. Medicine self-reference as part of operations of a health system is concerned with the problems it solves or fails to solve, and how it carries on communicating about that inside the system. It does not matter if all the scientific knowledge required for delivering a baby in an ectopic pregnancy is perfectly described in a book on the shelf, if a doctor in a particular circumstance cannot implement the solutions as needed.

As operation of health systems, medicine has to do something about the patient in front of the doctor. The patient is under some sort of legally established protection, granted rights to the service or covered by the ethical principles of medicine. Medicine as a scientific field does not need to worry about any patient in particular because the scientific investigation can choose its topics according to its quest to find the truth. The norms regulating the scientific field are attentive to the validity of the investigations and their results, while the norms of the

health system regulate entitlements, protected rights and obligations. The science system can live with its promises (theoretical science is still science). But a health system must live out of the responses it provides to concrete health problems. By the same token, as a legal system cannot leave a plaintiff without a final legal decision, a health system cannot ignore any patient; these are matters of justification of the very existence and self-reference of these systems.[2]

We can say that the same argument is also valid for public health, as a sub-discipline in the scientific system (with several interfaces with diverse scientific disciplines such as sociology, economics, political science, anthropology, and so on), and public health as set of concrete practices implemented by public health organizations in a variety of settings; the latter dealing with concrete problems, and the former dealing with the design of models and scientific methods to test them.

Now we move on to our three questions.

(1) How does the history of medicine show the development of self-reference?

Medicine showed specific self-reference peculiarities in each of the four historical periods, as we briefly discussed in the sub-sections for each period in Chapter 4. Based on that, we try to bring those elements together in the following summary narrative.

At a point in time, when a semantic set is constituted with selected meanings distinct from those in other semantic sets in the same society, we can say that self-reference is distinctive in their making.

---

2 Quoting from Bynum (2008, p. 115), "Claude Bernard (1813–1878) […] summarized his own research career, as well as developing a philosophy of medical research, in his classic *Introduction to the Study of Experimental Medicine* (1865). It remains a book well worth reading. In it, Bernard argued that the laboratory was the true sanctuary of medical science. In the hospital, where sick patients need care, and the number of variables means that observations are only piecemeal, no real experimental science can flourish. Only in the laboratory can the experimenter keep variables constant, so that changes can be unambiguous".

The orientation for the selections of meanings needs to be preserved together with the selected meanings. The orientation designates where the meanings are to be picked from and their connectivity with the already selected meanings. The consistency and cohesion of the meanings have to be kept while new selections are made and the meanings are deployed in communications.

Let's try to be less abstract. At some point in time, the body was constituted as an object of observation with the intention to see how and maybe why diseases appeared in it. The body was then understood as a medium (even if not acknowledged this way) where disease forms could take shape. This object, the body, was the concrete base to which meanings could symbolically connect – could refer to. The body thus became the base of references – references denoting and supporting distinct narratives about the observed diseases.

Taking a brief detour, there was nothing ontological in attaching specific semantic meanings to the body as a symbolic medium. The same body (as medium) became the basis for the development of very different semantic structures leading eventually to ancient medical knowledge systems as diverse as Chinese medicine, Ayurveda medicine in India, Western medicine and many others, with the semantic schemes unfolding in different directions, developing multiple articulations proper for each differentiated knowledge system. In Luhmann's terms we can say, "there is no compelling pull from the object" itself to adopt any specific semantics, privileging a specific direction.

All this is to say that at those times, a community of observers of the body developed,[3] sharing similar interests in talking about the body and its dysfunctions. The semantics they used became characteristic of their particular way of paying attention to the body and their concerns with it. Maintaining the semantics thus became a way of recognizing common interest and belonging to the group, which implied also sharing an epistemological orientation towards the elected object of observation (the

3 Similarly, those seeking artistic representations of the body also gathered together around the same time in Greece.

body) and, furthermore, preserving the memory of what had been observed and talked about.

If this has happened in such way, this does not mean that there was a decision-making process deliberately deciding and establishing such a community and semantic universe. Surely the whole process unfolded over generations and centuries, and many such communities may have started and ended without leaving any trace. Finding historical evidence of such initial stages may be an impossible task, as only the eventually consolidated successful and agreed upon practices, as seen in the Hippocratic writings, became evidence from the past. The description therefore only briefly describes processes of long maturation; an evolution with fortuitous and unforeseen events, and, as Luhmann says, of "highly improbable" developments.

When we see in Hippocratic narratives the Greek doctors visiting patients in their homes, listening to their complaints, and taking detailed notes of what they observed in the bodies of the patients, long evolution surely had already happened, without leaving vestiges behind. The Greek doctor went to the patient with an idea about the object he was going to look at, and the points of observation he would be most interested in checking. He took his notes following the already expected standards of such recording, and later on shared them with small group of doctors who were also interested in discussing their own observations and notes.

Those who could read and write and may have had access to the notes of the doctors could use the same words without sharing the same observational intent and rigour. In this sense, the use of the vocabulary was not yet fixed as the prerogative of doctors; no one could claim exclusive rights to use it. Whoever knew the words could use them as they wished. It would not have been possible to name anyone as a charlatan then; even if the distinct groups shared such a judgement about outsiders, such agreements would not have had repercussion beyond the group itself.

This was the scenario when the closure of the semantic space had not yet taken place. Without the closure, no group could claim ownership or exclusivity over the use of the terminologies, and there was no possibility of developing any self-reference to be endorsed by the society at large.

The self-reference would be for the members of the group, recognizing themselves only. A semantic universe distinctly used and preserved by a segment of a population, but not acknowledged by the society at large, could only work as a thin base for internal self-reference in those groups, with circumstantial external acknowledgement when a doctor may have been remarkably successful in treating some cases.

The point is, the meanings and words they used still "belonged" to the society at large without exclusive use being assigned to anyone. Therefore, the only sustainable self-reference then was that of belonging to the group of those who used the same meanings and also shared the same principles and intentions for the formulation and deployment of those meanings.

The members of the group had a good idea of what they were doing – looking at diseases and trying to describe them. Hippocrates and his followers intended to separate the meanings they used from the then current superstitions and common beliefs driving people to the temples to please the mythical God doctor Asclepius, giving prayers and offerings for the desired healing.

We cannot see any system operating at that time. All we can see is a vocabulary being developed, without the regulated circumscription of its use to recognized singular groups of users, and also not linked to a dedicated set of practices that only those specific groups could perform. In other words, a system could not exist without the semantic space and the designation of rules of inclusion and exclusion from the system – that is, rules for designating who was recognized as a valid user of the semantics. This recognition is a crucial step in the establishment of self-reference, as being where the validated meanings were used in communications.

Delimitation of semantic space arises together with self-reference, as cause and consequence of each other, in a tautological self-reinforcing relation. This also establishes the hetero-reference – that is, the other side of the distinction, denoting what is external, the outside of the semantic space, and therefore belonging to the environment. These conditions for the genesis of systems were not yet fulfilled.

Tracking the appearance of self-reference requires assessing the conditions that could bring it about. These conditions, as we have repeatedly noted, should include the distinct semantics, the operatives communicating using those distinctions, and their social recognition. The communications thus would by themselves select those admitted as valid, which would then reoccur, along with other communications likely to be considered valid. Self-reference is thus the acknowledgement that those conditions have been met. They are prerequisites of self-reference that have also embedded self-observation functionality, which made it possible to perform the selection of valid communications. Self-reference operated with self-observation. We talk about individuals and groups observing themselves, observing their own communications, and communicating about the way they communicated.

To sum up, we recall that self-reference is closely related to how a developed social function system makes sense of itself with the meanings it selects and produces. Luhmann uses the term autopoiesis to denote such process. A fully developed system would therefore show the following observable attributes of self-reference:

- Self-reference – seen in narratives of the system explaining the system (the system speaks about itself);
- Self-reference – seen in a system's ability to construct and change such narratives about itself (the system speaks about itself in its own terms);
- Self-reference – seen in exclusive prerogatives for construction of the self-reference narratives (the system is the only one able to speak about itself in its own terms);
- Self-narratives – communications by the system saying what it is;
- Self-reference as opposed to hetero-reference – the system acknowledges being hetero-referenced from other systems' perspectives.

Fast-forwarding to centuries later, an embryonic system, developing its self-reference, acquires new functionalities. Once doctors could recognize each other by their practices and communications, and therefore communicatively reproduce the semantics they used and the self-ref-

erence they constructed, they could allow themselves internal disputes about their findings, approaches and understandings.

For instance, doctors could become interested in knowing how accurate the identification of symptoms was, and how stable the predictions and prognosis based on that. From that point, around the end of the Middle Ages, medical thinking realized the crucial relevance of complementing observation at the bedside with post-mortem examination of patients' bodies. This raised questions regarding the correctness of the clinical conclusions, now possible due to the confidence medicine developed of its own capacity to investigate and find answers, and allowed the mobilization of efforts to get permission and authority to open the bodies and examine and make observations about them. The self-referred embryonic system needed to find ways to confirm for itself whether its hypothesis was correct, strengthening the criteria for assessing the validity of its communications.

Internally, among themselves, doctors created spaces of communication of their points of view, with doctors participating and taking sides in the debates. That led to advancing self-observations and self-references concerning criteria for judging opinions. The following quotation illustrates the process:

> „French innovations permanently left teaching hospitals with two regular events: the daily ward round, in which a senior clinician, followed by junior doctors, medical students, and a nurse, would see and discuss each patient at his or her bedside; and grand rounds, in which interesting "cases"' would be presented by a member of the junior staff and analyzed by someone from the senior hierarchy, in front of a large gathering of students and doctors at all levels of experience. Often, after the presentation of the patient's history and clinical course, and the discussion of the differential diagnosis, the autopsy findings would be revealed by a pathologist, and the whole life and death of the patient put together in a seamless whole." (Bynum, 2008, p. 62)

The acclaimed settings for communications among peers with legitimacy to validate them, was one of the ways the embryonic system

set about reproducing itself. The system's internal dynamics would be alien to outsiders. Outsiders' main interest was with the outcomes of patients' treatments, with no concerns regarding the merits of polemics they could not understand.

The system needed to establish stable processes to confirm for itself its presuppositions and understandings. Medicine was fully in charge of the assessment of its judgements. By that time, for instance, the theory of humours was no longer sufficient to explain what was being uncovered by the autopsies. It was therefore necessary to expand the explanatory models. Medicine was observing itself, assessing its methods of observation; in other words, it was observing how it was observing.

The Social Systems Theory calls this *second-order observation* – observation of observers, or observation of itself observing. Newer and better explanatory models would replace the old ones and medicine was ready for that step without being threatened by the risk of losing self-reference. New self-references were being made.

From the fifth century BC to the fifteenth century AD, to be socially recognized as a doctor, someone would have to be able to talk about humours and how they were responsible for what was happening to a patient. That communication would be acceptable to others also claiming to be doctors. By the fifteenth century, such communications would not be enough. Other references had arrived in the semantic universe and were acknowledged in every field where medical observations and treatments were performed. The claimant would have to add a number of other elements and considerations. Doctors could criticize their own communications and claim the validity of broader horizons.

With the enhanced capacity to communicate about itself, medicine acquired autonomy and differentiation in a way that made it impossible for other knowledge disciplines to talk about what was considered medical business. These developments also counted a number of key points for strengthening self-references such as university medical courses, practices at the hospitals, corporative organizations to protect those with university diplomas, and so on. With its assured internal self-reference, medicine could claim its own domain vis-à-vis other domains in society, and gain recognition.

In the sequence, once firmly established in universities, hospitals, guilds, associations, and so on, the forum for debating and communicating about medicine allowed new types of questions to be formulated. After anatomical-clinical and physiological-pathological fields were consolidated as descriptive narrative of diseases, etiological questions of actual causes of diseases were formulated and nosologies organized classification schemes.

Through acknowledging the limitations of the explanatory models available at the time, medicine "learned" how to do it again and again, how to develop and bring new paradigms under scrutiny as soon as new understandings of particular pathologies of particular body parts were described. By being able to perform the revision of its self-references, medicine opened the door to new discoveries and new revisions in increasingly differentiated fields, with newly distinguished factors and hypothesized causal relations between pathogenic factors and complex body structures now described in minute detail.

At this point we need to make reference to an additional relevant topic of self-reference of medicine, the Hippocratic Oath. That was a remarkable step with profound influence on medicine's self-reference in many senses. It established self-reference on ethical grounds, setting the acceptable limits of behaviour for those aspiring to be recognized and present themselves as doctors.

The acceptance of the oath orientated the self-identity of the individuals, as well as the mutual recognition among those seeing themselves as members of the particular category of doctors. Members of the category acquired a fixed reference to assess anyone willing to claim to be guided by and behave according to the oath.

On the other hand, the oath also set references for outsiders to judge and scrutinize those claiming to be doctors. The oath to some extent drew a line according to which anyone could assign, even if tentatively and imprecisely, real doctors and charlatans to either side of the distinction.

The oath had long-lasting effects on the setting of self-references and self-identity, covering thousands of years. In systems terminology, the oath became a component of the programme for self-reproduction of

medicine. It became a programme to create an identity to be continually reproduced throughout generations of doctors. It was a stable representation of what doctors were supposed to be. Anyone claiming to be a doctor could be asked: "are you a doctor? So, you must know the oath, right? Please tell us what it says". Although no longer used in graduation ceremonies, it has inspired similar oaths currently replacing it.

To conclude our discussion on the development of medicine self-reference, we need to talk about a few additional aspects. Self-reference also points to what may come next, in the sense that it establishes criteria to select what is to be accepted as self-reference and what is to be rejected as hetero-reference. Self-reference thus sets the boundaries of the system, at the same time indicating what would need to be further explored and, on the other side, rejecting what would no longer be considered.

Although creating and striving for stability, self-reference is also in continuous need of reconfirmation, re-evaluating what was selected and rejected, and reproducing those steps. Self-reference does not become "ontological", so to speak, or definitive as an inherent fundamental essence. The elements of self-reference are contingent, so they can be revised, replaced and can also be different from their actual content.

As a system becomes more complex, it needs to develop ways of incorporating complexities into its self-reference. A complex system generates opportunities for challenges in specific matters of its elements. But localized challenges do not put the broader self-reference of the system at risk, even if the self-reference attached to the challenged elements needs to be resolved. Therefore, small changes do not necessarily affect the overall stability of the self-reference definitions.

Speaking in less abstract terms, medicine developed ways of dealing with its complexities; the creation of specialities is one of those strategies. Specialities are recognized as part of medicine. However, radical changes in the main themes of a speciality would not constitute a challenge for the self-reference of medicine. When autoimmune diseases were identified, for instance, they constituted a paradigm change in the realm of infectious diseases and immunology. But this did not have a similar impact on general medical self-reference. Self-reference there-

fore includes the possibilities to create internal divisions (specialities for instance), hence keeping complexity "tamed".

In conclusion, we can say that by the third period the self-reference of medicine was well established. That constituted a model for other disciplines, developing in line with an orientation of providing specific healthcare, identified nevertheless as different from strict medical practices (such as dentistry, nursing and physiotherapy).

(2) How did medicine self-reference become the foundation of the self reference of health systems?

Although strictly linked to the development of medicine, the historical emergence of health as a social system does have its peculiarities. Health systems could only be established after medicine achieved the degree of self-reference and semantic development described in the previous section. However other conditions combined with medical evolution led to the appearance of a socially differentiated health system.

The emergence of a social function system (to recall, health social systems are social function systems in the terms of the Social Systems Theory), occurs once a number of preconditions are achieved:

1) The society has reached the function differentiation stage, where different social functions (political, economic, legal, religion, art, and so on) are established as systems (see Chapter 2).
2) The new function system has acquired the exclusive domain of a specific binary code (health/illness) orientating all communications inside the system.
3) Based on the binary code, a network of meanings constitutes a semantic universe of exclusive use by the new system (meanings established in a nosology for example).
4) With its prerogative use of the meanings it creates and preserves, the new system constructs its own self-reference with which the system internally self-reproduces its communications (operational closure).
5) The other function systems in the society, with their respective codes and semantic universes, are recognized as belonging to the environ-

ment as opposed to the system, and as such they can be observed by the system.

6) By observing the other systems and being observed by them, the system can enter into structural couplings with any other system that it considers relevant to itself.
7) With the coupling processes, the system becomes socially relevant for other systems, and at the same time the system's self-references incorporate internal communications in connection with those couplings.[4]

By the early nineteenth century, almost all these conditions for the establishment of the health social system were fulfilled. To briefly explain, developed societies were already structured according to the functional differentiation. Conditions 2, 3 and 4 were also achieved, as medicine had already developed its well-defined exclusive semantic domain, which only medicine could reproduce. Conditions 5 and 6 also met the requirements based on the previous achievements. However, condition 7 had to develop further.

Our thesis is that while fixing its attention on the bodies of individual patients and, despite the remarkable development of knowledge of the body's anatomy, physiology, pathology and so on, the therapeutic arsenals were still insufficient, and medicine did not have much to offer to society at that time. Besides, etiological knowledge had not progressed much; diseases were identified and described in detail, but their causes were not clear.

Furthermore, even more important for establishing the health function system, there was very little understanding and few solutions for the illnesses affecting many in society, particularly epidemics with their devastating effects. Medicine in this sense offered too little as far as the social needs of the time were concerned. Without firm and valuable solutions for society, the incipient system would linger on with little recognition.

---

4 Chapter 2 has a summary of the theory and explanations of these concepts. See also the Glossary at the end of the book.

The crucial development arrived with a better understanding of the causes of diseases and the constitution of public health, with tools and knowledge to address the health problems of the collective, tackling populations rather than individual patients, and, of high importance, achieving the recognition of the political systems, realizing the powerful potential of the societal-wide implications of public health policies.

We develop the arguments in favour of this hypothesis throughout this section. To start our narrative, we return to some historical points in the evolution of medicine. A comparison with the legal system can be useful here. The legal system, together with the political system, was possibly the first social function system to become differentiated in human societies.

Consider the lawyers as professionals dealing with and defending cases on behalf of their clients in courts of the Roman Empire or earlier. A lawyer could not exert his profession alone, he would need first of all laws to refer his cases to. Laws, in whatever form they may have existed, had to be enacted as politically binding decisions – that is, decisions made by a political system – whatever regime and power structures in which they were decreed.

Besides that, a lawyer would have to present his case in front of a jury or judge, who would also listen to the other side of the dispute (probably another lawyer), and then make a deliberation, which would have legal weight and meaning, would be imposed and compliance demanded. Compliance would be enforced by the mechanisms made available to the legal system by the political system. The picture could be a bit more complicated with the inclusion of the appeals courts. Without these legal structures, lawyers could not exist.

A doctor, contemporaneous to such lawyers, would not have any of that; in fact, he would not need similar structures. Doctors were not required to submit their cases to a court, or argue with opponents, or justify their procedures in line with enacted policies; none of that. In ancient Greek society an individual interested in performing healing practices, with some communication skills and capacity to interact with sick people, listening to their complaints about bodily dysfunctions, with some knowledge of herbs, massages, baths, diets, and so on, pos-

sibly including prayers and invocation of godly powers, could claim the role of doctor/healer and conduct his business, as long as he could find clients interested in his services and willing to pay for them.

Such a set-up persisted until the final century of the Middle Ages about 2,000 years later, when formal medical education became recognized and offered as university courses. Then, becoming a doctor was hard, requiring a high level of literacy and money to finance the expense of the training. Those doctors were conscious of the specificities of their category and the necessity to be distinguished from the charlatans or surgeons who were then considered craftsmen with the same low status as barbers. In that context, the development of interactions between medical professionals and the political power was convenient for both, as exemplified by the eagerness after the French Revolution to bring doctors into the fold, with official public initiatives and spaces for medical practice and the training of medical professionals (see Chapter 4).

We can say that the medical profession, politically endorsed and promoted, became the backbone of the health system in the making, lending to the soon-to-be health systems the essential distinctive semantics, practices and self-references, endowing them with a differentiated social identity with specific meanings.

In accordance with the acknowledgement of belonging to a politically supported category, medicine was granted space for practising and for its self-reference reflecting its legal and political recognition. The medicine of universities' research and training combined with the medicine of public hospitals and political recognition, gained a status granting it the legitimacy and freedom to examine bodies, alive or dead, as medically justifiable, and was authorized to carry out experimentation of techniques of examination, diagnosis and treatment.

At that time, the growth of the embryonic health system was dependent on the reproduction of medicine, which medicine could perform with its self-reference and self-observation functionalities. By already being able to perform its operational closure (reproducing meanings created by itself), medicine would make possible the operational development of a health system where it would occupy the core reference. With its specific meanings, medicine could perform the selection of opera-

tions, keeping inside the system meanings adequately communicated with validated semantics and excluding the rest. Only medicine could judge medical practices and communications and that was key to establishing the health function system.[5]

Medical semantic development was thus complete and the health system could be built on that. Also, medicine was no longer only practiced by individual doctors on their own with individual patients. Medicine was becoming institutionally located and its practices were acquiring social recognition and visibility. Hospitals became the privileged, but not the only, site for that. The French Revolution envisioned a system covering large parts of the population with services according to communities' needs. This reveals the attention the political system had started to pay to health as a collective problem and medicine as a possible socially acceptable solution (see Chapter 4). We can say that the terrain was fertile and prepared for the sowing. But something was still missing.

At the beginning of the nineteenth century, a number of developments met the remaining requirements. We can list the most relevant ones: (1) better understanding of epidemic surveillance and control measures (John Snow and cholera as water-borne disease); (2) better understanding of links between poverty, living conditions, sanitation and diseases (Chadwick); (3) civil registration, epidemiological data and statistical analysis (Chadwick); (4) whole-life events and hygiene (maternal, infant and child care, housing, lighting, fresh air, disposal of dead bodies, and so on – Johann Peter Frank); (5) inoculation and vaccination (Edward Jenner). The notion of prevention was embedded in all these fields.

Also common in all those was the orientation of what was understood as the "Public Health Movement", which eventually found recognition and a place in academic disciplines as well as political spaces as

5 The same applies to any other function system. Only the legal system can make valid communications with the legal/illegal binary distinction; only the art system can make credible statements according to the distinction art/non-art; only the science system can communicate with acceptable claims of whether something is scientifically true/false; and so on.

public health authorities. The recognition of social factors (brought into the medical consideration), for which the discovery of the transmission of germs and diseases gave strong evidence of the need for programmes and actions of wide social coverage, made incontestable the understanding that health was definitely the business of the state.

We may say that nothing would have happened if medicine had not developed its self-reference, which could be at the core of the health systems. An image that may come to mind is of a nucleus around which larger structures could genetically develop.

Medicine was already differentiated as a discipline with a clear shape among other scientific disciplines. But the practice of medicine as a service to society, as curing or alleviating diseases in individuals, could only be configured as a social function system by adopting broader perspectives of social relevance (not only individual but population-wide) for addressing, preventing or alleviating disease in societies.

When public health concerns acquired political relevance and became prominent, medical professionals were already in hospitals and communities, studying diseases and providing healthcare services. Medicine made available the knowledge about how prevention could be tried and implemented. Medicine provided the fundamental knowledge of diseases to structure and guide the public health orientation. Public health needed that foundation from where it could expand the social focus and reach of medical practices – by social we mean the complex network of interrelated and intertwined sets of factors such as political, demographic, economic, cultural, educational, religious as well as environmental.

It is clear that medicine became and has remained the core of the health system, while the health system itself, with its public health self-reference, developed and incorporated additional concerns and specific semantics. Within the development of the health system there are innumerable communication operations related to a diverse range of disciplines. The development of the system around its medical core and its binary code (healthy/sick) made it possible to connect, incorporate and reproduce expertise such as nursing, physiotherapy, dentistry, pharmacology, biomedical engineering, as they and their specificities

became constitutive parts of the health systems, contributing to daily care of patients. This is addressed in the next section. But crucially, the health system emerged with the ethos that society as a whole had to benefit from healthcare knowledge and practices, and public health occupied that space of meanings.

In short, we can say that without medicine's consolidation of its communications (based on the distinct healthy/sick binary code) and related semantics, there would not be a health social function system.

(3) How did the coupling of the self-references of medicine and public health create health systems?

In *Law as a Social System*, Luhmann (2008a, p. 423) says: "An external, scientific description of the legal system does justice to its subject only if describes the system as a system that describes itself and constructs theory about itself".

With this quotation about the legal system's self-descriptions in mind, we can reflect on whether the same applies to the health system and its components: medicine and public health sub-systems. We may say that a description of a health system as a system should be made by the health system itself in the first place.

From the beginning of legal institutions in ancient times, lawyers operated (communicated) with others in the ambit of legal practices – legislators, judges, attorneys, jurors and so on – according to the roles they played at the time. In contrast, only by the end of the third period (eighteenth century) did doctors start to receive support, with a number of roles and schemes involving non-medical health professionals.

Today, the practice of medicine is surrounded by expertise of professionals such as pharmacists, nurses, laboratory and radiology technicians, physiotherapists, psychotherapists, nutritionists. The complexities thus established need to be orientated in a systemic way; a system has to be in place, reflecting a comprehensive unit and its role vis-à-vis the society, all under the same fundamental binary code healthy/sick.

The practice of medicine changed, and it is no longer just an encounter between a doctor and a patient; a lot has to be arranged around

those encounters, beforehand, during and afterwards. Many inputs, tools, resources need to be assembled (and essentially communicated about) to what is no longer simply an interaction that starts and ends during the meeting between patient and doctor. Health-themed communications now are used in many instances involving different professionals. This clearly indicates the necessity of having a system in place, requiring meaningful communications, in predictable fashion, corresponding to expectations and preserving stable meanings.

Being at the core of the health system, medicine now requires all the services that surround together with medical actions; the system makes all actions meaningful. A health system, as a self-referential system, constructs its own image and organization, creating a complex unit of which medical treatment is a component – the core one, but still a component among many others. The self-reference of the system has to be constructed and continuously maintained, incorporating medicine but not limited to medicine's exclusive concerns.

On the other hand, society is everywhere in the health social system. Public health is in charge of dealing with observing society from the point of view of the health system. The sub-system of public health evolved and differentiated itself from medicine in the internal environment of the health system, actualizing the overall orientation of the system towards society – that is, constructing the references by which the system can recognize and justify itself to itself and to other social systems, vis-à-vis the society the system has to assist. The health system recognizes itself for both the delivery of medical care and the delivery of public health operations for the health of the collectivities.

Very early in the history of medicine, as noted in previous chapters, the political powers of the time understood health as government matter, as societies needed healthy soldiers, cities to be kept reasonably inhabitable, and epidemics contained. Progressively, around the third period (also explained in the respective section), medicine also became a matter for educational, economic and legal concerns, each considering health and medicine from their own perspective. The "interfaces" and coupling with these other systems, had to be elaborated with the observation of their concerns, with public health taking shape by setting ten-

tative questions and tentative answers, and evolving as a distinct set of knowledge and references.

Telling this history in other words and with a few more details, we can say that while medicine carried on finding diseases and treating individual patients, public health (in a broader sense) had to find ways of dealing with:

1) the political system on matters of what would be justifiable as collectively binding decisions concerning the health of the population from which the political system were continuously pursuing legitimacy;
2) the legal system on matters concerning the legal mandates of health systems related to the rights of patients, the limits of professional practice, and so on;
3) the education system on the exclusive control of the health system, and the limitation and validation of the reproduction of healthcare skills and competences;
4) the scientific system on the advance of knowledge in fields of health and treatments of interest, as far as those advances are intertwined with on-going healthcare practices and the outcome of the incorporation of new knowledge into the health system's communications and actions;
5) the economic system as economic transactions might impact access, supply and provision of health services, and therefore has effects on distribution of risks of disease, outcome of treatments and deaths.

To this list we could also add interactions of public health with the functional social systems of religion, media, art, and so on. But we do not go into those details here.

In a broader sense, public health, keeping its central focus on populations and risks (as we have previously discussed), and therefore using the secondary binary code of at risk/not at risk (the primary code being healthy/sick), observes the other systems and communicates with them at the level of the respective organizations. The examples are:

1) the political parties, communicating about political support for policies of interest for both the political and the health systems;
2) the legal system, on what concerns applications of law to populations at risk, where rights to access to care have legal implications and entitlements;
3) the education system, on decisions about the number of professionals to be trained, new schools and courses to be opened, and so on, in correspondence to the population's needs and health service coverage according to society's risk profile;
4) the scientific system, on updating communications and new knowledge, on field trials, and assessing scientific evidence for optimizing reduction of risks;
5) the economic system, on sale and purchase transactions performed by the health system's organizations with potential effects on the population's health risks.

Of course, such broadening of public health horizons may sound unusual for the common characterization of public health as only the sets of functions a Ministry of Health takes care of, with the essential purpose of preventing disease (vaccination, surveillance, epidemiological controls, water and sanitation monitoring, drugs controls, and so on).

But here we support the notion that planning and dealing with all issues that may have an impact on the health of the population are matters of public health concern, be it carried out in the ministry or among private enterprises (pharmaceutical industry, insurance companies), academic institutions or others. The health system encompasses all communications where the medical (healthy/sick) codes are used, and also all communications on topics of increased or decreased risks of getting sick and dying, and their population distribution.

Everything that might be relevant for the collective in terms of producing or avoiding diseases and disease risk factors, assuring (or not) access and provision of effective (or not) healthcare, falls within the domain of public health, which public health can and should have a say about. All those possible areas of concern are currently, in one way or another, found in institutions reflecting, fulfilling and reproducing this broader

public health role. Some may say that we are bringing the whole field of health management into the realm of public health, and we would agree with that.

Furthermore, of key relevance among public health functions is the self-reflection of the health social system – that is, how the health system defines itself, ascertains what it is and what it is not, and its duties. To demonstrate why the health system should be concerned with legal, political, economic, educational or scientific deliberations (as well as other social system decisions), the health system needs a vision of itself and of the environment it tackles, including the diverse causal relations of disease occurrences, treatments and preventions.

We need to recall Luhmann's theory, where he establishes that a function system cannot communicate with another, because they do not use the same codes (see Chapter 2). However, organizations, a type of social system, can communicate with other organizations when they have within themselves sections specifically related to the same function system. For instance, the legal department of company A can communicate with the legal department of company C because they use the same legal codes and can therefore understand each other's communications. This therefore allows for the organizations of the health system to communicate with organizations of the legal, political and other systems. Being part of the environment of the health system, the other systems (political, economic, educational, legal, scientific, and so on) are observed by the health system, which is likewise observed by them. The organizations in each other's environments can communicate with each other on matters they have in common, arising from their mutual observations.

In addition to that, different departments within the same company can talk to each other because, belonging to the same company and having the overarching membership identity common to them, they share the company's orientation concern. For example, in a company, the legal department can talk to the engineering department, each recognizing the use of their respective distinct semantics; however, they have common understanding when matters of the autopoiesis (self-reproduction) of the company is at stake. The legal department may, for instance, warn about the legal implications of decisions the engineering department

wants to make. The interests of the company offer the guiding rules and channels for communication between the departments.

Following from these points, we can understand that the health system, particularly its public health sub-system, can engage in communications with organizations of other function systems when health matters are at stake. For that, the public health sub-system observes and represents the other systems inside itself; the public health sub-system sustains internal communications about the matters observed in other systems.

The public health sub-system can select the themes of communication with the political, scientific, legal systems, in relation to which it internally communicates and makes decisions about. Public health departments select and prepare the topics to communicate with organizations/departments outside the health system. The health system is thus preserved. While communicating with organizations outside the health system, the public health system will always communicate from its own perspective, always a perspective constructed within the public health sub-system in terms that are meaningful to it.

The selections and constructions the public health sub-system performs in these processes are constitutive part of its self-reference. Without self-reference, public health would not have a position from where it could observe its environment, observe other systems and organizations in it, never mind communicate with them, because it would not be able to differentiate itself from its environment.

The crucial historical moment for the constitution of public health (as discussed previously) was the French Revolution. What that revolution brought about was a new way of meaningfully structuring the provision of healthcare, creating entitled citizens, and bringing the reproduction of medical knowledge in line with social concerns of the revolutionary ideals, in the process of dismantling the old privileges of the aristocracy and creating a new order.

In fact, medicine's treatments were almost as poor as they had been for more than thousand years, but knowledge of the human body had expanded considerably, and the consolidation of new institutions like universities with medical schools, and guilds representing doctors' in-

terests, created new forms of social recognition of medicine (new forms of medicine communicating to itself and about itself).

At that time, the political system took matters of health as part of its concerns. The political system produced decisions on issues of social relevance that the system considered necessary for its legitimacy. That included regulation of doctors' training, hospitals, coverage and distribution of services. Medicine could not advance itself in those domains without political endorsement and authorization of medicine's purposes. That corresponded to the institutionalization of public health, first within the political body itself, and then in the soon-to-be-established social health system with distinct public health administrations.

The public health sub-system thus created was in charge of constructing and presenting visions of where, how and why state intervention was required and would make a difference. Even if institutions were still being established, the construction in practice of public health self-references was activated.

With advances in the observations of epidemics, establishment of surveillance schemes, inspections, vaccination, hygiene, prevention, and measurements of distribution of diseases, risk factors and medical services, a universe of rationale and argumentation for interventions became available. In that process, public health constructed its own self-reference in terms of its unique capacity to identify social health problems and solutions.

Public health thus portrayed itself as having the keys to the door to better health systems and better population health, proposing planned distribution of services, structures, resources, and suchlike, aimed at maximizing benefits. Public health now has consolidated command of the tools to elaborate and justify the solutions. It can advise the political system on the decisions that need to be taken. However, the Social Systems Theory reminds us that political systems, as autonomous social systems operating in the context of societies' function differentiation, still retain their own ways and the exclusive prerogative of deciding on political matters.

As final words on the topic, while elaborating its scenarios, plans and justifications, public health has to keep intense awareness, considera-

tion and communication with the medical sub-system, understanding the requirement to preserve what has been achieved and its aspirations. Any plans and proposals have to be as meaningful for the public health sub-system as for the medical sub-system.

The health system cannot afford to have its two sub-systems in conflict. The self-reference of the two sub-systems together create the self-reference of the health system. The health system has no self-reference constructed independently from those two self-references.

We can summarize this section as follows. With medicine as its "nucleus", treating individual patients, the health system needed the "membranes" to separate the system from its environment, and to work the "interface", regulating the interactions with the environment. From a Social Systems Theory perspective, although imperfect, this "cell-like" metaphor may give a tentative picture of health systems with the two sub-systems. It is instructive, however, to explain that the "cell" metaphor is misleading because it gives an appearance of "solidity" or "concreteness" to the health system, while we are in fact dealing with systems made of meanings – communicated meanings. Social systems are constituted in the realm of meanings, and the distinctions of "nucleus" and "membranes" is made by meanings. All of that is to say that the health system emerges as a unit with these two sets of meanings (medicine and public health) combined with their own self-references.

PART IV

Chapter 8: Concluding Remarks

Why are these topics relevant? How does the understanding of the historical configuration of health social systems help us improve them? These are our themes in the concluding remarks. A summary narrative of the evolution of systemic features of medicine and public health is our starting point in this chapter.

In the evolutionary stages of the first two periods and most of the third period, the uncertainties an individual doctor had to deal with were kept as his own exclusive concern. The doctors considered examination, diagnosis and treatment of patients their own individual business. From a certain point however, still in ancient times, examination, diagnosis and treatment progressively became scrutinized by other doctors. Later on, with formal medical education and the setting up of guilds and associations of medical professionals, the scrutiny became progressively rooted in more institutionalized roles. By then medicine had developed its own control of training, licensing and practising, and had established the correct use of exclusively medical semantics. At that stage, these were *systemic* features being made operational. The uncertainties an individual doctor had to deal with were no longer matters of concern only for the individual doctor; these concerns now belonged to the system too, even in its embryonic stage.

The system had to make sure that all doubts were addressed in a systemic way; doctors' doubts became medicine's concerns. Medical knowledge had (and still has) to find the answers. It moved on to the stage where, among other initiatives, commissions needed to be formed and informed; all possible courses of action considered; all possible evidence

gathered; and then a system-endorsed decision could be made, which became the responsibility of the system as much as of the individual doctor.

Currently, there is a vast literature on every single aspect of diagnosis, prognosis, examination, treatment, procedures, and effectiveness, comprehensively addressing individual patients' particularities. These studies and measurements are recorded for a huge diversity of settings and conditions where patients are living and being treated.

Health professionals are often required to read through a vast amount of published evidence on the issues they are concerned with. Nevertheless, doctors are no longer supposed to have all this information in their minds. That is impossible; they cannot even keep up-to-date knowledge of all studies in their specific specialities. They need to use search engines and devices for selection of materials they need, gathering manageable samples of relevant texts. The task has more than ever become one to be dealt with by the systems; the systems have to assure availability, access and support for dealing with an ever-increasing literature.

We can also say that, in huge contrast with the early stages of the profession, medicine has become less and less the medicine of the individual skilled doctor, fully knowledgeable on the matters at stake. We now see the medicine of teams of clinicians, specialists, surgeons, and so on, aided by information technologies mobilized when needed, plus a multitude of supporting professionals with their varied specialized tools (clinical laboratory, radiology, imagery, physiotherapy, pharmacology among others).

The complexities already uncovered and requiring further attention are beyond the capacity of any individual professional. Doctors do need the health system, as much as the health system needs them. The system brings about the communication channels and connections to guarantee stability for handling the ever-increasing complexity of the topics.

We therefore see medicine (and the health system as a whole) facing two environments: the internal and the external. The external environment of the medical sub-system encompasses its objects of investigation (the body, its structures and functions and the causal relations

of diseases), and the internal environment of the sub-system comprises diverse specialities and fields of information with their specific semantics. The internal complexities of the semantic universe of the system are continuously undergoing a dynamic expansion, aiming at representing inside the system (in its semantic universe) the complexities being continuously observed and tackled in the external environment (the human body).

However popularized medical information has become in recent decades, with many publications and movies for consumption by the non-professional public, and increasingly easy access to lay and professional publications online, it is still not possible to compare this dissemination of medical information with the speed and volume of new themes constantly published in specialized vehicles. As already said, this represents a considerable challenge to medical professionals willing to keep themselves up to date with the new developments. No doubt, the gap between non-professionals and professionals remains wide and is most likely increasing. The terminologies have become too highly technical for widespread understanding.

The complexities are indeed vast. Artificial intelligence (whatever this oxymoronic term means) arrives at this juncture as a tool for processing large amounts of data and finding indications for suggesting diagnoses and treatments in some circumstances. But far from solving all issues, it brings additional ones – for instance, to "negotiate" adequate fitting before AI can find suitable niches, without removing the ethical and legal responsibilities of the doctors who make the final decisions and carry out the treatments.

However, awareness of the limits of what medical attention could deliver for the well being of populations became clear at the point when health systems were taking shape as function systems in the nineteenth century. It was acknowledged then that the attention to individual disease was only telling part of the story. There was a universe of factors contributing collectively to individual's vulnerability and exposure to pathogenic factors. Furthermore, accessing services, getting the necessary treatment and surviving the illness were dependent on many causal factors present in the system. Medicine was not expected or supposed

to address such wide horizons beyond the context of diagnosing and treating individual patients. Public health was thus created with the mission of scrutinizing those horizons and finding collective solutions.

Looking at distribution of risks (of exposure to disease, getting sick, gaining access to diagnostics and treatments, getting cured, dying – as explained earlier, the term risk can be applied to all those possibilities), public health opens itself for consideration of all factors that may be involved in framing and addressing risks.

As opposed to medicine and its focus on the body and the diseases affecting it, public health needs to look for and find risk factors in the environment (physical, economic, social, cultural, and so on). It needs to consider elements that by definition do not belong to the medical realm, and rather pertain to other knowledge fields and science disciplines, as well as other social function systems.

For that reason, public health can enter into structural coupling with other systems, deploying semantics that are understandable for the other systems. Public health can develop communications demonstrating that some decisions on education, economics, social policies and political systems may have detrimental effects on the health of the population or may achieve positive results. In this sense, public health both frames problems and arguments, and indicates solutions other systems can consider in their specific decisions.

Medicine would not be able to do that, because medical semantics are strictly concerned with diseases in the body and how structures and functions of the body can be affected and restored by treatments. No other system could enter into structural coupling with medicine on those matters (except the science system). Nevertheless, when the issue at stake is identified as disease, medicine has exclusive deployment of the semantics of diagnosis and treatment.

So, as conveyed at many points in this book, public health makes health a social system, to which medicine belongs. This is the thesis we hope we have fairly and clearly explained. We may say that public health brings the *social* to *social* health systems, and addresses health as a social matter.

To conclude this summary, we need to talk about *self-reference*, the term that appears in the title of the book and is of great relevance for us to understand health as a social system and its historical constitution.

We have shown in this book the emergence of self-reference in medicine, when it became the sovereign of its semantic domains, and therefore could self-criticize and self-reproduce the meanings it selected as pertinent to it. Medicine became what medicine said it was. We have also shown how public health recognized its task of addressing risks and risk factors and how it established its domain in the use of instruments and tools (indicators being one of them) to formulate for the health system what the health system was and how to observe its developments. By creating the health system's image of itself, public health also creates its own self-reference, which it can adjust and change according to its own criteria.

It may sound confusing that these two self-references, so distinct in their scope and focus, can jointly compose the self-reference of the health system. This seems to create a tension: how could the two self-references be harmonized? Is there a third self-reference – the health system's one – that differs from the other two? The answer to the second question is no. The self-reference of a health system is either referred to as medicine (addressing sickness) or as public health (addressing collective risks). Surely the interconnections of these two sub-systems are permanent and continuous. For instance, public health can contribute with relevant findings from epidemiological surveillance; medical information about diagnostics and results of treatments are also relevant for public health estimations of a population's risks.

Thus, there is no reason or justification for separating these two fields as if they belonged to different systems, or as if there was a third system comprising both of them. As mentioned earlier, and needs to be emphasized again, public health has among its attributions the role of developing the image of the health system as a system for the health system itself.

A health economics approach may tell us there is no permanent and unique solution to the question of the optimum resources allocation for delivering care to single individuals, according to their specific needs,

and at the same time achieving a general allocation based on collective rationale – that is, maximizing the reduction of risks for populations (a sort of Pareto optimum). This is why universal health coverage is a problematic concept. Individual healthcare and population health risks are hugely complex dynamic universes; they have to be addressed by selections, for describing allocation options and for delivering necessary healthcare on a daily basis. Making selections requires facing reality in some ways.

But let's continue our reflections on the two main sub-systems of the health social system: medicine and public health. A particular relevant way they are different is how external observers see them.

Public health is the "constructed" face the health system presents to society. It explains what the system is, what it does and what it is aiming at. The indicators reveal the picture in terms that can be understood both technically and by those without technical knowledge. Public health narratives are designed for diverse audiences. The narratives may enter the discourses of political parties, establishing political aims for the health system. They can circulate in non-specialized media, promoting a sense of what is at stake and how the system is responding to the challenges as far as the health of the population is concerned. The narratives can also enter discussions in communities, NGOs, associations of patients, health insurance policy-holders, and many other types of social organizations, in line with their specific agendas. The narratives may be incorporated into the expectations of population groups claiming rights. We cannot examine here the full diversity of communications that may be generated with public health narratives.

We can metaphorically say that public health is the "dialogue face" that the health system offers to organized society (including other social systems such as the political, the educational, the economic, the legal) to explain the health system's orientations, operations and results. This "dialogue face" also listens and observes expressions and reactions. If a hospital needs to be built, if recruitment of health professionals is needed, if acquisitions of advanced health technologies must be justified, if the termination of a health programme has to be explained, then it is the responsibility of the health system, oriented to population risks, to ar-

ticulate the narratives in correspondence to what the health system says it is about to do or has done.

However, to grasp the health system as a comprehensive whole, we need to reflect on the fact that the efforts made in the public health subsystem is only a small fraction of the communicative dynamics on the medical (service delivery) side of the health system. Here, countless communications happen on a daily basis. The public would not easily understand the technical language used in those communications. They are essentially technical and the sentences are mainly meaningful for those who have been adequately trained. This section of the health system operates inside its semantic space, with very limited opportunities to be meaningfully observed from outside. However, we can say that this is where most of the life of the health system actually happens. The immense universe of communications taking place in this side of the health system needs to be continuously reproduced, and thus maintained without interruptions.

The health system cannot expose – let's say to the political system – what happens inside the medical sub-system. A politician who is not a health professional would walk the corridors of a hospital and would not understand much of what they would see or hear from the professionals in their routine work. The politician, like the political system in general, would need the figures and explanations public health teams put together to explain why that hospital requires, for example, a day hospital building constructed to alleviate the pressure of the demand.

To that end, public health officers make an abridged version of the system, reducing complexities, showing the hospital in terms of concise indicators, making the health system visible and somehow "tangible". The politician can then meet their peers and discuss what they have seen and the explanations they received during their visit to the hospital. They now have a sense of what the health system they are dealing with is about. Such complexity reduction could not be performed by medicine. It does not operate in such a mode; it seeks to reduce complexities in its field of concern – diseases and treatments.

But the politician saw the health professionals circulating, communicating and interacting with patients, how busy they were and how in-

tense the life in the hospital was. The politician could therefore realize that they could not deal with the health system in that setting, lacking the appropriate meanings to communicate, and even if they had, that would not give a view of the system that is constructed with narratives and numbers portraying specific aspects of the health system. But the politician could also realize that the health system was both the health professionals and their daily business and the public health professionals, putting together understandable numbers and narratives. The health system is therefore the collaboration of these communications.

Another aspect the politician may notice is that public health narratives can be presented with alternative scenarios, with contingent priorities, programmes and projects, adaptable to changes in the political scene and new aims and targets. There are important trade-offs to decide on. Meanwhile, the actual medical service has an implicit recurrent dynamic that must be maintained while the diseases and capacities to treat them remain more or less the same in the short and medium term. In contrast, public health priorities may change according to political commitments and the resources available; a new government may want to project its health agenda in novel direction. Notwithstanding, health services should at least remain available to patients arriving at health facilities.

These two different dynamics have influence on the self-reference of the health system, which has to incorporate both time horizons in its references. As practice, medicine focuses on the present and the patient to be treated now; while public health, checking the health risks in the past, tries to figure out how to reduce the risks in the future.

The system as elaborated by public health, with numbers and narratives explaining and justifying what was done and achieved, and what comes next, is the most visible face of the self-reference of the health system. Public health orientates itself towards the external interactions it needs to cultivate and preserve. Meanwhile, the self-reference of medicine (including all health service delivery professionals) are internal matters of the system and professional organizations; they do not require external approval, support or endorsement of their technical judgments.

In this sense, the self-reference of the health system looks inside its own communications as well as the communications coupling with other systems. The health system does not have an identity or self-reference independent from those two self-references. That would not make any sense. In one way or another, the health system will always refer to itself either as medicine (including, as we have emphasized, all health practices delivering healthcare to individuals) or public health; as treating the diseases of individuals or addressing issues in order to reduce the health risks of the population.

These understandings have implications for any project intending to strengthen health systems. It is necessary to grasp the different self-references any project focuses on – medicine or public health or their connections. Any strengthening of the medical side of the system needs to be represented and included in the updated self-description the public health sub-system produces. The advances in both sides of the system need to be recognized and incorporated into their specific communications. When the public health side uncovers evidence of important changes in the profile of diseases, the medical side of the system will develop clinical awareness of the new challenges it faces. These are simple examples. In many ways the public health views of the system and the medical approaches to patients are to be communicated and understood by both sides. The strengthening of the system will need to explicitly pursue these aims.

Ways of looking

This section is a dialogue between two health systems professionals. We call them Alter and Ego.

Alter: I read your book on the history of medicine and self-reference of health systems. I am not sure whether I found something to help me in my assignment in the health system of country X.

Ego: Perhaps if I say to you that the book speaks from a "second-order observation" position, this may help you.

Alte: I remember I came across that term, "second-order observation", while I was reading.

Ego: Yes most likely. Second-order observation is observation of observers. In this sense, the book talks from the point of view of the social system's observers observing "first-order" observers of the health systems – that is, mostly public health professionals.

Alter: If I understand you, this means the book has been written for those who observe health systems. The public health professionals should therefore consider themselves observed [smiles].

Ego: [smiles] Assuredly. Kindly observed though.

Alter: Could you please be more specific? I guess "second-order" observation drags us into abstract realms. Could you give us a few precise hints of what you are saying?

Ego: Look, public health professionals observe the health systems they work in. They talk and write about the characteristics they see, even when they do not have a systems approach, they have a "feeling" of what the system is. They describe programmes, structures, functions, operations and so on that comprise the system. They often use the word "system" meaning different things; seldom with the same meaning we use here.

Alter: Well, thinking about the book and the observers … I have the impression that there were more levels and layers of observers. Am I right?

Ego: Yes, you are quite right. I should explain that better. Perhaps we can even say that in this book we are operating at the third level of observation, if we consider the health professionals dealing with patients are the first-order observers, the public health professionals are the second

level, reflecting and dealing with the aggregated observations of the first level. The public health professionals observe the first-level observers. And indeed we are at the third level, observing and reflecting on public health approaches. If we take only the second and third level into consideration, as we do in many parts of the book, we could say that we are the second order of observers of public health as the first order of observation. I hope this is not confusing.

Alter: Ok, that is fair enough. But then what does this "health social systems" of yours add to the picture?

Ego: I will need to unfold the explanations of a number of points. But for a start, let's say that if you see the "first-order observers" (public health professionals) describing the system, you need an observation point from where you can see them as an element of the system they describe. They are doing self-reflection "on behalf" of the system. This is crucial. The Social Systems Theory helps us to see that.

Alter: I see. Yes, I remember the section where the book talks about public health creating narratives about the health system, explaining what and how the system does what it does. And those narratives are used in a diversity of contexts, including political parties, academia, media, communities, and so on. I am fine with the idea that the health system is seen as described, but the descriptions could be different. Other agents could perhaps use other meanings.

Ego: Precisely! But more important is that the health system relies on its self-descriptions not on other people's narratives. That is crucial for its self-reference and identity. When we as consultants advise a health system, we must not forget that we are external observers and the system observes itself. Their observations are more important for the system than ours.

Alter: Sometimes, while I was reading the text, I felt a bit uneasy about the metaphors you used. For instance, on this theme of self-reference

and identity. The metaphors give an impression of a system as an agent, having a will, a "personality", so to speak, an ego, a big one, fully conscious of its domains, surroundings and decisions. Isn't that a bit strange?

Ego: [smiling] Yes, indeed, we are so completely aware of our introspections, as our own and unique identity, that we find it difficult to figure out this idea of identity without an ego. A central distinction Luhmann has developed in his theory is that conscience reproduces itself based on thoughts, while a social system's self-reproduction is based on communication. Communication needs the consciences and therefore there is no social system without coupling with consciences.

Alter: So far so good. So...?

Ego: I do not have direct access to your thoughts. In fact, to anyone's thoughts, to be clear. Through communication we get a glimpse of what is in someone else's mind. But miscommunications and misunderstandings are common.

Alter: Sure, there is no doubt about that.

Ego: Nevertheless, with the "high improbability" of reaching common understanding [smiles], bridging the unbridgeable gap between minds, the consciences engaged in communications create or refer to meanings that can be preserved and communicatively reproduced. Communications offer the marvellous functionality of recursively confirming (or not) the selection of the communicated meanings. We can always go back to what was said or written and ask for clarifications or, even more important, express agreement or disagreement. Communications keep open the possibilities of yes or no, and the unfolding of the conversation goes in one direction or the other.

Alter: Let's see... the intuition is that through the linkages of communications the system acquires a sort of stability and permanence.

Ego: Exactly! You can imagine how revolutionary the invention of writing was. It strengthened the stability of the meanings in unthinkable ways for those who relied only on oral communication. Even if the authors of the narratives explaining a health system have long forgotten what they had said [smiles], the texts are still there for anyone to scrutinize.

Alter: Using an analogy, we often hear about written works acquiring a life of their own, sometimes totally disconnected from the consciences that first constructed the texts.

Ego: Yes, and we can also figure out how in a complex hospital context, for example, many expectations and actions are based on ongoing communications, linking past, present and future communications according to the expected meanings, to coherently follow from those already uttered. I say a complex hospital but this is also valid for all health facilities, no matter their size, and institutions.

Alter: Ok, if a communication links to others, reproducing expectations and meanings, we can say that a system is at work. Right?

Ego: I could not have put it better. Perfect! Stop the communications and the system disappears [smiles].

Alter: I guess we can see that communication is how all those things related to self-reflection, self-reference, self-description, self-reproduction come into being. And all these things have stability but need to be constantly reproduced and reconfirmed. The systems therefore live with the tension between what it strives to preserve (the selected semantics and narratives) and the ambition to develop new expectations and incorporate new meanings.

Ego: That is the point. When "first-order observers" observe their system, they need to generate narratives to be communicated and hopefully understood by those in the internal audience of the system (the other

health professionals) and some relevant addressees in the external audiences.

Alter: Maybe you want to say something about complexities and meanings, recurrent themes in your text. Why are they relevant for those working with health systems?

Ego: Before that, let me emphasize a few things. Those working in low- and middle-income countries with health systems that are not able to offer the whole range of benefits that medical science has already established should observe key aspects of the history of medicine and public health.

Alter: Ha! Yes, that is part of our main concern in this conversation.

Ego: The evolution of the health system is a history of communications evolving through remarkable stages. At the extremes we see the huge transformations that took place: from making simple, direct observations on the body surface, to highly complex explanations of diseases; from nonexistence of health professions, to highly institutionalized practices of a multitude of professionals; from unregulated reproduction of knowledge, to university-based training and thorough professional self-regulation; from a free-for-all exercise, to matters of high relevance for the states. The historical perspective should remind us that in spite of the fact that in the twenty-first century, medicine and public health have spread their knowledge throughout the globe, there are many countries at different levels of constituting their own health systems.

Alter: Wait! I am sure you are *not* trying to say that there are countries with health systems that resemble the medicine practised by the Greeks or in the Middle Ages, right?

Ego: Absolutely! Having reached the level of integration of communications seen on the global scale, obviously we do not find health systems

with the characteristics of the embryonic health systems of the past. No, this is no longer possible. But we may see absence or weakness in structures that were of crucial importance for the evolution of medicine at some point in its history. For instance, the establishment of guilds and professional councils, which exerted influence of great importance for the development of self-regulation independent of the universities. In countries where the number of doctors is very small, with sometimes only one doctor of a given speciality, it is very difficult to create such collective overseeing boards with effective general consequences for the quality of the services provided.

Alter: I can also mention, for example, countries where doctors in rural areas work in isolation for years and hardly have a chance to meet and talk about medicine with other doctors.

Ego: Good example! If there are no communications between doctors, they have to rely on themselves individually, practising the medicine they learned several years previously. Remember, the text mentioned the establishment of the routines of doctors and their students seeing patients together in the hospitals in France. They then introduced daily sessions of communications about cases, which is still very much practised everywhere. In countries like the ones you mentioned, doctors have little or no chance to benefit from such fundamental practices. Remember [smiles], the observer of the system needs to trace and investigate communications; communications are the lifeblood of the system. For the good of the system, doctors should communicate with doctors and other health professionals as often as possible.

Alter: As we are running out of time, maybe you can say something about complexity and meanings.

Ego: Yes, complexity is indeed an important theme but also a tricky one. Complexity is not an empirical object. It tells us more about the observational capabilities than about the observed subject. Complexity is configured where the system recognizes that there are many known unknowns

and possibly also unknown unknowns. The observer of systems should develop the awareness about the level of complexities possibly present from the point of view of all those who work in the system. Without measuring or translating complexities into some metrics – which by the way is not possible in the sense of complexity we use here – the observer can still reflect on the crucial steps the systems take. On one side, the system strives to reduce internal and external complexities where they can overwhelm the capacity of the system to process information, but on the other side the system looks for ways of becoming more complex, and therefore capable of addressing greater complexities in its environment.

Alter: If I may elaborate on that, you can say whether I am on the right track. Let's see. When a health system opens a service it did not have before, such as radiotherapy, on one hand it becomes internally more complex, given the additional semantics introduced in the communications among professionals, which may not be understandable to many of them. On the other hand, the environment of the system also becomes more complex, because what was not recognized, diagnosed or treated previously now becomes part of the image of the environment the system deals with. The newly recognized diseases open a number of possible connections to causes and treatments; this will bring to the system new questions about what is going on in the environment and what the system is doing about that.

Ego: In that way, complexities are simultaneously reduced and increased [smiles]. Tricky isn't it?

Alter: Oh yes! Absolutely! So, let's handle complexities with care [smiles]!

Ego: Let me give you another practical and real example of where the notions of complexity can be helpful. A middle-income country is going to get a loan from an international bank to purchase equipment for its 200 or so maternity hospitals. The country is struggling with high maternal and infant mortality rates, so the public health strategy is intended to decrease the risks. Well-equipped maternity hospitals with trained staff

are supposed to be able to change that picture. Initial estimates based on inventories and needs assessment of all pregnancies and births identified a staggering number of many thousands of items to be procured, including all sorts of devices and disposable medical items for wards, surgical theatres, intensive care units for mothers and babies, incubators and so on. Not only the diversity of items but also their quantities would represent a formidable challenge for the ministry in charge of the procurement process. Obviously, it was beyond the capabilities of the ministry to handle all the steps of specifications, tendering, selecting bids, agreeing contracts with suppliers, receiving and checking the items, and distributing them according to the needs of the patients. So, complexity had to be reduced. How was it done? The list of items was shortened to 30 essential items of high relevance in obstetric care [smiles].

Alter: [smiles] Indeed that was quite a reduction of complexity!

Ego: Luhmann has an interesting way of talking about complexity. You may have read it in the book and in the Advance Topics section. Complexity is vital for the systems, which need to find a balance between *redundancy* and *variety*. These are technical terms from cybernetic science.

Alter: Too much redundancies and everything becomes rather boring [smiles]!

Ego: [smiles] Yes, the system becomes complacent, so to speak. Although redundancies are useful for setting clearly where the system is and where it can return to in case of surprises, it may set the system on a risky path, when enforced internal repetitive unanimity may lead it to fatal errors while addressing new conditions in the environment. So, a balance between redundancies and variety has to be found.

Alter: I could also say that too much variety and the system may get lost in the sense of being overwhelmed by the complexities it created internally.

Ego: Perfectly said! And excessive internal variety can cause the system difficulties in dealing with the internal challenges. Imagine how difficult it could be to negotiate all the demands of those 200 maternity hospitals, when some got more and some got less; or some got what they did not need at all, while some did not get the minimum they needed. It would be really tough.

Alter: And the interesting concept of "requisite variety" seems to be useful for grasping this topic, yes?

Ego: Well remembered! "Requisite variety" is what a system needs to be able to make adequate and reliable internal representations of the environment it is dealing with. In our example, the staff to handle all the formidable administrative tasks, if the procurement went ahead based on the first estimate of needs listing thousands of items, would not have the "requisite variety" to be able to represent internally in the system all that would be necessary to communicate and implement that option.

Alter: So… the conclusion is …

Ego: A public health specialist should be aware of these problems, particularly in cases where it is virtually impossible to estimate for each item or even sets of items the advantages in terms of decreasing the risks of the population at risk. The expert should make a general estimation of the positive impact of the investments and pray for the best [smiles].

Alter: Yes, there are quite a lot of complexities to be reckoning with. This is where public health becomes art rather than science [smiles]. So, I guess, we can talk now about meanings, right?

Ego: And meanings go hand in hand with complexities. Meaning is not a state of mind or something realized through introspection, or a psychological experience. No, in Luhmann's formal terms meaning is a distinction between actuality and potentiality. The actual meaning of a fruit is that it is not an apple, or pineapple, or watermelon [smiles]… If I say to

a blind man "there is fruit by the window", he will wait for additional explanations to then select from all the meanings associated with the word "fruit" those that may be meaningful to him. With additional information he will then be able to associate meanings related to size, texture, taste, weight, price, recipes, time to consume it before it is spoiled, and so on. If I say "banana", he will attach a number of those extra meanings, according to his experiences and interests. This is just an example of what in fact many contemporary philosophers and linguists have called attention to, highlighting the network of meanings fundamental for making any particular meaning meaningful.

Alter: Yes indeed, it is a banana [smiles].

Ego: A meaning is the selection of an option among many possibilities. The meaning is produced by the actual as opposed to all possible meanings that are set aside and dismissed. So, meanings are immersed in complexities. As communication is continuously flowing and dealing with meanings, it is important to be aware that all meanings are made by selections, and those engaged in communication are making selections all the time, both while uttering messages and receiving them. A message is not assuredly a truthful conveyor of a precise meaning because the two sides of the communication are independently carrying out selections. So, if we work with communications, we need to be aware of the meanings and contingent selections born of complexity that do not always coincide as intended by the communicating sides.

Alter: Perhaps we can leave that at this point and move on to the last topic, perhaps the one you see as more important to be clearly understood.

Ego: Yes, thanks. It is very important to make clear that public health works fundamentally with the distinction *at risk*/*not at risk*. The risk is defined as related to a potentially undesirable health condition (or outcome). We can say that "at" is about (refers to) the population showing a certain vulnerability, or exposure, or being potentially harmed, by a

specific condition. This is at the heart of any preventive activity. Public health interventions only make sense if they reduce risks.

Alter: Yes, I guess this is fairly acceptable.

Ego: But the issue is that public health has to go outside the health universe of meanings to find the attributes of that "at" [smiles]. The characterization of populations is not a core matter of public health. It needs to borrow meanings from other fields of knowledge such as social science, anthropology, political science, economics, management sciences, legal sciences, systems science. What public health does is the precise characterization of the risk being studied, and then with the borrowed concepts tries to find out the possible factors present in the societies, the famous "social determinants of health", for instance. Public health cannot ignore the perspective of risk, which is the reason for its existence. Without attention to risk, public health narratives become narratives of the scientific domain the concepts come from. If, for instance, we talk about a particular health system in terms of governance, voice and accountability, environmental impact, and so on, without referring to how exactly that has an impact on the risks of getting some disease or not getting the treatment required, we will be making political speeches and statements for the media, not public health narratives. Awareness of this is important for all levels in the system.

Alter: Maybe you still can tell us something about "closure" as a valuable concept of the Social System Theory.

Ego: Closure is indeed another central concept of the Theory. By closure we understand that only health systems can make legitimate communications using the semantics recognized by the system. The historical evolution of medicine tells precisely the story of how doctors' communications progressively became the exclusive domain of doctors. We tell this story in the book. The semantics became increasingly complex and only recognizable by those adequately trained for communicating in those terms. The closure closes, so to speak (smiles), the semantic universe of

the discipline, making it exclusive to those entitled to use it. In that way, the system creates its internal environment of communications, where all meanings are distinctively used and also constitute the system's self-reference.

Alter: Can I take this chance to ask: how precisely can this help a public health professional in the field?

Ego: Wherever public health professionals are involved, they will be able to assess if the semantic closure has been properly established. This includes the use of the semantics only by those entitled to: the correct use of the terms; the correct articulation of the terms in communications between professionals; the appropriate use of the concepts in narratives to the outside; and so on. Public health programmes need clear and stable communication of meanings. Within the closure, those who are supposed to understand the communications should be able to do so. The closure is only partially made for keeping the outsiders outside [smiles], but it is mostly intended to ensure that those who are inside are indeed inside – are fully able to understand and participate in communications with the specific meanings of the corresponding semantic sets.

Alter: I believe we covered the topics I had in mind. I understand that Social Systems Theory, as a theory, does not give us guidelines or protocols on how to proceed once we are in the field observing systems, nor does it give us frameworks for mapping contexts, processes or networks. Self-reference is not a trivial concept and self-observation brings a number of methodological challenges. Still, there are no step-by-step instructions on "how-to-do", or methods and procedures for data collection and data analysis in this book. Social Systems Theory rather indicates to us the orientation of the "second-order observation", which we may also call *ways of looking*.

Ego: Well said! If I may add a point…

Alter: Yes, of course.

Ego: For the professionals operating within public health sub-systems, and inclined to look for inspiration in the social systems approaches, I would suggest exercising what you called *ways of looking*. Pay attention to communications – without them, there is no system. Pay attention to observation and observers; check who observes and who is observed, and how observations are communicated through the system. Check when self-observation and self-description are carried out; this will be self-reference at work. Also, keep in mind that the communications are making and reproducing the system at the same time. And finally, never lose sight of health risks and risk factors, keeping track of how communications about risks appear, circulate, reverberate through the system and are incorporated into the self-reference of the system.

Alter: Ok! Thank you.

Chapter 9: Advanced Theoretical Topics

In this chapter, we address advanced discussions on Luhmann's theoretical approach to self-reference (Luhmann, 1990) and other topics, reflecting on how the concepts contribute to our discussions about health as a social system.

Decision, tautology and paradox

Although this topic may sound too abstract, we invite the reader to enjoy, if possible, the humour between the lines.

It may sound rather obvious, but it is worth noting that medicine cannot negate the distinction (healthy/sick) on which it developed. Likewise, public health cannot dismiss its fundamental distinction (at risk/not at risk). However, the self-references of both sub-systems do not require them to reflect on the validity of those distinctions. They can be problematized outside the health system by an observing system such as the scientific system or in philosophical discussions. Surely, though, the health system does not need to challenge itself that much; it does not need the instability generated by shaking its own foundations.

However, as we are in an observing system, a few epistemological reflections may illuminate the scene. More than a factual portrayal of the reality of the world, a distinction, any distinction, such as health and sickness, is attributable, recognizable and certifiable through communications.

Distinctions are essentially communicable. They also are logical operators with mutually exclusive opposing sides. That is the "nature" of distinctions.

However, between the attributions made by the distinction while being communicated, and the reality of what is communicated about, there are spaces of indeterminacy and gaps. A doctor using the four humours scheme to communicate observations could be understood and accepted in ancient times, but not in the eighteenth century, although the diseases could still be the same. The basic distinction healthy/sick is still the same, but the ramifications on the connecting side (illness) have grown progressively complex with the addition of increasing numbers of ramifications.

The fundamental distinction nevertheless is kept throughout an infinite number of communication operations, preserving its validity intact. The communications not only keep the distinction alive for the communicative operations but also admit it as valid and the very reason for the occurrence of the communications themselves.

We can venture an apparently complicated sentence: we can say that communication happens because communication communicates the reason and basis for communicating. Medicine (as well as public health) established its sovereignty to decide the validity of its communications. Doctors throughout their training, at any historical stage, acquire the role to communicate about diseases and also communicate about the validation of medical communications.

A medical communication is thus considered valid if a doctor says the communication is valid. For those allergic to tautologies, this statement may trigger a reaction. Yes, a tautology is indeed established.[1]

The same with the tautology "I decide as I decide". If we like, we can further expand the sentence into: "I decide without being subordinated to my decision because I myself made that decision in the first place". An

1 In his book *Essays on Self-Reference*, Chapter 7, "Tautology and Paradox in the Self-Description of Modern Society", Luhmann (1990) presents his views on the topic. Our discussions in these final remarks are based on that chapter.

individual makes a decision and then no longer feels bound by the decision because that person has made the decision. An example is someone deciding to go on a diet but not being bound by that decision because they made it in the first place (a good excuse). This can be further expanded into: "if I am subordinated to my decision I am not sovereign enough to make it. I can only be either sovereign or subordinated, not both".

This play on words tells us that, worse than a tautology ("I decide as I decide"), the tautological entanglement can further develop into a paradox ("If I am subordinated to my decision, I am not sovereign enough to follow it never mind make it"). The well-known example of tautology: "this sentence is a sentence", and of paradox: "this is not a meaningful sentence", helps to illustrate the difference.

"The decision is a medical decision because I am a doctor." This example seems to tell us that we can live with a tautology, because it does not prevent operability. We can continue; a medical decision can be made and accepted as such. While a paradox (such as "a medical error is not an error because it is not medical"), even a solvable simple one such as this, has paralyzing effects. Paradoxes have to be somehow solved before action can proceed.

Going back to tautologies, we can say that "the tautology – medicine is what medicine says it is" can be an acceptable statement. The tautology reveals the closure of the medical semantics we have talked about at many points in this book. We may even venture here to say that the tautology brought about the closure of medicine; it was only after medical doctors became aware and fully accepted the tautology that medicine achieved its fundamental closure and therefore its self-reference. This process can be observed in many other fields of knowledge, and indeed in public health too.

So, self-reference is rooted, surrounded, entrenched in tautologies. Let's accept therefore that health systems can tautologically preserve their operability, maintaining that "the health system self-refers to itself because it is the health system"; furthermore, "only the health system can refer to itself as a health system while no other system can do that".

We can get a bit dizzy in this "house of mirrors", but we can find our bearings and do not get lost. This is because, still, we can remove the tautology from sight, avoiding the feeling of hopelessness that accompanies the tautologically provoked allergic reaction.

In fact, self-references are always constructed with elements and relations collected from realities. The meanings constructing self-references can avoid the recurrence of the tautological statements by focusing on the selected elements and relations. We need to talk about health and the health of citizens, and treat them, don't we?

Luhmann (1990, p. 127) says: "In a very general sense, systems avoid tautological and paradoxical obstacles to meaningful self-descriptions by 'unfolding' self-reference. […] circularity of self-reference is interrupted and interpreted in a way that cannot be accounted for." The paradoxes and tautologies have to remain latent, observable only by external observers, such as examiners of PhD theses.

Notwithstanding, the health system is strongly attached to the narrative public health constructs with its selected meanings. The health system has plenty of narratives for "self-description"; therefore it does not need (or have time) to look at the mirror and state, "I am a self-referred system". Despite the fact that it is indeed a self-referred system, it is better to focus on public health's constructed narratives, and remove from concern the fact that the narratives are contingent and therefore the system could have formulated its identities differently. Regardless of such existential conjectures, the queues and waiting lists of the hospitals and health centres keep growing longer.

Luhmann calls our attention to the meaning of meanings. Social systems operate with communicated meanings, which constitute the only "building block" of any social system. He defines meanings as the unit of actuality and potentiality. This means that a meaning has an actual selected definition (actuality) from the complexity comprehending all other possible definitions (potentialities).

Meanings are contingent. They can be different; "they are not necessary or impossible", and that is so because meanings emerge from com-

plexity.[2] Where meanings are embedded in complexities it makes them contingent; there is always a surplus of possible denotations to choose from. In a sea of moving, submerged meanings, one of them will manage to temporarily stay on the surface and be selected.

Nevertheless, public health has a duty to fulfil. As an autopoietic subsystem it has to preserve and reproduce its codes with the means it creates itself. It has to tell what the health system is. The narratives have to be plausible and understandable. They have to be operative and prescriptive, and even, if required, presented in a legal and normative format. They need to be implementable and have observable results. They have to address the population (society) concerned, and optimally make possible for the population to see itself reflected in the presented narratives. The survival and reproducibility of public health rely on the accomplishment of these requirements.

So the stage is set and the play will unfold. The audience is out there waiting to hear what public health is going to say about the health system and the health of the society. Public health cannot simply go up on stage and state: "I am the only voice of the health system". That would not be enough. A voice that says it is a voice is only stating the obvious.

Studying self-reflection

In this section we also focus on the self-reference of health systems and particularly public health. The idea is to suggest that future research topics examine more closely the realization of self-reflection in practice.

As we have said at a number of points in this book, including in the previous section, public health has the attribution to explain the health system to the health system. Public health uses concepts and indicators for that.

---

2 In Chapter 6 of this book, in the section "Meanings and complexities: what the theory tells us", there is a presentation of the topic of meanings and complexities, where the reader will find more detailed information.

However, in the last two decades a number of attempts have been made to describe the health system as a system. These initiatives originated from academic institutions and international organizations such as the WHO.

Thousands of articles can be found in the literature on topics such as strengthening health systems and health systems thinking. These efforts were intended to make available to the health systems of the world conceptualizations of health as a system and how that could be incorporated into the way the health systems could look at themselves. The intention was and still is to provide a conceptual arsenal with which the health systems could develop some sort of self-reference.

However, the success has been very limited because the references still lacked the key orientation to focus on self-reference per se. To describe the health system as a system, public health needs to develop a self-referential description where it sees itself as a system – a subsystem developing self-narratives as a system, and aware of that. This sounds like a quite demanding task, doesn't it?

In their Ministries of Health (or similar institutions), public health departments all over the world have been developing the self-reference of their respective health systems. We have said this a number of times in this book. The remaining issue nevertheless is that those self-references made with public health indicators have a "blind spot", so to speak. The public health department represents and constructs narratives explaining the operations and problems of the health system it looks at. The department makes pictures of the internal environment of the system (with indicators about production of services, productivity, costs, efficiency, internal organization, and so on) and makes pictures about the environment where the system operates (coverage, epidemiological profile, achievements of policies and programmes, community participation, and so on).

Those pictures can be comprehensive enough for the aims of the health system and for its capacity of dealing with complexities. But the "blind spot" is still there, if we can find it.

The academic efforts to describe the systemic features of the health systems have not got very far while applying the frameworks found else-

where in other fields. The exercises have not been comprehensive and were only tentative applications of imported frameworks to specific details of the operations of health systems.

Those theoretical frameworks were not developed enough to incorporate the crucial theme of self-reference. The results have always been expressed in simple narratives of external observers (mostly academic institutions). As far as the internal processes of health systems are concerned, there are no examples of explicitly self-referential descriptions, expressing full awareness of the systemic features. Surely, the health system's narratives are still self-referential, from internal and external perspectives, as representing portrayals of the system as the system sees it. However, almost always, those narratives are not recognizing and acknowledging that it is the health system talking about itself.

Costa (2023) offers extensive discussions on the frameworks used for systems thinking, including those proposed by WHO (2007) in *Everybody's Business*. In that book there is a section dedicated to the groundbreaking *six pillars* framework. The argument is that the scanty theoretical references were either narrow or did not see systems as such, or both. The *six pillars*, for instance, deals with health systems as if they were large enterprises; in this sense it employs notions developed in management science about five decades previously, with the same functions renamed using health-related titles. Let's say, the *six pillars* talks about: (1) medicines, vaccines and technologies; (2) health information; (3) health service delivery; (4) health workforce; (5) leadership and governance; and (6) financing. Obviously these pillars were analogous to, respectively: technological composition; information; production and selling of products; labour; management; and finance. Management science had already established that decades before. No notion of system informed such understanding, never mind systems' self-reference.

The "blind spots", we can now say, is in the selections ministries' public health departments have made for the illustrations of their health systems. The selections separate what the departments want to portray and what they prefer to leave aside, in the shadows and out of sight. Rarely, or maybe never, do the departments take the initiative to acknowledge their selections as selections of a self-referential system that can only operate

with its selections. Selections are made but the act of making them does not need to be told in the narratives.

The observers separate themselves from what they observe. In order to carry out observations, the observer stays in the "blind spot" as the author of the selections and distinctions deployed in the observations. The observers can either observe themselves using distinctions and making selections (second-order observation – observation of the observer) or observe with the deployment of distinctions and selections (first-order observation).[3]

Nevertheless, the system is not required to expose its self-references to that degree. Therefore, self-reference is not revealed as such. What has to be revealed is what can be presented as concrete reality, not as choices that are contingent and could be different. Reality has to be acknowledged as real, not a matter of choice.

Interestingly, Luhmann (2015, p. 45) says "los sistemas, en su autotratamiento, desarrollan formas de aprehensión de la complejidad no accesibles al análisis y simulación científicas"[4]. In our understanding of this statement, the health system, through the public health sub-system, is capable of assessing aspects of itself that external observers from the scientific/academic system would not grasp or analyse, as the observation horizons are different.

However the health system does not yet speak the language of systems, and maybe does not know how and why it should talk about itself as a system. The health system (meaning the public health department

3 The observer can also observe the selections and distinctions used, but that is second-order observation – observation of the observer observing. In second-order observation the observer sheds light on "blind spots" and recognizes them. For that, however, distinctive distinctions suitable for applications at second-order level of observations will be needed. Nevertheless, "blind spots" can never be ruled out; they are unavoidable and necessary for observing at any level of observation.

4 Our literal translation is: "the systems, in their self-reference, develop forms of apprehension of the complexity that are not accessible to scientific analysis and simulation".

we have been talking about) does not see itself in terms of systems operations, or even as a sub-system concerned with its own self-reference. In fact, neither external nor internal observers see the system; they see the internal and the external reality they construct, and make sure that the fact that the constructions were their own remains in the "blind spot". So they do not see the self-referential system. Not yet.

Social differentiation, form of problem and complexity

In Chapter 2 we mentioned that the differentiation of health as a social system was built from within the health system itself, with medicine first and public health joining in and carrying on the reproduction of the meanings. We may also repeat that medicine built its differentiation by sticking to its medium, the human body, and to the orientation of its problem formulation about diseases, their causes and treatments.

Luhmann (2022, p. 84) says: "the discipline [he is talking about sociology as scientific discipline] most general semantic point of reference is not its title nor the most general form of its subject matter, but the problem formulation that constitutes the discipline". In the case of sociology, the fundamental question would be "how is social order possible?" He proceeds by saying that such a question already admits and does not doubt the existence of social order. The question is thus asked within a certain social order that makes it possible and meaningful to raise such a question. In this sense he speaks of a question that has been already answered because it takes the existence of the topic for granted.

But in the same text, in relation to the same fundamental question, "how is social order possible?", he talks about *unanswerable questions* at the root of scientific disciplines. As such, these are questions that can be asked again and again, focusing each time on nuanced differences and new specific themes. These questions can make the scientific endeavour withstand, continue and progress, remaining actual and valid. Enduring questions as such are advancing scientific knowledge.

We may say that similar questions are asked in relation to medicine and public health as scientific disciplines. The questions could be: "how

is individual health possible?" and "how is population health possible?" Surely these questions can be formulated where experience of such a thing as "health" is individually and collectively already recognizable. We therefore may say that these problem formulations constitute medicine and public health as scientific disciplines, looking for answers that to some extent have already being devised and experienced. Nevertheless, in face of new patients, new diseases, new risks for collectivities, these questions have to be permanently actualized and respecified

The form of the problem – that is, the form of the question – a repeatable formulation, can therefore sustain the whole edifice of a scientific undertaking, promising an answer will eventually be found. In Luhmann's (2022) terms, the progress of a scientific enterprise is maintained by unanswerable questions – questions that also do not argue about their fundaments (social order, as well as individual and population health, are already assumed to exist or be possible; and are unquestioned fundamental preliminary answers).

Medicine made its closure around its fundamental question; driven by the complexity medicine itself was disclosing, dealing with and paradoxically expanding. The problem (the unanswerable question) is the way the complexity is brought to the table and served in adequate digestible portions. Let's now talk about complexity.

In the history of medicine the will to know and the will to improve were omnipresent. Making these orientations permanent and possible, there was also the will to communicate, about what was being uncovered and about the results of treatments; in other words communications about the formulated questions and the provisional answers.

And there were always complexities – complexities of the subject being observed: the human body and its diseases. Complexity is the attractor, the inexhaustible source of things to be known, things seemingly waiting and "inviting" to be known, uncovered and revealed: the complexities of the mysteries of life itself.

It is clear though that we are talking about systems of meanings, not transcendental essences or ultimate truths. The complexities of the human body and its diseases had to be represented, translated into mean-

ingful communicable statements inside the semantic universe medicine has been working on and continuously rebuilding for itself for centuries.

The questioning about health has been incessant, as has been the attention to the symbolic medium, the human body, which also has continuously manifested a huge variety of diseases, in apparent unstoppable and largely unpredictable fashion. Therefore, so much has had to be learned! There is so much complexity to address! And communicate about!

These open complexities in the medium medicine observes, the human body, capture medical attention and efforts to construct semantic representations inside the also increasingly complex universe of medical meanings.

In other words, complexities have been growing on both sides – in the external environment where the complex object is continuously explored, and internally, where the apparatus of communicative structured semantics communicates about the represented complexities. In Luhmann's words:

> „We assume that complex systems are able to develop and maintain a system-specific order only under the condition of a higher level of environmental complexity. In this sense, a kind of ecological complexity pressure is not only the origin but also the operating condition of complex systems." (Luhmann, 2022, p. 89)

Complexity is thus productive and generative; it multiplies its horizons, expanding in many directions. It provokes, keeps and sustains the ultimate question: "how can health be possible?" Like a flame, at the same time illuminating and creating shadows.

As operational characteristic of systems dealing with complexities, the system needs to strike a balance between *redundancy* and *variety*. *Redundancy* consists in the repetition by the system of well-established meanings, creating the sense of concreteness and stability the system needs. On the other hand, *variety*, opening up the system to new meanings, makes it better able to address the unending complexities of the environment.

These two functionalities give the competences the system needs to deal with its internal and external complexities (see Luhmann, 1995, pp. 172–174). *Redundancies* protect the system against losing its selections and meanings, offering secure foundations for action. The system becomes less dependent, exposed or vulnerable to the circumstances. But it is also the case that *variety* protects the system against a restrictive and harmful unanimity, leading to fatal errors. It is of the nature of communication that it cannot prevent challenges, negating the validity of what was communicated – what, as opposed to *redundancies*, opens up the possibility of inclusion of *variety* into the system. The system is not built for acceptance alone; therefore a degree of *variety* is fundamental for making a system better able to address the environment's complexities and changes. Too much *redundancy* leads the system to become unable to face new challenges. Too much internal *variety* and the system may lose self-reference.

To sum up the dynamics of increasing/decreasing the internal and external complexities, let's say that: too much *redundancy* and the system moves towards an increase in complexity to counteract the repetitive inertia; too much *variety* and the system drives towards reduction of complexity. Any autopoietic system is engaged in such dynamic processes.

We can thus conclude this section by saying that the *form of problem* is a technique of approaching complexities; it establishes the orientation towards finding possible solutions, and at the same time allows for the indeterminacy inherent in complexities, the recognizable diversity and contingencies.

A question such as "how is individual or collective health possible?" implicitly acknowledges diversity and complexity; it gives room for the uncertainties of the unknown side of the complex subjects, and at the same time provides orientations to narrow down, even if provisionally, a selection of items and themes to be distinguished, observed and handled.

The general question remains unanswered and open to new formulations, open to be respecified, without vanishing when specific answers are found. It rather builds new formulations on answers previously obtained.

Furthermore, as the human body and its diseases open up universes of huge complexities, addressing the health risks of populations and distribution of healthcare services is equally hugely complex. So these complexities constitute formidable reservoirs for scientific exploration and expansion.

By being aware of the systemic manner of formulating problems and proceeding with enquiries, the disciplines acquire relevant self-references. As public health has among its responsibilities the task to explain to the health system what the health system is, its questions and answers make clear the social differentiation of health systems and their self-reference.

Symbolic medium and second-order observation

We saved for this section of the book a brief discussion of some important points of Luhmann's theory. The discussion should help us to see the basis and implications for using the concept of symbolic medium, and to understand the meaning and relevance of second-order observations. We develop these topics in the following.

In the section on symbolic medium, in Chapter 3, we talked about distinguishing the conceptualization of health as a symbolic generalized medium of communication and the body as a symbolic medium. Here we add a few more thoughts on those difficult but rather interesting themes. Perhaps the reader will find it useful to read those previous pages before continuing, because the following discussions start from there.

The body

Metaphorically speaking, we can say that the *body*, as symbolic medium, only appears in the health system as words, as symbols, and symbolic representations as meanings, because the health function system, like all the other social function systems, is made up of communications, nothing else. It could not be different. Likewise, diseases can only be part of the communications of the health system (be considered inside

the health system), if described in words and therefore communicated as meaningful semantic descriptions of forms distinguished in and attached to the symbolic medium of the body.

It is important to grasp these notions. The human body, as a concrete object in the environment of the health systems, does not enter the health system. The reader should keep in mind that medicine as sub-system of an essentially communication-based function social system exists in the realm of meanings. Only meanings – more precisely communicable meanings – make medicine. This is perhaps a tough statement to sell, but perfectly understandable within the Social Systems Theory. Only the symbolic body – in other words, only the meanings referring to the concrete bodies and their diseases – can be the object of attention and theme of communications in the health system. Indeed, bodies and diseases are concrete occurrences in the environment, in the real world – there should be no doubt about that – but in the communications about them, they still consist fundamentally of meanings, nothing else.[5]

We also need to bear in mind that communication has a comprehensive span, not only including the conversations a doctor has with another doctor or the conversations among health professionals. Communications can involve utterances in any health context but also countless written formats, such as records in patients' files (electronic or paper media), medical statements, results of exams, reports of procedures, assessments of health programmes.

Returning to our topic, as symbolic descriptions (expressed in words), the diseases can only be inscribed and related to an equally symbolic substratum (a symbolically represented body); that is, a meaningful representation of a disease has to be linked to and based on a

5 One could argue that medical professionals, maybe surgeons particularly, carry out actions not communications. But we need to consider that before, at planning stages, during the procedure, and after the surgery, there are lots of communications going on. Surgeons often consult with each other and interact with other specialists and auxiliary staff all the way through. We could say that if surgeons were forbidden to communicate, they would not be able to perform any surgery.

symbolically represented body. We are dealing with bodies and diseases in a universe of communicable representations, no matter how closely linked the meanings are to the empirical realities. While treating a patient, doctors are so close to the actual body and its ailments that they lose sight of the fact that they are communicating among themselves, and signs and symbols acquire reality-like materiality with references to textures, consistencies, shapes, weights, odours, and so on. But, still, the only way the health system can communicatively deal with diseases is through the meaningful descriptions of diseases in the symbolic medium of the body.[6]

Throughout the history of medicine, the body has remained a symbol, "container-like", so to speak, with all forms found in it and described in words as constitutive structures and functions; in such a symbolic medium, the diseases were inscribed and also "symbolically carved" as forms corresponding to observable evidence of dysfunctions.

To conclude, we should remember that the human body also appears as a symbolic medium in other function systems. For instance, in the system of arts it appears in association with representations of aesthetic values; in the political system, as possibly policy-relevant numbers, such as in armies, schools, housing, workforce; in the legal system, as granted and bestowed with rights (even dead bodies have rights); in religious systems, as having sacred attributes corroborated in prescribed rituals. All of these are examples of symbolic domains.

In Luhmann's characterization of medium and form, he explains that the constitutive elements are *loosely coupled* in a medium but *tightly connected* in a form. Because of that, a form can be perceived while a medium remains abstract, unperceivable as such. A medium is only

6 In her book *The Body Multiple*, Annemarie Mol (2002a) reports the findings of her research on the representations of arteriosclerosis and how it differs according to the orientation of the professionals. Surgeons, angiologists, clinicians, physiotherapists tend to describe different things when they refer to patients with arteriosclerosis, the identification and cause of the disease, and how they should be treated. This gives us a clear example of how health professionals are indeed communicating in the realm of meanings and symbolic representations.

perceived in the forms it acquires. A medium cannot be apprehended in itself.

These definitions create some difficulties for us when we try to characterize the body as a symbolic medium. The question that may arise is what are the elements that are loosely coupled in the body medium and tightly connected in the disease forms. To solve this conundrum we need to consider a number of things. Indeed, there are many concretely observable bodies everywhere. Those are neither medium nor symbolic. Regardless of the potential symbolism any of them may acquire in certain contexts, in their simple immediate existence they remain biological beings or just occupiers of space. In contrast with those, the symbolic body is a meaning, a representation referring back to the empirical one. As medium, such representation is full potential, an almost infinite reservoir of countless forms that can be configured in symbolic connection to it. As a symbolically "opaque and blurred intangible mass", the body can nevertheless receive and "display" precisely defined, but inevitably also symbolic, marks and structures.[7]

We can say that the *loosely coupled* elements in the body as a symbolic medium are all the components that medicine analyses in search of identifying the diseases afflicting patients. The medium includes all that can be seen in samples of blood, urine, faeces and other tests. It includes enzymes, hormones, diverse biochemical and organic compounds; electrical signals; tissues, and all anatomical features; multiple types of cells with their diverse functions; genes and DNA; antibodies; as well as everything that can be observed during external clinical examination and internal exploratory procedures and surgery. The human body is a hugely complex assemblage of a vast number of interrelated elements. These elements acquire particular configurations specific to each disease. A disease therefore is a *rigid coupling* of some elements in a peculiar composition – that is, a *form*. The elements are "loosely" coupled in the symbolic medium as healthy bodies have an excess of possibilities to continue be-

7 The operations, syntaxes, rules, articulations, and so on in the symbolic realm are matters of linguistics and semiotic studies.

ing healthy, with a large range of variations in all elements in the same individual as well as across individuals.

The gains from seeing medium and form in this way can be explained as follows. It helps, for instance, to follow the history of medicine as a continuous struggle to identify forms and form-constituting elements in the human body, and to enrich its symbolic representation, reflecting the empirically observed complexities. These understandings also help us to see the universality of medicine as built on the unique, singular, stable and permanent symbolic medium – fundamentally the same body wherever medical knowledge has taken root.

If we try to consolidate all that in a few short sentences, we can say that the symbolic medium of the human body is a massive repository of the complexities of all symbolically represented elements that can be observed in a healthy, sick or dead body. The health systems, as well as any social system, cannot address complexity as such, but medicine has developed selective approaches to extract from huge complexities only the designated relevant elements, and observe them in concrete bodies as deemed suitable, orientated by the symbolic representations it constructs itself.

The symbolically generalized medium of communication (SGMC)

We briefly talked about "health" as a SGMC in Chapter 3. Now we add a few points. In our contemporary society, a reference to health systems elicits many associations. In the health systems we can be professionals, patients or populations exposed (or not) to health risks. Outside it, as observers of health systems (researchers in the academy for instance), we may realize that something valuable for the society is generated, which does not refer exclusively to medical care.

In previous chapters we talked extensively about public health as a universe of meanings and communications in itself. So when we hear that a health professional (not only doctors) has expressed an opinion or given advice on any health matter (not only medical), that message carries the validity claim of being constructed within the health semantic universe of matters of legitimate attributions and responsibilities of the

professional concerned; that grants attention and predisposition to acceptance of the opinion that would not be the case otherwise.

The same opinion expressed by the pastor of a church, or a schoolteacher, or actor, or reporter, or sportsman, or policemen, would not be identified as a valid and legitimate health message; in these cases doubts could be easily raised and the message dismissed.

In this sense, health is a SGMC, the main characteristic of which, as Luhmann described it, is that besides being a medium through which messages can be conveyed, it increases the probability of acceptance of the message. Those entitled to communicate through the SGMC of health receive attention and recognition that other mediums of communication cannot grant for health matters of concern.

From time immemorial, suffering patients were – and still are – willing to follow any advice on what they must do, whether given by a traditional healer, a charlatan, a priest, a shaman, or a doctor. Suffering brings acceptance of help, no matter if eventually it proves to be deadly. Incontestably, in a direct doctor/patient interaction, doctors are granted space. But not only that; when we look to public health interventions, where the suffering may not yet be widespread or visible to the society, there is also acceptance of the assumed value of an intervention explained to the population. Mass vaccination is an example of that. COVID-19 vaccination campaigns are vivid memories for all of us. In fact, all preventive interventions can be analysed in the same fashion.

The issue here cannot be explained only by the fact that such interventions are matters of state power and decisions of political systems; we can see such a reception also in private health insurance schemes, for example, where the insurers see prevention as beneficial for the insured as well as for the scheme. The plausible argument communicated through the health medium (the SGMC) becomes more convincing. Imagine if the finance director of an insurance scheme or of a Ministry of Finance is the one who publicly makes the health arguments? It would not be as convincing as coming from someone seen as entitled to communicate through the health medium.

An additional feature of such a medium, as mentioned in Luhmann (2016, p. 130), is that a good medium for functional systems presup-

poses that the constitution of forms or conveying of messages, as for the SGMC, do not wear out the medium. In fact these medium are renovated. As a SGMC, health is renewed by new communications, which create new possibilities of communicative connections and their acceptance.

Importantly, the propensity for acceptance distinguishes the concept of the SGMC from the previous theme of the body as a symbolic medium. These explanations should have made the difference clear. While the SGMC is a medium essentially for communication purposes, the *body* as a symbolic representation is of epistemological relevance and constituent of medical knowledge itself.

Health as a SGMC is renewed and kept open for subsequent use as a recognizable conveyor of health messages with implicit validity claims; while the *body* symbolically represents the on-going accumulation of its actual and potential components.

Second-order observation

In "The Code of Medicine", Luhmann (2016 and 2023) acknowledged that *health* fulfilled all requisites to be considered a social function system. However, he called the *health system* the "system for the treatment of the sick". In that publication, Luhmann seemed to present a narrower notion of what health systems actually are. This is understandable, considering that in the late 1980s, when he wrote it, many critical health systems issues were not yet well grasped and widely discussed, and the term "health systems" was not generally used with the meanings of today. Nevertheless, it seems that he did not investigate deeply enough the extensiveness of "public health". He did not see the countless operations now performed as public health functions.

We can say that there are two problems with characterizing health systems using the words he chose: (1) More than "treatment of the sick", medical practices involve complex processes of identifying the sicknesses, treating them (for sure), but also preventing recurrences, avoiding severe occurrences, as well as considering predispositions,

prognostics and susceptibilities;[8] (2) Although not covered by the name he selected, prevention activities focusing on healthy people are also part of the same system. So the current commonly used name "health system" is better suited because it embraces all procedures and programmes taking place in the system (not only medical treatments).

There is a plethora of public health preventive interventions in which medical attention may be minimal or even absent. We can mention, for instance, occupational health (looking at reduction of work-related health risks such as exposure to chemicals, excessive noise, muscle strain and stress); sanitary inspection (looking at reducing risks related to expired, improperly stored or spoiled food and medicines, and so on); sanitation (such as looking at control of health risks related to water-borne diseases and sewerage); surveillance of disease vectors (such as mosquitos, rats, insects and poisonous animals); *one health*, as an integrated approach to human, animal and environmental health; epidemiological surveillance (monitoring society's profile of diseases, the trends, including violence and accidents, and outbreaks of epidemics); quality monitoring (for certification, licensing and accreditation of health facilities); monitoring of radiation risks (for eliminating health risks of users of radioactive materials, X-ray, and so on); community health and health education of communities (making communities proactive for reduction of health risks); and many others. Besides all that, public health also includes all the operations related to planning, implementation, monitoring and evaluation of health services and programmes, comprehensively covering all fields of health administration and decision-making at the level of organizations and health institutions. All these fields require multiple specific training and not necessarily a medical diploma. In all of them, the concerns in focus are the risks for populations, not the diagnosis and treatment of individual patients. Furthermore, and particularly, public health is exactly the component of the health system that makes possible the reflection of

8 In some circumstances, leaving a patient untreated can be justifiable, with such decisions being taken entirely within the scope of responsibilities of the health system.

the system about itself as a system. The term "treatment of the sick" does not comprehend any of these practices and horizons.

Nevertheless, we find in Luhmann's theory many valuable concepts for our understanding of health systems. The reader may recall the concept of "second-order" observation mentioned at several points in this book. The reader may also remember the two binary distinctions: healthy/sick and at risk/not at risk, which for our subsequent discussion we can refer to using the concept of "double coding", also found in many of Luhmann's texts.

We mentioned that the health system operates with two binary distinctions: the original *healthy/sick* code and the additional one, *at risk/not at risk*, developed later. Table 9.1 summarizes our views of these distinctions:

Table 9.1: Health systems' double coding and focus

	Healthy/sick	**At risk/Not at risk**
Individual	Medicine	Medicine
Collective	Public health	Public health

Medicine focuses on individual sickness as well as the risks and probabilities of an individual being treated developing, for example, side effects, recurrences, co-morbidity or vulnerabilities in consequence of the actual disease or related to/in consequence of the treatments they received. Public health instead focuses on the occurrence and distribution of health and sickness in populations (including ratios between sick and severely sick, such as malaria cases and severe malaria cases), and estimates the probabilities (risks) for populations – however risk can be defined: as exposing collectives to negative health conditions, diseases and effects. This double coding comprehensively covers all possible characterizations of attention and efforts deployed in either of the sub-systems of a health system.

With these distinctions, public health can perform its observations of populations and design risk-reducing procedures to improve the prospects, without treating diseases but just addressing risks and risk factors. This is the essential preventive nature of public health, looking for ways to prevent the human collective's health conditions from worsening.

In relation to the deployment of binary codes, repeating what was explained earlier, one side of the binary code leads to further communications within the system; it is the connective side, linking one communication to the next, while the other remains as the reference/reflection side. Within the medical sub-system, there is not much to communicate about a healthy individual; the communication then stops. However, within the public health sub-system, healthy individuals can be further distinguished with the deployment of the at-risk/not-at-risk distinction, at risk being the connecting side of the binary code. Central to public health concerns and communication is the identification of populations at risk and consequently the preventive actions to reduce the risks. The not-at-risk side remains the reference/reflection side.

We can say that public health often needs to perform "second-order observations", by which it considers what the medical side of the system has been doing. It goes without saying that doctors perform "first-order observations" on the bodies of their patients, while public health observes what medicine observes (observation of observers), and from that estimates the population's risks – frequency of diagnosed diseases, for instance, being a way of approaching risks. Although doctors also often perform second-order observations (when for instance they observe each other), public health performs broader analysis of trends in diagnostics and general outcome of treatments dispensed by the medical sub-system. With the reference to the attribution of sickness, but considering its own distinctions related to distributions of collective risks, public health can work on ratios such as sick population/general population; severely sick cases/sick population; population exposed to specific risks/general population; population sick due to exposure to specific risks/population exposed to specific risk; sick population get-

ting needed care/sick population; sick population cured/sick population treated; and so on.[9]

By performing "second-order" observations of what the medical subsystem does, public health can construct its measurements of risks, using the arsenal of formulas for calculation of indicators. For instance, public health can generate indicators such as efficiency, effectiveness, equity, performance and coverage, as presented in Chapter 6, characterizing and establishing attributes of the services made available, provided and received by the collective. In short, public health constructs social projections of services the medical sub-system generates.

As public health looks at the collective (including the healthy ones), not the individual, we can say that health systems are in fact more complex than Luhmann described in the mentioned publication. In practice, in any health institution, public or private, the separation of the two subsystems is very clear not only because of the differences in the codes they use but also the horizons of their observations. Every decision in the public health realm is concerned about reducing risks in some way; all decisions (all programmes, all initiatives, all implementations, all policies, and so on) are only justifiable on these terms. By doing that, public health can reflect on the health system as a system, contemplating it with the mission of reducing society's health risks and therefore, at some point in the future, perhaps eliminating diseases, or making societies free of diseases – that is definitely a health system's undertaking, not a medical task.

9 Following from the presentations in Chapter 6, most public health indicators can be interpreted as "second-order observations", because they correspond to observations of observations carried out beforehand by other observers. For example, effectiveness and equity are estimations often based on outputs and outcomes of services delivered on the healthcare side (medical) of the system.

Public health, self-reference of a scientific domain

In this section we reflect on the epistemological fundaments of public health. As a scientific field, medicine sits well within the broader domain of biological science; undeniably, human bodies are biological beings. However, public health has a complex place in the scientific realm.

As stressed throughout this book, what distinctly and essentially defines public health is the notion of risk. Risk is the central concern of public health. Risk as chances of occurrence of diseases (or aggravated conditions) in collectivities; risk as the probability of illness and illness-related occurrences among members of a collectivity.[10]

Risk is not a being, an object, an event or an entity; it is not a given phenomenon in the empirical world. Risk is a relation, a probability of "something" happening or not. If this "something" happens as diseases or worsening of diseases, the diseases themselves are to be dealt with within the medical domain (ultimately in the realm of biological beings), while the probabilities as indicators of threats to the collectivity have to be tackled as risk and public health matters of concern.

Risks are not observable as such, despite the fact that risk factors can be observed and must in fact be empirically established, even if tentatively. Risks can be estimated, calculated and used as guidance for prevention, but they carry indeterminacies and thus confidence intervals as any probabilistic estimates do. Both in hindsight as risks of past events or risks of future ones, risks are calculated chances, approximations based on the data available in the present.

We could say that medicine and public health have two contrasting ontological assignments. The health systems need both orientations. Medicine approaches an ontologically constituted entity: the human body with its concrete ailments. Public health approaches populations, looking for health risks to tackle. However, risks are not concrete and not as unquestionably present as the bodies and their diseases are. Risks are estimated with observations carried out in the past. The risk

10 We suggest returning to the sub-section on "Risk" in the "Concluding remarks" section of Chapter 6. That section should be read in conjunction with this one.

estimated in the present refers to the past; its projection into the future for justifying preventions does not embody an ontological reality; the predictions may be confirmed but may equally not be. On the other hand, the body of a patient cannot be dismissed as mistaken, even if the diagnosis and treatment were later proved to be wrong. The ontological status of the patient as such, no matter the accuracy of the diagnosis, is incontestable.

These two contrasting positions are relevant for any health system's practical concerns and decision-making. The health systems are concerned with both tangible and intangible entities. As observer, the health system constructs them based on the distinctions it deploys. However, on one side the observer has an object to focus on, and on the other side it has a presumptive "abstraction" it establishes, one that is always moving from the past into the future.

Risk factors – the elements and potential intervenient contributors or co-generators of disease occurrences – have to have plausible causal links to the diseases they presumably determine or aggravate. The causal links between risk factors and diseases may be direct, in the order of biological interaction between organisms or organisms and environment (toxic waste for example), or indirect, as facilitators of the direct causal organic interactions. Social determinants are not pathological per se, but may lead to increases in the chance of exposure and direct contact with pathogenic factors, or may prevent or reduce the chances of delivery of preventive and/or curative care.

From 2010, the WHO incorporated into its parlance the concept of social determinants of health, gathering non-medical factors under the banner of the "social". Broadly, the frame includes diverse types of factors. From the WHO webpage:

> The following list provides examples of the social determinants of health, which can influence health equity in positive and negative ways:
> – Income and social protection
> – Education
> – Unemployment and job insecurity

> – Working life conditions
> – Food insecurity
> – Housing, basic amenities and the environment
> – Early childhood development
> – Social inclusion and non-discrimination
> – Structural conflict
> – Access to affordable health services of decent quality.
> Research shows that the social determinants can be more important than health care or lifestyle choices in influencing health. (WHO, 2010a)

Unquestionably, this framework comprehensively covers many possible determining factors.[11] The WHO's concern with *health equity*, as expressed in the quoted paragraph, seems to be the basis for assembling and calling such diversity of factors "social determinants". Certainly, "equity" can only be addressed or referenced in a social scheme of references. Equity is essentially discernible as a social phenomenon of distribution of opportunities across population segments.

However, calling all the listed factors "social" is a stretch too far and mischaracterizes social attributes, which might be determinant of illnesses without primarily having "equity" effects. Many factors may affect all members of a given collectivity, leading to poor health without configuring inequities, but still remaining a social determinant and intervenient factor. Poor education and cultural misconceptions, for example, may lead to cases of preventable diseases regardless of the structure or stratification of the society. In short, one can see the *social* without necessarily looking through the *equity* glasses.

Nevertheless, we take inspiration from the idea of "social determinants of health" and consider the distinction between *biological/social* (or, with similar denotation, *medical/social*). In other words, *non-medical* means *social* – that is, sets of *non-biological* contributors to the occurrence of diseases – to be understood as intervenient *social* causal factors,

11 We could add others, for instance in the political field, such as lack of accountability and participation in health system decisions.

where biological determinants (that is, medically identified) are those affecting the individual organism, and social determinants affect the collectivity, presumably facilitating the biological connection. We can live with such understanding so long as we nevertheless recognize that with such use of the word *social*, countless phenomena, which are in fact addressed by specific scientific and technical fields, not necessarily covered by socio-science.

We understand that when we search for risk factors and we make incursions into the realm of social determinants, a whole universe of factors can be brought to the fore for consideration, as such use of the term *social* includes a huge range of variables in the fields of economics, culture, education, law, management, politics, religion, occupations, and so on. Ultimately, these variables shape the social factors leading to the *risk* of getting sick and/or the sickness being aggravated, but they have their own specific "causality mechanics".

The main message we want to convey and focus on is: we may say that public health is a scientific domain absorbing other disciplines' paradigms into its observational approaches. Emphatically, we may say it seems that public health would not be able to figure out its "object", *risk*, without deploying concepts borrowed from a plethora of paradigms and potential determination links found in other scientific domains.

As previously mentioned, risks cannot themselves be observed, as they are purely relational constructs. In themselves, risks are invisible, intangible. We can therefore say that *risk* has no ontological essence. In public health, *risks* are always defined, determined and constructed. Referring to risk and projection of risks into the future, Luhmann (2008b, p. 72) says: "Like the overall concept of probability, every measurement is fictitious and thus nonbinding – at any rate when it is a matter of statements about the future […] no one already lives in the future". They are constructed themes and cannot be themed without the presumed causal factors and causality relations.

To talk about *risks*, public health refers to the *biological* and/or the *social* causal plausibility. From diverse scientific domains, public health takes variables and their ontological character (as concrete objects, beings, events, measurable attributes, social entities, and as empirically es-

tablished variables). The borrowing brings to public health not only the conceptualizations but also the empirical approaches developed in the knowledge fields the variables come from.

In simpler words, economic factors, for instance, deemed to have influence over the causation of a disease in a given collectivity, are brought into the public health field of reflection with their empirically defined dimensions as actualities. The validity of the borrowed concepts and variables can only be questioned within the ambit of the science that established them. What public health can nevertheless do is reflect on the soundness and applicability of those concepts and variables for addressing risks and risk factors.

Taking a brief detour, the concept of *resilience* is a good example. Resilience was formulated as a scientific concept back in the 1970s, and was consolidated in the 1980s with the works of Dr Norman Garmezy and others, studying factors strengthening the mental health of children and adolescents living in stressful contexts. These were works in the field of psychology and psychopathology (see Shean, 2015). The concept was imported into social science, and from there brought into public health, arriving in the literature by 2015 (see Barasa et al., 2017), where it has been used to refer to survival of health systems in difficult contexts. The concept was seen as an inherent attribute of the systems, leading to increased chances of survival under severe constraints.

The importation of concepts brings with it the formulations elaborated in the science from which the concept originated. The question the science that formulated the concept answered included what, when, how and why something became resilient. In importing the concept, the science has to go through a process of assimilation and adaptation of the concept, endowing it with a distinct outlook. More specifically, in the public health domain it is necessary to ask questions and make the concept work in relation to reduction of health risks. It is necessary to argue whether a system showing specific signs of resilience (to be clearly described – impressionistic descriptions are not enough) also shows signs of reduction of health risks in the covered population. We must keep track of the risks.

If resilience is achieved with maintenance of behavioural patterns that do not decrease health risks, it might still be defensible in the territory of social or political science, but not in the field of public health. Of course, it all needs to be treated empirically. Resilience in itself is not an unbounded positive value. Qualifications need to be made because bad habits are also resilient.

We say here that there is nothing wrong in borrowing concepts, but the work that has not yet been done is to relate the concept of *resilience* with clear observation of risk reduction. Without linking the imported concept with the core public health business of reducing populations *at risk* (or preserving them in the *not-at-risk* position), those studies remain in the realm of social science or public management disciplines. So we might say that we, public health professionals, borrow concepts at our own peril.

Nonetheless, social determination of illnesses entered the modus operandi of talking and thinking of public health. The role of the social determinants of risks of getting sick (or not getting the necessary treatment) is attributable to the presumed causal links of the variables deployed in the study. The social determination is dependent on the attributions the study will try to observe, find evidence of and later on possibly tackle.

In contrast, the effectiveness of biological determinants is on more solid empirical ground. Public health interventions will need to, somehow and at some point, address the biological plausibility directly, checking the chances of success of collective prevention of diseases. If the water is contaminated and the communities are consuming it, the biological link of source and health damage can be established without resorting to social determinants; the same for preventable diseases for which vaccination should cover as many people as feasible regardless of social factors.

The dual order of causality, the *social* and the *biological*, requires the combined observation of the social and biological environments and the factors producing or contributing to the production of the diseases in "both" environments, so to speak.

At the end of the day, as we have been repeating ad nauseum, any public health indicator is concerned with risks, risk distributions and reduction of risks. The actions – better still, the public health interventions – will need to find justifications and feasibility in the dimensions of biological and social factors associated with the risks in focus. The justifications should show how improvements in indicators such as equity, efficacy, effectiveness, efficiency, coverage and disease profile are attainable by tackling biological and social risk factors. Keeping in mind that risks can only be tackled through such factors.

What we have done in this final section is what the Social Systems Theory calls second-order observation or, in other words, observation of observation – that is, how public health can observe itself as observer, performing therefore self-reflection. Such practice is the way forward for the development of public health self-reference.

To conclude, we have a few additional words. We assume that *biological plausibility* is optimally confirmed through clinical trials. *Social determinant plausibility* is confirmed through epidemiological studies. Clinical trials and epidemiological research can and often are carried out together. Epidemiology is the ultimate ideal calculator and estimator of risks.

In any public health intervention, even if only interested in, say, managerial (training of health district managers for example) or economic objectives (increase in the health budget for instance), and far away from the biological dimension, the outcome must in one way or another be measured and justified in terms of decrease of risks or risk factors.

As an example, let us consider managerial initiatives promoting models such as decentralization or adoption of horizontal hierarchical structures. These models, and their sociological fundament, economic rationale, or legal and political justifications, become incorporated into public health knowledge when links to reduction of risks are, even if tentatively, established. Otherwise they cannot be considered as belonging to the public health universe of meanings, remaining instead in their original field of knowledge.

This understanding has correspondences in all public health indicators presented in Chapter 6. All of them ultimately should refer to assessments of risk (or risk factors or exposure to risk factors, or all together), even when the calculation of risks is extremely difficult – such as when equity indicators are used to judge differences in access to healthcare among different segments of the population without specifying the diseases and risks involved. Even if not precisely measured, access to healthcare surely has effects on the health risks the population faces. Such statements can therefore be accepted as belonging to public health.

Sometimes it is necessary to use "proxies", such as lack of access to healthcare, as the risk factor when the diseases and their frequencies cannot be reliably specified. It can be reasonably assumed that difficult access to healthcare can indeed be risky for the collectivities experiencing it. In fact, public health works with estimations, guesses and assumptions that are justifiable when risks are incontestably present but cannot be calculated or unquestionably estimated.

The thing is, when comes to public health decisions, the total inventory of risks to be dealt with for optimization of solutions cannot be achieved in a definite and incontestable way. There are always other possible narratives, as our discussions in Chapter 6 about the indicators and their contingency showed. The maximization of reduction of risks across the board is too difficult to achieve or proved to have been achieved. The health system has to content itself with the feasible solutions that seem to have reasonable prospects of achieving significant risk reduction. It may not be proved to be the best or the optimum, but it is feasible and carries guarantees that benefits will be achieved.

When indicators related to organization, sustainability, financing, governance, environmental impact and others are used, the considerations of risks are indeed implied. Plausible biological links are in one way or another at least suggested or assumed.

It has been asserted elsewhere that public health publications hardly discuss the theoretical references they are based on. It is understandable that public health does not have its own theoretical basis and relies on theories, models, concepts and frameworks from which it borrows vari-

ables. However, this should not necessarily be the case. We believe this reflection is necessary to advance public health theoretical thinking.

Public health has to be more concerned with the issues of risk in a more explicit and focused manner, also understanding the challenges of decisions based on risk expectations, and the chances of occurrence or non-occurrence of the predictions. In this sense, there are risks in dealing with risks. A decision-maker deciding on the basis of foreseen future risks is bound by their decision and by the need to confirm the predictions.

> „Furthermore, on the social, normative and time dimension of risk, the decision maker may resort to norms by which decisions are formulated and acquire normative basis endowing it with extra probabilities and potential for achieving future results. Likewise scarce resources are allocated in ways of tackling the estimated future risks in a hopeful bet that the allocation is optimizing the use of available resources. All to be sensibly articulated and justified in terms of the aimed risk reduction." (Luhmann, 2008b, p. 69)

Will public health one day be able to develop theoretical themes of risk without resorting to other disciplines? This is a difficult question. At this point in time we can only say that prevention can be as close to the biological plausibility as, for example, vaccines are. However, all preventive actions will always require a degree of *social determinant* composition through which the collectivity under the "menacing cloud" of a specific risk can be identified and helped. The outlooks of the collective and its social determinants will somehow be "constructed" through its social, anthropological, cultural, environmental, economic, political, religious, legal and educational attributes. All social collectives have or can be assumed to have such types of attributes. Risks not only are detectable in such contexts of attributions of socially structured identifications, but also can only be addressed through those attributes when the social enters the consideration of the interventions to be implemented. There can

be no public health interventions that are not, in one way or another, social.[12]

Finally, we say a few words on the borrowing of concepts from other scientific fields. There is no problem in borrowing. Physics has borrowed – and still does – a lot from mathematics. But the scientific domain importing concepts should be aware of the borrowing and be critical about it. Otherwise high "interest" can bankrupt the business. The "imports" should be "exchanged" with "currencies" of the discipline "importing" the concepts. Humour aside, we need to be mindful about where the borrowing takes us, and do not lose sight of "risk", our core theme, which we have legitimate responsibility and role to preserve. When, for instance, we say that higher maternal mortality rate seems to be associated with societies with low income and education levels, we know that we are bringing from outside the health realm the social notions of income and education, concepts over which we do not theorize or claim historical authorship. We borrowed them; and self-reference tells us that there is nothing wrong with that. Although possibly one day the owners of the original concepts will raise questions such as: "this and that concepts are no longer what you say they are; we have already changed them". Then what shall we do?

12 For those interested in rich reflections at the intersections of anthropological and public health studies on health risks themes the book edited by Panter-Brick and Fuentes (2010) has a wealth of contributions.

Appendix

References

The consulted literature includes books in English and Spanish, and books on the history of medicine and on social and health systems. The titles are separated in two sections: the first has the titles mentioned in the book, and the second the recommended further reading.

Mentioned in the book

Balabanova, D. et ali. (2013) – Good Health at Low Cost 25 years on: lessons for the future of health systems strengthening, The Lancet 381 (9883).

Baraldi, C., Corsi, G. and Esposito, E. (2021) – Unlocking Luhmann, Bielefeld University Press.

Barasa E., Cloete K. and Gilson L. – From bouncing back, to nurturing emergence: reframing the concept of resilience in health systems strengthening, Health Policy Plan. 2017 November 01; 32(Suppl. 3)

Barona-Vilar, J. (2023) – Manual de Historia de la Medicine, Tirant Humanidades, Valencia, Spain

Bynum, W. (2008) – The History of Medicine: a very short introduction, Oxford University Press, New York

Canguilhem, G. (1978) – On the Normal and the Pathological, ed. D. Reidel Publishing Company, London

Costa J, (2011) – Is Tajikistan too poor for Health Insurance –, Medicus Mundi Schweiz Bulleting 120, reader of Swiss TPH Spring Sym-

posium on Improving Access through Effective Health Financing, Basel, Switzerland

Costa, J. (2023) – Health as a Social System, Luhmann's theory applied to health systems, an introduction, transcript Verlag, Bielefeld, Germany

Craig, R. (2021) – Bloodletting and medical practice, https://medicine.uq.edu.au/blog/2021/06/bloodletting-and-medical-practice)

Crosland, M. (2004) – The Officiers de Santé of the French Revolution: A Case Study in the Changing Language of Medicine, Medical History, 2004, 48: 229–244

Foerster, H. (2014) – The Beginning of Heaven and Earth Has no Name: Seven Days with Second-Order Cybernetics, ed. Fordham University Press, California

Foucault, M. (2003) – The Birth of the Clinic, an archaeology of medical perception, ed. Routledge, London

Fox, R. L. (2022) – The Invention of Medicine, from Homer to Hippocrates, Penguin Books, Milton Keynes, UK

Global Burden of Disease (GBD) Institute for Health Metrics and Evaluation (IHME); Available from https://vizhub.healthdata.org/gbd-results/

Jamison, D. et ali. (2006) – Priorities in Health, The World Bank, Washington D.C.

King, M. and Thornhill, C. (2003) – Niklas Luhmann's Theory of Politics and Law, ed. Palgrave, New York

Knudsen, M. (2012) – Structural coupling between organizations and function systems: looking at standards in health care, chapter 5 in The Illusion of Management Control, A Systems Theoretical Approach to Managerial Technologies, edited by Thygesen, N., ed. Palgrave, London

Luhmann N. and Fuchs S. (1988) – Tautology and Paradox in the Self-Descriptions of Modern Society, Sociological Theory, Vol. 6, No. 1 (Spring, 1988), pp. 21–37

Luhmann, N. (1990) – Essays on Self-Reference, Columbia University Press, New York

Luhmann, N. (1995) – Social Systems, ed. Stanford University Press, Stanford
Luhmann, N. (2007) – La Sociedad de la Sociedad, ed. Herder, Cuidad de México
Luhmann, N. (2008a) – Law as a Social System, Oxford University Press, Oxford
Luhmann, N. (2008b) – Risk, A Sociological Theory, Transaction Publisher, New Jersey
Luhmann, N. (2013) – Introduction to Systems Theory, ed. Polity Press, Cambridge
Luhmann, N. (2016) – El código de la medicina, chapter 6 in Distinciones directrices, ed. Centro de Investigaciones Sociológicas (CIS), Madrid
Luhmann, N. (2018) – Organization and Decision, ed. Cambridge University Press, Cambridge
Luhmann, N. (2022) – The Making of Meaning, from the individual to social order, edited by Morgner, C., Oxford University Press, Oxford
Maina, T., Mwai, D. (2019) – Tracking Health Resources Using National Health Accounts, in: Macfarlane, S., AbouZahr, C. (eds) The Palgrave Handbook of Global Health Data Methods for Policy and Practice, Palgrave Macmillan, London.
McLaughlin, E. (1941) – The Guilds and Medicine, historical background, Annals of Medical History, Philadelphia, US
Mol, A. (2002a) – The body multiple, ontology in medical practice, ed. Duke University Press, USA
Panter-Brick, C. and Fuentes, A. (editors) (2010) – Health, Risk and Adversity, ed. Berghahn Books, USA
Scott, J. (1999) – Seeing Like a Sate, ed. Yale University Press, USA
Seidl, D. and Becker, K. (2006) – Niklas Luhmann and Organization Studies, edited by Seidl, D. and Becker, K., ed. Liber and Copenhagen Business School Press, Frederiksberg C.
Shean, M. (2015) – Current theories relating to resilience and young people, VicHealth, Melbourne, Australia
Spencer-Brown, G. (2015) – Laws of Form, ed. Bohmeier Verlag, Leipzig

Varela F., Maturana H. and Uribe R. – Autopoiesis: The Organization of Living Systems, its Characterization and a Model. Biosystems 1974 5:187-96

World Health Organization (2007) – Everybody's Business: Strengthening Health Systems to Improve Health Outcomes: WHO's Framework for Action, WHO Geneva

World Health Organization (2010a) – Social Determinants of Health, https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

World Health Organization (2010b) – Monitoring the building blocks of health systems: a Handbook of indicators and their measurement strategies, WHO Geneva

World Health Organization (2022) – Health System Performance Assessment, a Framework for Policy Analysis, WHO Geneva

Further reading

Babini, J. (2017) – *Historia de la medicina*, editorial Gedisa, Barcelona, Spain

Baecker, D. (2006) – *The Design of Organization in Society*, chapter 9 in *Niklas Luhmann and Organization Studies*, edited by Seidl, D. and Helge, K., ed. Liber and Copenhagen Business School Press, Frederiksberg C.

Borch, C. (2011) – *Niklas Luhmann, key sociologists*, ed. Routledge, New York

De Savigny, D., Blanchet, K. and Adam, T. (2017) – *Applied Systems Thinking for Health Systems Research, a methodological handbook*, ed. Open University Press, London

Gargantilla, P. (2023) – *Historia de la Medicina, un recorrido por los hitos de la ciencia médica a lo largo del tiempo*, Pinolia, Spain

Knudsen, M. and Vogd, W. (editors) (2015) – Systems Theory and the Sociology of Health and Illness, Observing healthcare, ed. Routledge, New York

Luhmann, N. (1998) – *Complejidad y Modernidad, de la unidad a la diferencia*, ed. Trotta, Madrid

Luhmann, N. (2002) – *Theories of Distinction, re-describing the description of modernity*, Stanford University Press, Stanford

Luhmann, N. (2014) – *Sociología Política*, ed. Trotta, Madrid

Luhmann, N. (2015) – *Sociedad y sistema: la ambición de la teoria*, ed. Paidós, Barcelona

Luhmann, N. (2017) – *La Economia de la Sociedad*, ed. Herder, Cuidad de México

Moeller, Hans-Georg (2012) – *The Radical Luhmann*, ed. Columbia University Press, New York

Mol, A. (2002) b – *Cutting Surgeons, Walking Patients: Some Complexities Involved in Comparing*, in Complexities, Social Studies of Knowledge Practices, edited by John Law and Annemarie Mol, ed. Duke University Press

Ortiz, I. (Editor) (2017) – Historia de la medicina, atlas ilustrado, Susaeta ediciones, S.A., Madrid

Porter, R. (Editor) (2002) – Medicina, la historia de la curación, de las tradiciones antiguas a las practicas modernas, Lisma ediciones, Madrid

Rasch, W. (2000) – *Niklas Luhmann's Modernity, The Paradoxes of Differentiation*, ed. Stanford University Press, Stanford

Seidl D. (2004) – *Luhmann's theory of autopoietic social systems*, Ludwig-Maximilians-Universität München, Munich School of Management,

Seidl, D. (2006) – *The Basic Concepts of Luhmann's Theory of Social Systems* chapter 1 in *Niklas Luhmann and Organization Studies*, edited by Seidl, D. and Helge, K., ed. Liber and Copenhagen Business School Press, Frederiksberg C.

Sheather, J. (2019) – *Is Medicine still good for us?* Thames & Hudson, London, UK

World Health Organization (2009) – *Systems thinking for health systems strengthening*, WHO Geneva

World Health Organization (2017) – *World Report on Health Policy and Systems Research*, WHO Geneva

Yuste, C. and Arrizabalaga J. (2020) – *Eso no estaba en mi Libro de Historia de la Medicina*, Libros en el Bolsillo, Spain

Brief Glossary of Luhmann's Terms

System: Luhmann said: "system is the difference between system and environment" (Luhmann 2013, p. 44). A system is a result of this distinction, by which an observer can assign the place of observation either inside the system or in the environment. A social system does not appear without its environment. A system can internally consider its constitutive distinction, making for itself an internal representation of the environment that concerns it. Anything in the environment that is irrelevant for system's operations is not represented inside the system.

Autopoiesis: Imported from the works of the biologists Humberto Maturana and Francisco Varela (1974) and applied by Luhmann to Social Systems, autopoiesis is the condition of existence of social systems by which they produce the means for their own reproduction. A social system either performs its autopoiesis or does not exist as a system. Furthermore, no system can take care of the reproduction of any other system. Social systems are continuously orientated towards assuring the preservation of their self-reproduction. Autopoiesis is a permanent drive, performed at each operation of the system; 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 the element constitutive of social systems; the "lifeblood of the system". Luhmann's concept differs from the traditional transmission model of communication according to which "Alter" sends messages then received by "Ego", completing the communi-

cation link. In contrast, for Luhmann communication has three components: content (information), utterance and understanding. Utterance is the physical emission (sound vibration, visible printed characters, light signals, tactile braille marks, etc.), and content is the conveyed information (within a shared semantic universe of meanings). Understanding is the unit of content and utterance, and is what make possible the interlacing with the subsequent utterances and contents. Understanding includes misunderstanding. Communication allows for recursive confirmation (by which those communicating can go back to what was said and confirm or not the messages) and self-validation, interlacing past, present and future communications, stabilizing meanings and the systems that rely on them. For Luhmann, society is the totality of communications taking place, and no communication takes place outside society.

Contingent and double contingency: Contingent is said of something that could be different, i.e. it is neither necessary nor impossible. The term appears in relation to observations, selections, communications, decisions, etc., which, for being contingent, can always be different. All communications happen in a condition of double contingency, whereby both sides, while communicating, perform their own selections over what is said, listened and replied to. The selections each makes can be different and include the possibility of rejection of the communicated messages. For having the same possibility on both communicating sides, communications are therefore doubly contingent.

Closure and Operational closure: A system only deals with the information it internally produces. What is observed in the environment becomes information once the perceptions are selected and internally processed. Operational closure correlates with autopoiesis as the reproduction of a system happens within its operational closure. The selection and validation of observations and communications are prerogatives 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 otherwise possible, the boundaries of the two systems would collapse and the system/envi-

ronment distinction would no longer be valid. Operational closure however allows for a system's capacity to observe itself and also observe other systems. For simplicity, in a number of sentences in this book we shortened the terms "operational closure" and used the word "closure" alone, with the same meaning.

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 their communicative operations. *Function systems* communications are based on specific binary codes of communications (see next point); *organizations* (Luhmann, 2018) 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 does not exclude anyone in the society and anyone can take part in its communications at some point. Differently, organizations select those they identify as members (employees for instance); and only members can take part in decision-making communications. Interactions are communication systems of two or more people in face-to-face (or virtual) meetings; once the communication is over, the system ceases to exist. Society members may take part in several social systems.

Binary codes: Binary codes are essential for building *function systems*. Each *function system* has its own specific binary code (legal/illegal, healthy/sick, government/opposition, art/not art, etc. respectively for the legal system, health system, political system, and art system). The binary codes are fundamental to each respective *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* apart from the system of law. The words legal and illegal may be used in any circumstances but they will not carry the weight, legitimacy and consequences, if used outside a legal system. *Function systems* are thus based in simple binary codes that

nevertheless 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, while keeping their closure. In Luhmann's terms, systems organized under their respective constitutive closure do observe the others and by doing so "irritate" or "are irritated by" the others, creating expectations and reacting to other systems without losing their distinctive separation from them. *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 it, and in the process operate in a coordinated manner as seen from an observer's point of view.

Social differentiation: In Luhmann's grand theory, the current stage of evolutionary transformation of societies' structures is characterized since the eighteenth Century by the existence of several operationally closed and differentiated *function systems*. Differentiated *function systems* (law, politics, economics, health, religion, art, media, science, 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 initial primitive *segmented* society stage, individuals were assigned to social structures by their place of birth or life. In the subsequent more complexly *stratified* societies, individuals were assigned to stratus within the society (such as castes or noble versus peasants families, etc.). In the modern societies, differentiated in *function systems*, no function system has the central role and preponderates over the others; 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. Observation has blind spots, which are the deployed distinctions themselves. However, a distinction can also be observed, but that requires another distinction assigning the distinction to be observed to the marked side of the new distinction.

Second-order observation: Another important reference in Luhmann's theory is the work of Foerster (2014), who conceptualized the notions of an observing system observing observers and, respectively, the distinctions they use in those observations. This includes self-observation. All social systems perform first-order observation and second-order observations, i.e. observations of observers.

Coding and programming: Coding orientates the communication. Binary codes offer the two sides of the distinctions the communication refers to, while always electing only one side, the connecting one. Programmes guide communications by conditioning the selection of themes and semantics, supporting communication connectivity with the chosen side of the binary distinctions.

Complexity: Complexity is attributable to observation; it is not an object in itself. Both system and environment have elements and relations between elements that surpass system's observation capacity. Complexity therefore refers to the unavoidable system's limits for making observations and making sense of what it observes. A system reduces complexity by making selections and focusing only on certain elements, excluding

the rest from consideration. Contingency is intrinsic to complexity, because while reducing complexity a system makes selections, which could be different. The environment is more complex than the systems.

Requisite Variety: The systems do not have the requisite variety, i.e. the necessary diversity of distinctions, for internally making comprehensive copies of its environment in all its complexities. Therefore they make selections and focus only on those elements they can distinguish. Without the requisite variety, a system leave elements and relations unaddressed.

Decision: A 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. A decision connects to another and the subsequent ones in such a way that those making decisions do not need to go back to the basis and evidences of the previous decisions; thus, decisions become premises for those that follow from them. This connectivity absorbs uncertainties, as the new decisions do not need to address the uncertainties concerning the previous ones when they were made and justified.

Uncertainty absorption: A decision already made does not have to communicate the uncertainties that surrounded it before it was taken, the ambiguities or doubts about evidences guiding the decision. A communicated decision does not communicate that it is also contingent, i.e. 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 operations of any organization.

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 can make legitimate decisions, or, in other words, only decisions taken by members are recognized as valid and relevant for the organization.

Paradoxes and removal of paradoxes: In Luhmann's terms, all distinctions are paradoxical for being a unit of a difference. Dealing with distinctions and corresponding observations and communications, the systems need to remove paradoxes from sight, focusing on only one side of the distinction at a time, without attending to the contrasting opposite side. Paradoxes carry the risk of preventing determinability and thus the loss of connection between communications therefore one side needs to be privileged. For instance, in dealing with the challenge of allergies and autoimmune diseases, the provision of care needs to stick to the sickness side and remove the indeterminacy of considering that the body is having a normal reaction although against mistaken factors or disproportionately. The symptoms of sickness therefore have to be treated. Iatrogenic psychiatric diseases, with aggravation of symptoms over the course of a hospital admission, may be considered another of such example. Systems' self-reference is also an example of where paradoxes must be avoided (see below the paragraph on self-reference).

Re-entering: Re-entering happens when a distinction enters one of its sides. For instance, when a system produces internally a representation of the environment (always partial) from which it distinguishes itself through the system/environment distinction. On the system side of the system/environment distinction, the system establishes a representation of the distinction itself.

Medium and form: Mediums can only be observed through forms. Forms can only appear in the medium that makes them observable. Forms consist of tight connections of the elements of the medium, while these connections are loose in a medium.

Symbolically generalized medium of communication: These are symbolic medium linking motivations and selections. For instance, the symboli-

cally generalized medium of communication (SGMC) of power, money, law, love, art and truth, among others, provide pervasive references in their respective systems, facilitating specifications and acceptance of communications. The SGMC increases the chances of accepting communications that otherwise would be highly improbable. For instance, the SGMC of power increases the acceptability of decisions enacted and respective messages emanating from those holding positions of power and recognized legitimate use of the respective SGMC.

Meaning: In Luhmann's terms, meaning is the unit of the difference between actuality and potentiality. In other words, the medium of meaning allows the creation of forms that differentiate between the actual selection and the other possibilities. The meaning of a word is the selection of an actual possibility vis-à-vis the others that remain latent and potential for not being selected.

Complexity and complexity reduction: A formal definition says that complexity is the condition of having too many elements and relations so that the elements cannot be related to all the others. The environment is always more complex than the systems. The systems observe their environment and try to reduce its complexities by selecting the aspects (elements and relations) that the system considers relevant for its autopoiesis. This is complexity reduction. Systems do not have the "requisite variety" to relate each environment's element to an element of the system. In other words, it is impossible for the system to map and represent all elements and relations of the environment. Because of that, systems have to make selections, reducing the complexity of the environment the system has to deal with. In this process, progressively, the system also becomes more complex, by refining and developing new selections. But the system's complexity has to be controlled or even reduced to avoid overburdening the communicative operations of the system and its capacity to coordinate its own elements.

Systems' self-reference: Self-reference is in the confluence of interrelated concepts, such as autopoiesis, operational closure, observation, selec-

tion and communications. A system can only carry out its reproduction internally, and the reproduction entails reproduction of consistent and meaningful communications. A system needs to distinguish, recognize and validate communications belonging to it, separating them from the others. Validation of communications cannot come from outside, therefore self-reference continuously operates with the distinction system/environment. Self-reference makes possible the identification of meanings that make sense and therefore can communicatively be reproduced by the system. Self-reference is defined in contrast with hetero-reference, as two sides of a distinction. Autopoietic self-reference does not require the system to have an exhaustive, complete self-description and description of its environment. For the self-reference, the system recognizes its limits and distinguishes what belongs to it and what doesn't. Semantic codes are deployed with symbols and signifiers the system operates with. Mistakes may happen, but the system has safeguards to keep its self-reference updated and, adjusted. The system deals with the paradoxical nature of self-reference, where referring to itself is referring to itself referring to itself and so on to infinity. To stop eternal loops the system elects an identity of the self, establishing that no further exploration is required. The patient needs to be treated and the system knows which side it has to pay attention avoiding the entanglement of eternal self-reflection in its self-reference. System's identity is crucial for system's self-observation, self-description and self-organizing. For identity construction, the system/environment distinction plays the fundamental role; what does not belong to the system belongs to the environment. The system relies on its internal representation of causalities observed in the environment as interrupters of self-referential loops. In Luhmann's (2015, p. 99) words,[1] "the system de-symmetrizes itself". To stop tautological risks of its own self-referred loops, the system creates an asymmetry, internally referring back to the representations of the environment it created itself, recognizing assumed inherent causalities of the environment (for example

1 N. Luhmann (2015), *Sociedad y sistema: la ambición de la teoria*, ed. Paidós, Barcelona.

the diseases and their causes), preserving in the process its operational closure and autopoiesis. In simple terms, the health system not only treats the diseases with meanings created by it, but also treats them according to orders of causalities the system admits and confirmed it has observed in the environment. The history of medicine shows how new causal paradigms replace the old ones.

Structural coupling of communication and consciousness: Luhmann speaks about independence of the autopoiesis of communication and the autopoiesis of consciousness, although having necessary mutual fundamental connection. There would not be communication without consciousness and vice versa. But their reproductions are carried out independently. Utterances link to subsequent utterances in the medium of communications, as well as thoughts link to ensuing thoughts in the consciousness medium. Alter do not communicate to Ego a copy of Alter's thoughts. The utterances require selections of what to say. On the other side, Ego decodes Alter's utterances with his own selections. So, Ego's thoughts are not a copy of Alter's thoughts transferred by communication. Luhmann speaks of an *orthogonal* relation between communication and consciousness. Their coupling do not eliminate their autonomy; the opposite, coupling rather requires their independent autonomous performances.

Index