Social Indicators Research Series 70

Filomena Maggino *Editor*

Complexity in Society: From Indicators Construction to their Synthesis

Springer

Social Indicators Research Series

Volume 70

This series aims to provide a public forum for single treatises and collections of papers on social indicators research that are too long to be published in our journal *Social Indicators Research*. Like the journal, the book series deals with statistical assessments of the quality of life from a broad perspective. It welcomes the research on a wide variety of substantive areas, including health, crime, housing, education, family life, leisure activities, transportation, mobility, economics, work, religion and environmental issues. These areas of research will focus on the impact of key issues such as health on the overall quality of life and vice versa. An international review board, consisting of Ruut Veenhoven, Joachim Vogel, Ed Diener, Torbjorn Moum, Mirjam A.G. Sprangers and Wolfgang Glatzer, will ensure the high quality of the series as a whole.

More information about this series at http://www.springer.com/series/6548

Filomena Maggino
Editor

Complexity in Society: From Indicators Construction to their Synthesis

Springer

Editor
Filomena Maggino
Dipartimento di Scienze Statistiche
Sapienza Università di Roma
Rome, Italy

ISSN 1387-6570 ISSN 2215-0099 (electronic)
Social Indicators Research Series
ISBN 978-3-319-60593-7 ISBN 978-3-319-60595-1 (eBook)
DOI 10.1007/978-3-319-60595-1

Library of Congress Control Number: 2017946843

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction

Complexity and Its Implications in Dealing with Indicators

The quantitative observation of reality and the aspects defining it (like wellbeing, quality of life, and so on) requires a multifaceted approach and a compound methodology.

The usual practice, aimed at quantitatively observing reality, is to section and divide the observation in single elements called indicators.

The risk of using indicators is to consider reality like a machine, made up by elementary components. Since the world is actually an inseparable network of relationships and reality is a self-regulating system, the approach to indicator construction should respect those characteristics and require relationships, schemes and contexts to be defined and considered.

No indicator can be considered separately and independently from the others. Each indicator is important, but what makes the monitoring exercise meaningful is represented by the relationships that can be observed and analysed between and among indicators. The integrated view allows the phenomenon we are monitoring to be located diffusely in the system of indicators. In other words, a (social, economic, environmental) system is an integrated set/totality which can be understood only by examining the features of the whole, in line with the well-known saying *the whole is greater than the simple sum of its parts*.

The systemic approach and view find their theoretical reference in the theory of complexity, which applies also to the world of indicators. Complexity is actually a mathematical theory technically known as non-linear dynamics. Its application to the study of reality allows the understanding of the fundamental characteristics of social phenomena by using a systemic view, requiring the identification of relationships, networks and organisational schemes.

In dealing with indicators, complexity affects the following:

– *The construction of indicators*. Consequently, indicators should be many in order to, in a systemic view, focus not only on a single element. Each single element should be considered as an integral part of a variety of elements related to each other, defining that reality; no element or no indicator has an intrinsic validity in itself; consequently, if we are measuring a phenomenon, e.g. wellbeing, we should be aware that it is not located in one single indicator but is a global characteristic of the group of indicators, even though each indicator has a meaning in itself; moreover, the dynamics of the system do not require a rigid selection of indicators.

- *The analysis of the indicators*. The analysis should respect the non-linear relationships among indicators and require a multi-technique and multi-method approach; since indicators are actually mutually related (i.e. each indicator is influenced by the others and influences the others), the analysis should not produce a result represented by a simple and single number but should produce a meaning; the relationship between two indicators yields a meaning and produces new exchange, new contacts or interactions; in this perspective, the analysis in the ambit of a system of indicators allows the system to generate itself.
- *The interpretation of the results*. The systemic characteristic of the relationships among indicators requires a particular attention in the interpretation of the results obtained through the analytic process; the attention should be based on the idea that any increase of complexity introduces more refinements, fragilities and uncertainties in the statistical analyses.

In this frame, the synthesis of indicators plays an important role in:

- Reducing the complexity
- Allowing analytical processes to be conducted at higher levels of the system
- Allowing easier communication of the results

However, synthesis should not be pursued inconsistently. We should avoid aggregating many indicators, inevitably producing a meaningless value.

For this reason, complexity should be preserved in constructing, managing and analysing indicators and should guide in the representation exercise (telling stories through indicators).

The complexity approach should guide not only academic researchers but other actors like policymakers. In fact, this debate always points out that dealing with complexity shows challenges which are institutional, methodological, statistical and technical.

The Volume

This volume aims at disentangling some important methodological aspects and issues that should be considered in measuring complex social phenomena through indicators and in dealing with those indicators in order to construct syntheses.

Even though apparently dealing with these issues is merely a technical problem to be faced and possibly solved by statisticians or information scientists, the construction of indicators presents also other crucial aspects to be considered, starting from philosophical and political concerns. The ultimate success or failure of constructing and using indicators depends upon, as Alex Michalos pointed out in many occasions, the negotiations involved in creating and disseminating the indicators or the reports or accounts that use those indicators.

The volume has 13 chapters organised in four parts:

The **first part** is focused on *conceptual issues*.

Alberto Peruzzi introduces some important epistemological issues related to the notion of complexity by discussing how its increasing use in social sciences actually warns against excessive expectations and abuses of the notion resting on an appeal to rhetoric. He aims at providing a step towards clarifying the scientific meaning of complexity.

Rocco Sacconaghi proposes to disentangle the problem of how to synthesise analytical data and illustrates how the phenomenological approach can contribute to an effective interpretation of the relation between heterogeneous elements, moving from a list to a synthesis without causing an undue homogenisation of the elements themselves.

The **second part** deals with *methodological issue*.

Marco Fattore, in Chap. 3, raises an important issue related to the so-called information-based policymaking and the role that socio-economic statistics and indicators play in that context. In particular, he discusses the role of big data and data science on future socio-economic statistics and their potential effects on the construction of social and economic indicators.

The editor of the volume enters into the merits of developing indicators in Chap. 4, by stating the importance of having a systemic view and illustrating the challenge, needs and risks of this exercise. Chapter 5 deals with the methodological issues related to the synthesis in a system of indicators, by distinguishing also between aggregative and non-aggregative approaches.

Kenneth C. Land, Vicki L. Lamb and Xiaolu Zang, in Chap. 6, face important questions related to the construction of synthetic indicators: Can properties of a society described by synthetic indicators be scaled across time periods and levels of analysis – from the whole system to subunits thereof? Starting from the idea that indicators describe complex systems like societies, they address the question within the context of two general sets of equations of state for complex systems. The first complexity model is a non-linear deterministic dynamics model, defined by difference or differential equations. The second one incorporates stochastic (uncertainty) elements into the model specifications, leading to the various classes of statistical models.

The **third part** explores and investigates different *technical issues* related to the construction of synthetic indicators. In particular, three main approaches are illustrated.

Matteo Mazziotta and Adriano Pareto, in Chap. 7, illustrate the consolidated methodology aimed at constructing composite indicators, by including this in the worldwide movement (“beyond GDP”) aimed at identifying the best approach for measuring wellbeing and by paying attention on the pros and cons of this approach.

Marco Fattore, in Chap. 8, illustrates how the synthesis of indicators can be managed by using a non-aggregative approach, based on partial order theory. He shows how the application of partial order theory can overcome the limitations of aggregative approaches. The proposed method, in fact, is able to deal with indicators measured through data ordinal in their nature and measuring phenomena not necessarily correlated to each other.

Arranged together with Michela Gnaldi and Simone Del Sarto, Chap. 9 starts from the idea that one of the aims of measuring social phenomena is to identify, quantify and possibly explain the differences between units of analysis (individuals or countries) starting from some characteristics. In this context, the most applicable statistical approaches are those which allow us to deal with two different analytical perspectives, clustering (or grouping) units into homogeneous classes, by taking into account at the same time the multidimensionality of the phenomena under study. One of such methods is the latent class (LC) multidimensional IRT model.

The **fourth part** includes some *particular experiences* in dealing with the synthesising exercise by using also concrete data.

In Chap. 10, Chang-Ming Hsieh starts from considering how the synthesising process can account for potential societal, cultural and/or individual differences in values associated with different facets or domains represented by the considered indicators. Such topic is continuously discussed in the field of social indicators and is related to the use of a weighting system in order to reflect individual different values associated with different life domains. As illustrated in the chapter, the topic has important implication not only at the conceptual and methodological level but also at the technical level.

The content developed in Chap. 11 allows Ludovico Carrino to compare different strategies in normalising indicators while building syntheses. He discusses not only the rationale of the different approaches but also the consequences of their adoption in terms of results which show how the normalisation process can play a crucial part also in defining variables’ weighting.

Giovanna Boccuzzo and Giulio Caperna, in Chap. 11, illustrate the construction of a synthetic indicator through the application of the non-aggregative approach proposed in Chap. 3. The application uses official data produced by the Italian National Institute of Statistics and aims at defining a measure for life satisfaction in Italy. The application shows how the partial order theory can represent an important resource for social statistics.

Chapter 13, authored by Tommaso Rondinella and Elena Grimaccia, focuses on the big challenge of comparing different territorial areas according to multiple factors to be synthesised. The proposed solution, requiring the adoption of different statistical methods, is applied on indicators regarding the Europe 2020 indicators involving European countries.

The volume has no presumption to be complete and is not able to cover all methodological approaches and technical applications. The topic is continuously evolving and also expanding the boundaries of interest.

Other reflections could be added.

Some of them include representation (possible use of graphics in the perspective of synthesising in a complex context) and communication issues (how to obtain understandable data, and results, how to correctly present them).

Another topic to be more systematically explored concerns how to analyse indicators in a systemic view and context. This process requires particular attention: indicators representing a complex reality should be analysed through a systemic approach by considering from one hand non-linearities and multiple reciprocal causal relationships and on the other hand the uncertainties. The latter issue is particularly delicate especially in the perspective of defining possible futures.

In other words, our work is still in progress . . . to be continued . . ..

Roma, Italy Filomena Maggino

Part I
Conceptual Issues

Chapter 1
Complexity: Between Rhetoric and Science

Alberto Peruzzi

Premise

What is called "complexity theory" contains a core of ideas of scientific relevance, surrounded by features which, although intuitive, are in mutual tension. If complexity really marks a turn in our scientific image of the world, that core should not be burdened by vagueness and ambiguity, associated with features in mutual tension. Rather than relying on the intuitive idea of complexity as a sort of universal glue between different theoretical approaches worked out so far, caution is suggested. The reasons for caution do not imply any choice of specific theoretical framework, but their consideration is preliminary to any such choice.

This caution reacts to widespread rhetoric which ends in adhesion to a cult of complexity, and cult is not science. Although rhetoric may help bring a change of perspective, it obstructs development of a theory satisfying those standards of rigour and testability constitutive of the scientific method ever since Galileo: a method which relies on observational and experimental procedures in terms of quantities to test (the axiomatic presentation of) a theory expressed in mathematical language.

That method has various features. One is measurability of the quantities we talk about. For instance, consider the notion of "length": its measurement needs the introduction of units, which though conventional have to satisfy constraints (a rigid rod must be length-invariant under transport and this invariance rests on theoretical assumptions); data don't tell us how to interpret them, and when more than one interpretation is at hand, it is the theory we adopt which makes the difference; the consistency of data with a given theory (in turn supposed to be self-consistent) neither implies its truth nor is it a warrant of its explanatory power. It is plain that any

A. Peruzzi (✉)
University of Florence, Florence, Italy
e-mail: alberto.peruzzi@unifi.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_1

mathematical model omits some aspects of reality, but this omission can be justified by the gain in explanation and prediction it ensures. No matter whether we adopt a verificationist or a falsificationist viewpoint, or a probabilistic variant of either, such a method is more demanding than qualitative taxonomy and both more prudent and precise than any background "philosophy". Since the seventeenth century the growth of scientific knowledge has relied on this method and no other path has shown itself more reliable or effective. In principle, the new perspective focused on complexity, intended as a notion common to natural and social sciences, is no exception. But is this really the case?[1]

To free complexity from the "noise" of rhetoric, we need to examine the constituents of the notion, with their different sources, motivations and examples. If we miss the composite nature of the manner in which complexity is held to be crucial for distinct scientific domains, we risk turning up the "noise". This risk is witnessed by much of the flourishing secondary literature on complexity in social sciences, educational studies and philosophy. That this increase in background noise has been due to both scientists and scientifically oriented philosophers – who, in the garb of prophets, have led their followers to treat "complexity" as a mantra – is a further concern.[2] The new dialogue between science and philosophy, initiated in the early twentieth century, gave rise to what is presently named "philosophy of science" and it might have been expected to produce a different outcome. Thus, research on complexity demands an epistemological analysis unbiased by enthusiasm or hostility; and while no sophisticated meta-theoretical outlook produced the "complexity turn", logical rigour remains indispensable to assessing its import and relevance, as under that rhetorical noise (and blur) there is one of the most original and fascinating developments of our time, something involving mathematical sophistication and careful protocols to store/process the data pertaining to theoretical models.

Here, technicalities will be by-passed but this is not to suggest we can dispense with them. Their omission is due only to the limited aim of the present paper, that of a conceptual analysis to help clear away rhetoric and confusion found in much talk about complexity. Once detected, these can be avoided. Once avoided, the scientific

---

[1]Of course, an adequate account of the scientific method calls for many more details, including the recognition that its features combine in different ways for different subjects; therefore, it is hard to prescribe one and only one way of "doing science". Yet these preliminary remarks already furnish a couple of suggestions for research on "QOL-exity" (i.e., Quality-Of-Life-complexity), namely, (1) there is no direct inference from any plurality of data to "Synthetic Indicators", (2) the hypotheses used in collecting and organising the data should be explicit and crisply stated. If the study of complexity is part of science, it does not justify any shift to a new-style, computer-aided, inductivism (the rumour surrounding Big Data is a case in point), exactly just the appeal to "complexity" does not, by itself, imply a new framework for research in the social sciences, still less one ensuring unprecedented advances. It rather points to a theory of Synthetic Indicators as leading parameters of a dynamical system to be precisely defined.

[2]Systemic thinking does not need any particular "philosophy". The ability to take into account the consequences of a decision, together with their unintended feedbacks, was always recognised as an ingredient of rationality, of direct relevance to any strategy in politics, business and conflict.

content and explanatory power of theoretical models of complex systems in the social sciences can be properly evaluated.

Two Razors

From the first attempts to form a unified image of nature, the explanation of a given body of evidence accessible through relatively restricted data appealed to entities which were not part of the evidence to be explained, e.g., after Democritus, the existence of different kinds of atoms, or, after Aristotle, the existence of stable combinations of the four elements, each admitting continuous variation of density. Metaphorically speaking, the more kinds of such entities, the thicker the beard on the face of theory. Such Inflationary tendencies in the positing of metaphysical entities to account for empirical evidence provoked practice in the wielding of no less metaphorical razors. Two are relevant here. One was honed in Ockham (England), the other in Santa Fe (New Mexico).

First, the famous Ockham's Razor (OR): named for the birthplace of one of the greatest philosophers of the Middle Ages, though the version that became famous was one William of Ockham never used. That version in Latin reads *Entia non sunt multiplicanda praeter necessitatem*, whereas Ockham wrote: *Numquam ponenda est pluralitas sine necessitate*. It is the former version which became standard: paraphrased into English, it means that the entities we suppose to exist should not exceed what is strictly necessary to account for the data. While not the authentic version of Ockham, it remains a reasonable suggestion, recommending that the number of kinds of entities appealed to beyond evidence be minimised. Equivalently, the razor expresses the idea of maximising "ontological" economy.[3]

If we stay with Ockham's actual words, OR can be updated as follows: *Numquam ponenda est complexitas sine necessitate* – i.e., never posit complexity, unless necessary. Most scientists are familiar with this kind of razor. The value of simplicity was classically stated in Newton's *Opticks* as a rule of scientific research, one since shown to be fruitful, and strong motivations recommend it. The adoption of such a razor promotes an austere habit of thought though it is not so easily wielded in accordance with a precise and consistent methodology. Inherent within it is also a dangerous propensity to treat the idealised model of a system as reality.

The Santa Fe Razor (SFR) is seemingly cheaper and easier to use but in reality more tricky. To keep Latin as the language for razors, it might be rendered as *Numquam ponenda est simplicitas sine necessitate* – i.e., never posit simplicity unless necessary. The SFR requires no less careful management than OR. It takes its

[3]It is precisely from a consideration of OR that one of the best books on complexity and its models starts: Bertuglia and Vaio (2011). This is recommended reading, especially for those social scientists who intend not just to build models by exploiting the machinery of complex systems, but also want a synoptic view of different methodological options.

name from the Institute founded by George Cowan in 1984, which has as its motto "Science for a Complex World". It was at the Santa Fe Institute that the by now widespread view of the economy as a complex system was first promoted, dating from a conference held in 1987. The SFR shifts our attention from numbers (of theoretical posits) to interactivity, from saving to spending, and from issues of conceptual economy as in OR to the large variety of aspects involved in understanding *quality* by means of *quantity*. But, it comes with a symmetrical and no less dangerous propensity to that deriving from OR: to take the foam for the sea.

Much contemporary debate pivots on the rigid opposition between these two razors. What if we think of them as complementary rather than mutually exclusive? After all, one could say, the SFR prompts us to look at simplicity as an emergent property grounded in complexity and the "necessity" it refers to concerns the underlying process which allows for such emergence, through correlations between the components. In the absence of such correlations, it would be an instance of "disorganized complexity", to use Warren Weaver's words. Thus, one might think, the second razor makes the first possible. But this scenario seems immediately to tie us in knots, as the issue appeared to be deciding which of the razors to use in social sciences at the other's expense, but now the two razors complement each other. And if so correlated, how can they be opposed? To go back to Popper's dichotomy between two perspectives in the modelling of systems, "clocks" vs "clouds", we could even look at a clock as resulting from a cloud (of atoms) as well as at clouds as the result of many (say, asynchronous) clocks: this suggests a means to overcome the divide between conservative and progressive minds, namely, between those who deny complexity is a science and those who rely on it as the keyword to understand everything.

But then the price to pay for a (non-rhetorical) appreciation of complexity consists of having to sit on the edge of not one, but two razors, and we cannot expect such an uncomfortable position to recommend itself to either skeptics or enthusiasts for complexity. The *bonus* is that once both principles, associated with OR and SFR, are taken carefully into account, much of the rhetoric surrounding discussion of complexity is, if not banished, at least brought into view. The *malus* is that the resulting picture is even less clear. Against the "fast thought" so frequently advertised to save the effort of rigour, classical "slow thought" has its rights. Thus, let's start clarifying the terms of the opposition.

Back to Origins

Roughly speaking, a system is complex when it's made up of many interrelated components which, owing to their number and/or the number of their mutual relationships, exhibit properties which are not only different from those of any component but also *are* irreducible to (properties of) the set of components. Such irreducibility is intended as lying in the "things" and the way their whole behaves: a complex system is one with a behaviour hard, if not impossibile, to understand and to compute in terms of the components (be they elements or parts of the system).

So, complexity has from the start both an objective and a subjective aspect. Moving from intuition to the concepts of physics: the trajectory followed in the state space of the system depends on a particularly intertwined structure of its components and/or a very large number of simultaneous interactions among an equally large number of components. With a further concept-stretching (to use Imre Lakatos' term), we can add another feature: a small difference here can produce a vast difference everywhere – the "butterfly effect" leading to the land of chaos. Accordingly, such a system is supposed to be extremely sensitive to tiny input-variations. The exact conditions at the source of the flow of "information" through the boundary, i.e., the interface of the system with the environment, turns out to be an essential factor for any forecast (as a constraint, it is assumed that periodic orbits must be dense).

Finally, some complex systems are capable of self-organisation -which implies a capacity for the system to enter stable states conferring tolerance to the above changes. Granted that self-regulation in the presence of changing background conditions may lead the system to stabilise on unprecedented regimes, both (a) the idea of a self-regulating system and (b) the idea of non-hierarchical organisation are typically associated with the system being complex. Ad (a): if the system is the output of a program, as in the case of cellular automata, the program is supposed to be given in advance (differently of a neural network lacking any external algorithm of correction) and yet its behaviour can display behaviour we would call "complex". Ad (b): the emergence of macro-structural properties implies a hierarchical architecture which induces constraints on micro-structural dynamics, with top-down feedbacks contributing to the complexity of the system.[4]

Each of these different aspects has been formally articulated in areas of research as far removed from each other as e.g. game theory and meteorology. But different interests gave rise to different notions and reveal further aspects of complexity, which call for different methods. If, as someone also claimed, being "complex" were not intrinsic to a system but relative to its description and the method adopted to probe it, such conceptual fission would be inevitable in the presence of many non-equivalent descriptions and methods; and since a theory arranging the set of alternative descriptions and methods within a unifying picture allows to confer the concept a grip on an intrinsic property, the claim that complexity is language-laden prevents the existence of such a theory.

Two years after the Santa Fe Institute, the Center for Complex Systems Research (Urbana-Champaign) was founded by Stephen Wolfram. Many other institutes have

[4]This is just an anticipation of conceptual difficulties to be analysed in the following. For the time being, let me add that, were complexity only an "approach" or a "style of thought", there would be no reason to worry about such difficulties. There is a reason, i.e., they are obstacles to overcome, if we care for a definition of complexity that is cross-disciplinary, not so much to grasp a Platonic essence as to sustain the search for a set of axiomatic principles. That such a set is yet to be found is a challenge not to give up, and the appeal to a new "style of thought" does not relieve us of the task.

been created since then, mainly devoted to complexity-focused research and development. Today there is a long list of centres which promote research on complex systems around the world. The number of journals having "complexity" or "complex" in their title is impressive and many books, from advanced monographs to popular science, cover a wide range of applications. There are also dedicated series, by publishers with a worldwide market, centred on complex systems, not to mention a number of websites and mail groups; and the growth rate in the literature on complexity has been rising for many years. This is noteworthy, considering it was in 1987 when a reference journal such a *Complex Systems* (also founded by Wolfram) first appeared.[5]

The first issue (1995) of another journal, *Complexity*, had an opening paper by Murray Gell-Man entitled "What is complexity?". Twenty years later, notwithstanding the subject's growth, that question is still with us and has become more compelling, because of the many ways in which complexity has been approached in in dealing with a rapidly increasing number of topics.

To provide an idea of the manifold "ingredients" associated with the notion of complexity one might think of listing a set of formal features. Such a project assumes that a unitary meaning has been shown to capture the generality expected of the notion. This has not yet happened. An alternative strategy consists of returning to the origins of the interest in complexity.

This is the way adopted here, for both contingent and conceptual reasons: an adequate survey of research trends on complexity, just as a "rational reconstruction" of their development up to the present state of the art, is beyond the length of one paper; moreover, detection of the different ingredients of complexity is favoured by looking at the seminal works, in which the ingredients have not yet been superposed, so to say, and, ultimately, the issue about the meaning of the so-called "complexity turn" was already present in the original literature.

Many of the works marking the tipping point in the turn towards an increasing interest in complexity appeared between 1968 and 1981, in roughly a decade. Even when they took the form of collections of previously published work, their impact was amplified (an effect in line with ideas stressed in the works in question!) and actually the underlying ideas were not born in those years. For instance, some specific "seeds" of complexity go back to von Neumann and Ulam's work on cellular automata, while Schrödinger and Turing demand mention in view of their pioneering contributions to the analysis of the notion of living system; Heisenberg had also contributed with his statistical investigation of turbulences, and other names could be mentioned as well.

But the new picture, with its "style of thought", did not come directly from those who anticipated one aspect or other of complexity: their contribution was recognised with hindsight, after the change of perspective which matured in the late 1960s and came to the fore in a set of seminal works during the 1970s. These works or at least the names of their authors are probably known to any researcher on

[5]The *Physica* series of journals had already been enriched in 1980 by *D. Nonlinear Phenomena*.

complexity who pays any attention at the sources of the notions in use. What is mostly relevant here is that a minimal list of books is sufficient to make apparent the different ingredients which jointly inspired a generation and became "classic" references.

General Systems Theory, 1968, by Ludwig von Bertalanffy.
Towards a Theoretical Biology, in 4 vols., 1968–1972, by Conrad Hal Waddington.
Steps to an Ecology of Mind,1972, by Gregory Bateson.
Stabilité Structurelle et Morphogenèse, 1972, by René Thom.
Synergetik: eine Einführung, 1977, by Hermann Haken.
Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuations, 1977, by Ilya Prigogine (with G. Nicolis).
Autopoiesis and Cognition, 1980, by Humberto Maturana and Francisco Varela.
Observing Systems, 1981, by Heinz von Foerster.[6]

Since our present aim is neither a complete survey nor a critical review, still less a chapter on the history of ideas in the twentieth century, the rich content and sense of exploration those works convey cannot be given adequate discussion here. Nonetheless, some hints may be useful.

Already in 1945 Von Bertalanffy, a former member of the Vienna Circle, had sketched a cross-disciplinary investigation of any kind of dynamical system and introduced his readers to the study of the nested hierarchy of open systems in nature as the new frontier of science, searching for laws that apply to generalised systems, "irrespective of their particular kind, the nature of their component elements, and the relation or 'forces' between them".[7] Waddington introduced the notion of "epigenetic landscape", rather than assuming the environment of a system as an independent variable, and proposed a non-standard way of looking at the micro/macro relationships in Darwinian evolution. Unlike the other authors listed, Bateson was not a specialist in one field, in search of a wider perspective. He acknowledged himself to be an autodidact, with only Nature as his book, in Galilean style. Yet his freedom of mind had a contagious effect, reinforced by brilliant prose, and the unusual connections between the most different topics he led his readers to consider, when seen in the light of general principles, contributed to a heightened attention to systemic complexity.[8]

[6]Let me emphasise again that for the most part the books listed were collections of previously published papers, but the joining-up of ideas thus presented boosted their impact. (The entries in the list will be included, in standard reference format, within the bibliography at the end of this paper.)

[7]Similar suggestions, pointing to a cross-domain analysis of hierarchical organisation, came from Herbert Simon in the 1960s.

[8]Some years ago, on the occasion I met Nora Bateson, I referred to her father as the "Twentieth Century's Socrates". The original Socrates was not expected to provide his unlucky interlocutors with any theory and those who sold his brainstorming as a theory were simply cheating. Yet his questions made Plato's theory possible. Indeed, Bateson was criticised for his lack of step-by-step arguments and his dealing with epistemological problems in a non-professional way, but many

Thommade use ofanalytic methods developed in a branch of differential topology (singularity theory), and relied on his own classification of elementary "catastrophes" for systems governed by a potential function, as a decisive resource for understanding the most various natural phenomena, with an emphasis on biology. In the succeeding years catastrophe theory became a hot topic. Though its applications to social sciences met objections, the mathematical models had a penetrating effect and remain a source of essential tools for dealing with any process through which a small continuous change ends in a sudden transition to a different regime, corresponding to a global "qualitative" change.

According to Haken, who regarded laser beams as a paradigmatic example of self-organised coherence, the variety of types of self-organisation in nature can be covered by a small set of unifying concepts, pivoting on "the enslaving principle" through which global order parameters reduce the degrees of freedom of a system's constituents. Prigogine's research on the thermodynamics of open systems led him to describe"dissipative systems" (i.e., systems having stability far from equilibrium) as the centre of a new view of nature, in which chaos also held a place.

Maturana and Varela proposed a view of any living system, from single cells to societies, as an autopoietic machine, that is, as a "network of processes" of self-transformation. Accordingly, the components of an organism are no longer parts (organs), but rather processes such that "(i) through their interactions and transformations continuously [they] regenerate and realise the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they [the components] exist by specifying the topological domain of its realisation as such a network", Maturana and Varela (1980, p. 78).[9] Influenced by Bateson, this view resulted in what at Santa Fe was called as a "Complex Adaptive System" (CAS) by John Holland and Murray Gell-Mann, taking account of its feedbacks, and in fact research at the Sant Fe Institute was focused on CAS's as the key for unifying other research trends on complexity, see Holland (1995). The underlying characterisation of autopoiesis had in turn a strong influence on Niklas Luhmann's view of social systems.

Finally, the last entry in the list presents ideas which seem to have been conceived first: the research of von Foerster, linking "second order" cybernetics and biology, had been ongoing for about 20 years before converging with the growing body of work on complexity. He also established the Biological Computer Laboratory at Urbana-Champaign in 1960, which became a crucible of ideas on self-organising systems.

As for explicit philosophical commitments, both von Bertalanffy and von Foerster were originally influenced by logical empiricism, Waddington by the philosophy of A. N. Whitehead. Bateson, strongly interested in cybernetics and

popular books on complexity have appeared in the 50 years or so since he wrote and none have matched his mastery of style.

[9]This space has its own dimensions and complexity enters the scene with the interactions among processes. Some updated formulations of such an approach can be found in Petitot et al. (1999).

itsapplications to anthropology, had a radically critical attitude towards the epistemology of modern times as “autocratic”. No less immune to twentieth century debates in philosophy of science, Thom gave a neo-Aristotelean inflection to his emphasis on emergent qualities.

In fact, the general biological inspiration of most of the seminal works listed above prepared the ground for a progressive link with emergentism, as a general philosophy, and in particular with the diachronic structure of cognitive development according to Jean Piaget, to whose ideas explicit reference is made by von Foerster, whose work, together with that of William Ross Ashby, played a key role in connecting research on cybernetics with systems theory. But von Foerster’s philosophical views also separated epistemology of complexity from mainstream philosophy of science, as independently did Bateson and Maturana and Varela, who gave expression to a dialectical view of a scalable hierarchy of systems: since each of these is anchored to its specific inner space, their position is nearer to an idealistic dialectics than to one of a materialistic kind, since no property of an observed system S is independent of the observing system S' (the supposedly neutral ambient space hosting S and S' is affected too). Subsequent work of Varela with Evan Thompson and Eleanor Rosch contains an explicit endorsement of a Buddhist worldview, see Varela et al. (1991).

The same years in which these books appeared also saw the paper by Edward Lorenz: “Unpredictability: Does the flap of a butterfly’s wings in Brazil set off a tornado in Texas?”, Lorenz (1972),[10] which had a worldwide effect in drawing attention to complexity.

Further developments in the 1980s introduced concepts and tools which strengthened the framework for treating complexity and led to new fields of application.[11] Among the chief advances were Stephen Wolfram’s use of statistical mechanics to study cellular automata, see Wolfram (1984, 2002), and Per Bak’s “self-organised criticality” deserve mention. The latter notion (introduced by Bak together with Chao Tang and Kurt Weisenfeld), corresponds to the conditions in which “mass effects” become possible and a system, so to say, lives on them, rather than dying with them, see Bak et al. (1987).

Until recent years few professional philosophers of science have paid attention to these ideas which from the 1970s onwards were becoming a source of new models of nature. One of the first exceptions was Mario Bunge, who in the fourth volume of his *Treatise on Basic Philosophy*, Bunge (1979), sought to embed an ontology of systems into the logical and algebraic framework of formal semantics.

---

[10]Lorenz published no book, even one of collected papers. His name (and research) became known to a large audience through the brilliant science writer James Gleick. His book about chaos, Gleick (1987), presented a fascinating collection of case studies intended to point at a theory left to the reader’s imagination.

[11]A chronological map of the history of research on complex systems is in Baianu (2011), p. 23, which also provides information about authors and lines of research not mentioned here. Let me remind the reader that the few and cursory historical references in this paper are supposed to provide an introduction to epistemological questions about complexity.

Though complexity was not his focus of concern, the rigorous setting of his presentation shows how much care is needed when we talk of theories, models and systems. This lesson was not taken to heart in some of the cursory reflections by expert researchers in particular fields who lacked professional expertise in philosophy of science. The risks exhibited in such writings are illustrated not only in the work of past decades. There is still a persistent habit to think that the care put into one's own work within the field of one's particular scientific expertise can be dispensed with in talking about the meaning of such work or that of others. Such an unscientific habit seems widespread especially when complexity is the subject of discussion. Combined with the widespread lack of scientific education on the part of philosophers, this has made it difficult for a serious dialogue between scientists and philosophers on the topic of complexity to get off the ground.

Curiously, the above list of books offered synthetic and mainly informal expression to ideas about complexity so as to reach a wider audience, but they had a solid epistemological background, and one at odds with mainstream philosophy of science. This was true of *La Nouvelle Alliance*, a book published in 1978 by Ilya Prigogine (with Isabelle Stengers), see Prigogine and Stengers (1978), in contrast with later general books intended to communicate the meaning of the turn to complexity to a large audience.

Indeed, discussion of further aspects remained confined to journals. This especially concerns a source of conceptual seeding of the field we have already mentioned, namely the work of von Neumann and Ulam on cellular automata (the very term "theory of complexity" seems to have been first used by Christopher Langton). John Conway's algorithmic world, LIFE, is an excellent didactic reference, illustrating emergence and stability of configurations in a discrete space and time governed by a small set of simple laws. Wolfram's research in this area led him to advance a general and philosophically controversial view, whereby such an experimental approach to computation yields "a new kind of science" whose paradigm is how simple algorithmic rules give rise to complexity, Wolfram (2002).

In retrospect, we can identify which concepts from these listed works came to play the role of co-ordinate axes in research on complexity:

1. *Co-evolution* of a system and its environment (be it natural or social), together with the variability of constraints acting on large-scale processes.
2. *Emergence* of relatively stable systems through self-regulation, by taking their consistency with the laws of thermodynamics as a reference model.
3. *Non-linearity* as a recurrent feature in the dynamics leading to systems of increasing complexity. (So-called *bifurcations* are a major instance of such a "sensitivity" to small changes amplified into divergent lines of evolution, so that, when these non-linear state-transitions repeatedly occur, the dynamics of a system becomes predictably unpredictable.)
4. *Morphogenetic laws*, at work everywhere, in the growth of crystals as well as in visual gestalts.

5. *Sudden phase transitions* (loss of continuity for specific values of control parameters), for instance in case of "conflictual" states, through which a system passes from one pattern of behaviour to a radically distinct one.
6. *Attractors* of different shape for different systems, each endowed with its own set of attractor basins, in terms of which both von Foerster's slogan "order from noise" and the concept of top-down causality can finally achieve precise formulation.

These ideas were not independent of one another and a rigorous presentation might reduce the number of axes. Each was rich in meaning and its range of applications went well beyond borderline "case studies" that might leave our images of the world or of science unaffected.[12]

The widespread impression was that a new "paradigm" had been born. But what exactly? Notwithstanding that the original lines of research had different motivations and were elaborated on different theoretical backgrounds, they shared the project of identifying concepts which *transfer* from one domain of science to another. That such transfer turned out to be less smooth than claimed does not alter the fact that a path was cut across traditional boundaries between sciences: not in virtue of any "reductionist" ontology, guided by the idea all kinds of entities within a domain can be defined in terms of a single basic-level kind, but based rather on a cross-boundary language, possibly associated with an abstract theory of dynamical systems, in which patterns of complexity find expression.

An Analogy with Category Theory

Patterns of structure, order, and complexity across domains call for a theoretical framework suitable for dealing with domain-independent features, and these call, in turn, for a notion of *universality*, which is something more than independence of specific hardware.

In this respect, there is an analogy with the focus on universality characteristic of the category-theoretic approach to the foundations of mathematics. Formal details aside, this cross-boundary character is shared by category theory and "complexity theory"[13] insofar as what both intend as *universal* is the outcome of invariant relationships that are also *trans-categorical*. The categories are, in mathematics, those of each specific kind of mathematical structure (such as the category of topological spaces or the category of Abelian groups) and, in the study of

[12]Each provided a condition on a system and, jointly taken, the set of conditions could orient to characterise complex systems, though none of them could be taken as sufficient. Some of them might also be non-necessary. For instance, a system governed by a set of linear equations can salso how a kind of complex behaviour.

[13]Inverted commas are due to the manifold ways of approaching complexity which thus far have not reached unification. In what follows I prefer to use "complexity framework".

complexity, those of each kind (class) of dynamical systems. Such "universals" are correlated with the emergence, respectively, of more cohesive totalities with respect to point-like sets of entities.

On the complexity side there are instances of different systems with the same kind of dynamical patterns; on the category side are instances of different kinds of objects and maps with the same categorical properties. In the first case, we find one and the same type of attractor for systems composed of different types of materials; in the second, within the different models of a given mathematical theory we can identify a generic model, through which any member of the intended class of objects-and-maps can be "functorially" obtained.

The underlying worry is not new at all. Already in medieval times a term was introduced for notions having a similarly cross-domain status: they were called "transcendentalia" (in English, "transcendental" [entities], not to be confused with "transcendent" [entities]). The main difference is that such a notion lacked any consideration of the maps involved as well as any consideration of dynamical aspects – concepts which are central to the categorical and complexity frameworks. In philosophy, the meaning of transcendental concepts changed after Kant, but that lack remained. In a nutshell, philosophers detected a problem without being able to solve it. It is all the more striking when the solution comes from scientific research which was not originally addressed to that problem. But in order to provide a firm solution, it must be the outcome of a theory. Now, category theory exists while no complexity theory is at hand. But what if they could work together?

It is noticeable that in both the categorical and the systemic framework the source of unity across manifold domains calls for a new language: one which does not identify a foundational project in the traditional sense – it proposes no new kind of elementary/underlying/ultimate ingredients. In the mathematical case there is no longer a uniquely defined ontological hierarchy and thus the unification of mathematics does not issue from a *reduction* of any theory to one basic theory – say set theory; in the systemic case, the demarcation between social sciences and natural sciences is no longer sharp or deep as previously thought and the unification in sight is not the result of a finally achieved *reduction* of the social to the physical – rather, there is a recurrent set of dynamical patterns, exactly as there is set of structural patterns across different areas of mathematics.

Can this analogy be relevant to scientific explanation in natural and social sciences? In recent years the pioneering work of Robert Rosen in mathematical biology has come to be acknowledged, and the width of his view that attempted to unite emergent organic complexity with category-theoretic notions has begun to be recognised, see Rosen (1987, 1991). One can doubt if the class of systems defined with this aim in view ("anticipatory systems") will be adequate to the task, but the conceptual resources to improve on his suggestion have now become available.

On the other hand, the similarities listed so far may suggest more than we are justified to claim. Long-range order out of small-range interactions is not the same thing as structure-laws independent of the elements. Moreover, an attractor for a dynamical system is an end configuration in the set of possible states and may not be unique, whereas a generic model for a mathematical theory is initial, in the sense that any other model of the same kind factors through it, and it is unique up to isomorphism. Thus, unless we confine the analogy to a special subclass of systems, much work remains to be done to clarify the manner and the extent to which category theory can provide a fruitful setting for modelling emergent features of dynamical systems.[14] But, as already noted, whereas category theory is well-defined, complexity theory is not, being just a class of models divided by one common language.

Bypassing the Dual Language

We return to the question: what can we learn from the seminal works on complexity? There is a specifically epistemological point to emphasise: the "turn" related to the systemic approach to complexity appealed to quantitative language in treating subjects which were formerly thought to admit only qualitative characterisation. Since the "complexity turn" is often presented as standing in radical opposition to the previous tendency of post-Galilean science, it should be emphasised that its appeal to mathematical models is in fact a continuation of the Scientific Revolution. Insofar as mathematical models of complex systems provide an explanation of hitherto unexplained facts, this represents a further accomplishment of that Revolution, rather than its reversal.

It may be objected that this argument misses the point, for it overlooks the *radical difference* in the scientific image introduced by a focus on complexity. My reply involves two steps. First, motivations for the claims in the previous paragraph. Second, clarification of the problems related to such "radical difference". This section will briefly deal with the first step, the two following sections ("Some doubts" and "The complexity of science and the range of compositionality") will deal with the second one.

[14] A personal remark may be telling at this point. In the early 1970s, as a student, my interest in complex systems was sparked by one of the most open-minded Italian physicists of that period, Giuliano Toraldo di Francia, whose courses in Florence treating foundational problems of physics I had the good fortune to follow. He gave me the opportunity to meet Prigogine and Thom. My research centred on topics, such as models of semantic cognition, which then appeared distant from applications of the complexity framework. But I was already searching for a general setting for the "cross-boundary" universality mentioned above. I found it in category theory. However the link between category theory and dynamical systems theory was at that time unclear to me, I later became aware that Bill Lawvere had already developed a categorical approach to dynamical systems Lawvere and Schanuel (1986), and that Rosen had explored applications of category theory to biological systems. Recent papers on category-theoretic treatments of dynamical systems have investigated issues of complexity from a perspective which draws on Rosen's work. Further advances are in prospect. But so far proposals to bridge the two theoretical frameworks remain too generic to provide an insight into specific open problems, or too tied to the study of particular systems which appear of little relevance for social sciences. This picture may change. If so, it will provide further evidence that my theses in Peruzzi (2006) apply to the *emergence* of cognitive patterns. See section Emergence below. One of the first papers in this direction was Ehresman and Vanbreemersch (1987).

From the seventeenth century Scientific Revolution onwards, the language of the natural sciences became progressively more mathematical. But the steady expansion of the appeal to quantities was accompanied by a recurrent debate: can a purely quantitative account cover every aspect of reality? The progress of physics (and chemistry) was made possible by the adoption of a formalism centred on the role of measurable quantities (supposed to be continuous and additive) subject to operationally defined standards in their application. Its success was so stunning that many scientists were led to identify the *real* with the *measurable*. Nonetheless, the search for (and characterisation of) *causes* in biology, psychology, economy and sociology remained confined to qualitative language. Anyone convinced that qualitative aspects are intrinsic to the characterisation of life, mind and society faced a dilemma. If such aspects are law-governed, either the laws can be expressed in quantitative terms or the whole framework of quantitative inquiry is conceptually insufficient, thus, if knowledge of such subjects is possible, our method of investigation has to be radically different.

This dilemma remains with us in social sciences, and also in psychology the debate has never ended. Something has changed, however. Quantitative methods have entered both fields and if a qualitative assertion admits an empirical test, the test is designed in quantitative terms.

There was an underlying tacit hypothesis common. Namely that *the use of quantities is committed to linearity, additivity, compositionality, all supposed to be reducible to the interactions of sharply identifiable (if not pointlike) constituents of a system*; any cause can be accordingly factorised and its effects are uniquely determined and hence predictable (in principle, with certainty). Determinism and predictability were strictly linked, if not simply confused with each other.

The rejection of such underlying hypothesis in the works mentioned above might have been expected to pose a fundamental challenge to the conception of irreducible qualities, since what the authors of the above seminal works were proposing was an image of the world strongly dependent on the study of variable quantities. What actually happened was different. The complexity framework was frequently taken as proof of the limits of quantitative science. But if qualities become *emergent quantities*, the dynamics which makes their emergence possible is expressible in equations just like the law of falling bodies or of electric repulsion. Thus, rather than bringing rhetorical slogans in support of the complexity framework or sarcastic remarks to dismiss it, the discussion would have benefited from a precise methodological analysis of the role of the two razors and their interplay.

What is at stake is more than an issue of language. The complexity framework by-passes the classical Quantity vs Quality debate, in a way reminiscent of the thesis of the transformation of quantity into quality familiar from nineteenth century dialectical materialism. But with this difference: the transformation is now formulated in precise mathematical language – that of the theory of dynamical systems. While not all aspects of dialectical materialism are preserved in this setting, it is nonetheless curious that the traces of that philosophical position have almost completely disappeared. This is strange when one considers that the complexity framework effectively undermines the fundamental opposition of "Quality" and "Quantity" which had been a recurrent theme in metaphysical thought in the Western tradition until challenged by dialectical materialism. To suggest how various have been the shapes that this opposition has taken, we mention some of them here.

Aristotle: mathematics is not suitable for the investigation of nature.

Knowledge of nature pivots on the essential qualities of any being: everything in nature tends to an inherent goal, associated with such qualities. The very motion of a body signifies that this process is ongoing – and motion is qualitatively sharply distinct from rest in an irreducible way. Moreover taxonomic classification is the paradigmatic task of science, whose aim is to assign every entity its place in nature – its proper slot in the great (and static) cabinet of Being. Equivalently, "real" definitions designed to capture essences are the aim of inquiry.[15]

Descartes: the mental is beyond the range of science.

Rather than through (qualitative) subject-predicate logic, the grounds of rationality take the form of algebraic equations. Science is a rational endeavour based on mathematics, and mathematics is essentially geometry, and geometry investigates quantities that can only be defined for something "extended" in space. Minds are not "extended": mental properties, concepts, thoughts, arguments and any other feature such as belief, desire, hope, have no length, area or volume. The conclusion, by contraposition, is straightforward.

Eddington: the scientific and the perceptual account are in mutual contradiction.

When we sit at a desk, what is in front of us is not one, but two things: the perceptible desk, as a full, rigid, piece of wood with a smooth, flat, surface of, say, brown colour, i.e., the macro-table, and the desk as described by physics, i.e., as a set of atoms bonded together, with the whole volume mostly made of empty spaces between atoms in constant motion and a surface which, if considered at the microphysical level, is not flat at all (and colour is not in the thing, but in which wavelengths of light are not absorbed).

Husserl: the quantitative worldview has led us to confuse nature with its mathematical models.

The natural sciences reduce anything to measurable objects. With such objectification, the subjective aspects of experience disappear, together with the constitutive process through which we assign meaning to what we say about nature: we become things among other things. In order to recover the very sense of scientific investigation – a sense not itself belonging to nature – a

[15] In Popper's view it was the persistence of this idea in the social sciences which was mainly responsible for their backwardness in contrast to the natural sciences. Note here, however, that if "complexity" denotes an essential quality of a system, Popper's distinction becomes blurred.

different kind of investigation is needed, one directed at the subjective roots of the very ideas of object, nature and science. But to identify these roots we have first to suspend, or "to bracket", beliefs and existential assumptions and such "bracketing" paves the way to focus on pure phenomena and their essences. This is the task of *phenomenology* and it must be free of any quantitative method, on pain of a vicious circle.

Wittgenstein: quantitative language is just one among others.

We are involved in a plurality of language games, none of which is entitled to primacy. Ordinary language is not one but rather a plurality of language games: the womb of every specific, technical, formalised, language, and in particular of any language dealing with quantities. Each specific language is legitimate as any other, provided the context of its practical use is made explicit. There is no master context and any socially shared context is equally valid.

All these views were ignored in the texts mentioned above (section "Back to origins") though in different ways. If we look at any object around us as a relatively stable outcome of an underlying dynamics governed by principles of self-organisation, and if we apply that conception to the mind itself, the oppositions stressed by Aristotle, Descartes, Eddington and Husserl all break down, and Wittgenstein's juxtaposition of a plurality of language games only concerns end-products. A new frontier for science is before us. This is the new frontier named "complexity".

The problem becomes how to re-conceive the relationship between Quantity and Quality. In the first place, the opposition is no longer static: qualities emerge from a dynamics in which only variable quantities are involved. For purposes of explanation and prediction a "qualitative" dynamics may be sufficient in many cases, as for instance, when the positive, null or negative derivative of a function provided the needed "qualitative" information on the evolution of a system. The exact values of the quantities involved are irrelevant to the dynamical behaviour in certain regions of the state space but may be extremely relevant in other regions due to non-linearity. This too is something to be explained in giving an account of the emergence of certain kinds of qualities from certain kinds of quantitative change.

Some Doubts

Ontology concerns "what there is" and in recent times ontology makes use of systems theory, but the dynamic view of anything real, a view which is part of contemporary science, asserts that "what there is" is a (possibly stable) outcome of many different, ongoing, changes. This step from statics to dynamics in treating the distinction of Quality and Quantity calls for precise examination of its meaning. But instead of efforts to make the meaning precise we frequently find mere rhetoric, opening the door to conceptual confusion.

Confusion can remain even when rhetoric is avoided. The recurrent idea of complexity as at the centre of a circle of so many topics that there will be "something for everyone" invites such confusion. Its acceptance spells the end of logic, which if it does not immediately bring the end of science, certainly prepares the way. *The End of Science:* so is titled a book by John Horgan, published right after his editorial "From Complexity to Perplexity", which appeared on *Scientific American* 20 years ago, see respectively Horgan (1995, 1996).

In that editorial Horgan expressed a negative attitude to complexity in asking "is the 'New' Science of Complexity more than soup yet?"(More politely put: the idea is fine, but what are the results?) while in the book he depicted contemporary scientific research, when not a mere a list of footnotes to "Newtonian" science, as a case of dubious literary criticism, or "ironic science". Horgan's paper fuelled a harsh debate. Its very title was laden with sarcasm, whereas that of his book carried a more dramatic tone. The mention of both is meaningful here, as those who reacted to the paper did not coincide with those who reacted to the book.[16]

Horgan presented complexity as a viral marketing campaign, which by means of a sequence of assertions of predictable unpredictability, nested like Chinese boxes, threatens to end science. Twenty years later, the growth of knowledge about complex systems deserves neither sarcasm nor drama. The use of the notion simply urges care, prudence and theoretical rigour when we take complexity as a source of scientific explanation. Horgan's call for a precisely defined concept of complexity is here endorsed. His complaint that complexity violates the "Newtonian Paradigm" as the royal road of science is not. We may note in passing that Quantum Mechanics already looks non-Newtonian, yet cannot be seen as "ironic science" considering the fact that it predicts precise effects made *certain* by the *uncertainty* principle. Thus there are two problems before us:

Problem 1. Which facts are *explained* in terms of complexity that could not be explained otherwise?

Problem 2. What *predictions* does complexity offer which could not be predicted otherwise?

An epistemologically informed reader will immediately spot a further question, preliminary to Problem 1, namely which model of explanation to adopt? As for

[16]On the conviction that (a) the achievements of science left about nothing substantial to be explained, Horgan claimed that (b) science is by now becoming postmodern, with complexity as part of this mutation. This claim was subsequently shared by some of those who advocated for (c) complexity as the new paradigm. Eminent proponents of (c) shared his idea that (d) mathematics is no longer the land of proof in announcing that (e) the deductive method has to be replaced by wide-ranging experimentalism through engineering of mathematical models. So far, none of (a)–(d) has been convincingly argued; rather there is evidence against each of them. In addition, there are doubts about the consistency of the conjunction of the four claims.

Problem 2: what do we mean by claiming that a complex system is "predictably unpredictable"?

Suppose the choice is, for brevity's sake, between *a* causal and *a* deductive-nomological model of scientific explanation (each admits variants, hence the indefinite articles). The very distinction between explanation and prediction relies on the asymmetry between past and future, which mirrors the intended difference between cause and effect. But if we adopt a deductive-nomological model the distinction fails (logical deductions involve no time-parameter). This might seem a side issue to the scientific status of complexity. Were it so, the matter would be simpler (in a Pickwickian sense). But it is no side issue: first, if the emergence of a macro-pattern out of a dynamics in which a huge number of micro-states take part contributes to the establishment of time asymmetry, this can be used as evidence against the deductive-nomological model; second, if both self-regulation of a system and co-evolution with its environment are considered, the feedback from global to local structure gives rise to mutual dependence, expressed in terms of equations which are time-symmetric (e.g., Volterra-Lotka equations for prey-predator dynamics, when the food supply for prey is constant).

No matter which model of explanation is chosen or how "predictable unpredictability" for an agent planning to intervene within the system is defined, any jointly consistent answers to Problems 1 and 2 must select some such option.

Such minimal awareness of the background epistemological issues is already a non-trivial matter when the task is to explain/predict emergence – and the relative stability of an emergent structure – for different kinds of systems. But then general questions have first to be addressed, such as "How is it possible that order emerges without violating fundamental conservation laws?" and "How can such an order endure, rather than being swiftly re-absorbed by a lower energy state?" In answering these questions, statistical considerations turn out to be necessary. Yet they cannot explain *why*, for a given system, one and the same set of order-types returns, rather than another. Can this explanatory gap be filled?

Replies to Horgan's paper soon appeared, some arguing that the gap can be filled with the help of the SFR (the search for hidden order where none seemed to be present).[17] In hindsight both the target of criticism and the position to be defended seem unclear, owing to the ambiguous reference of "complexity": now the general perspective of dynamical systems theory, now "non-linear science", now self-organisation. No doubt, these are all supposed to be aspects linked with each other within the complexity framework. But how exactly? Neither their intersection nor their union coincides with a unique definite theory. One sees here the rationale for the widely-endorsed view of complexity as a cluster concept.

For comparison, consider computability. It also found more than one formalisation (starting with Gödel's recursive functions and Turing machines) but

[17]As an example of these replies, let me recall Melanie Mitchell's paper, submitted to *Scientific American* but unfortunately not published there, see Mitchell (1996) and chapter 19 of Mitchell (2009).

they turned out to be equivalent in power, whereas the different notions of complexity, once precisely specified, turn out to be inequivalent. *Which* complexity-model is meant thus matters when an explanation, possibly the "best explanation", in terms of complexity for the behaviour of a system is our goal; and presumably also when a statistically grounded forecast about the future state of that system is made. Therefore, Problem 1 and Problem 2 can have definite answers only *provided* the cluster view of complexity is replaced by a meta-theoretical analysis, which presupposes core principles, together with cross-domain definitions. Only on this ground is an empirically guided selection between different notions of complexity possible. (Since such a core definition admits different completions, it ramifies.) The conclusion is that the success of the complexity framework in social sciences will turn on the rejection of the cluster view.

The Complexity of Science and the Range of Compositionality

Setting aside previous misgivings about the state of the art and the current lack of a rigorous theoretical setting at the required level of generality, the relevance of complex-systems-models concerns not only specific domains of scientific investigation but our understanding of the very "dynamics" of the formation and growth of scientific knowledge itself.

According to Thomas Kuhn, the historical development of science is an alternate sequence of states of relative stability ("normal science") and states of instability during which various proto-theories, originating from "anomalies" in the received view, are advanced undermining theories hitherto accepted (tacitly or unconsciously) as parts of one "paradigm". If two or more proto-theories develop, they clash and the collision ends with a winner, after which the field of inquiry settles into a new regime of relative stability.

Kuhn's idea, expressed in his 1962 book, *The Structure of Scientific Revolutions*, sparked much debate and criticism and was subject to various corrections and refinements in the following decades, but the main lines of his argument are now part of the "received view" in philosophy of science.[18]

The cycle of normal science and scientific revolution is supposed to recur time and time again. On inspection of the initial and end states of each cycle, the change is so radical and so great as to amount to a global re-organisation of knowledge. A "scientific revolution" has occurred (usually in a relatively small lapse of time) involving a "gestalt switch" sometimes compared to that seen in religious conversion. Not only are different principles involved, but typically principles expressed by means of a new system of concepts and a new language. All this is presumed to

[18]For other remarks on Kuhn-style epistemology in connection with complexity, see Bertuglia and Vaio (2011) chap. 4.

have overthrown the view of scientific knowledge as undergoing a process of cumulative growth, in which translations of previous languages/theoretical frameworks into new ones is always possible, so that at least part of the old view can be recovered within the new one.

Some aspects of this kind of process are well-known in dynamical systems theory and they are among the central topics of complexity. Thus we could exploit this framework to get a model of the history of science, e.g., as an epigenetic landscape (to use Waddington's terms), or a manifold punctuated by critical bifurcation periods, a model going well beyond an exclusively qualitative description of a recursive cycle of extended "normal science" periods (almost flat regimes) and short critical periods leading to a clash of world-views (turbolence up to a chaotic phase) followed by a jump into another "normal science" regime.

At best, the resulting model would *describe* the overall dynamics of scientific change. But would the model *explain* the change? This question pinpoints a worry that goes beyond the case under consideration.

In order to be more than a suggestive metaphor, such a model for the history of science requires a well-defined parameter space. If the set of parameters varies in time, we are (once again) in need of an underlying theory of such parameter-changes. At present we are far from having such a theory. Nevertheless, if the "complexity turn" is that revolution in the scientific image as claimed, an argument in its support would be the one just referred to, namely that complexity-based model of the growth of science would permit self-application.

This leads directly to a further question which the adoption of the complexity framework in social sciences poses. If Kuhn is right and we face such a revolution in sociology of science, it will presumably be impossible to retain the existing sociological knowledge (Kuhn's view included) within the new "paradigm". If the complexity turn indeed embodies such a revolutionary shift yet permits the preservation of at least part of existing knowledge in its intended meaning, Kuhn must be in error. But in that case support for the complexity turn based on its applicability to history of science falls away.

One might reply that since the complexity framework offers as yet a set of (overlapping) models, not a unitary theory, we are still at the stage of the clash of proto-theories, so that in exploiting the notion of complexity we must be clear which one we choose. Therefore the conundrum posed by the self-application of the complexity framework can be postponed. It identifies an epistemological question which can be by-passed in the ordinary practice of research. Now, suppose that, as Einstein said, scientists are "opportunists": they take whatever is useful to solve a problem, with no worry for the "purity" of the concepts to be used. For Einstein, this is no sin, but rather a virtue.[19] Thus, even if social scientists using models of complexity of different sorts are not sinners, the above epistemological question

[19] It seems Einstein didn't think the same when he felt the need to go to the root of the concepts of space and time in terms of an operationally sound definition of simultaneity – and his lack of opportunism was well repaid.

remains unanswered. If we don't care for an answer, we *are* opportunists (in using the model which fits with each given issue) and we *are not* (because we don't exploit the opportunities provided by different answers to the question, however tentative may they be, in order to reach the ending state of the revolution). If we do, we are called on to identify that conceptual core of complexity already mooted – and in that search we cannot remain opportunists all the time.

There is a reason why the present state of the art seems not to impose such an effort. For, if the difference between a complexity-model and any other model is *identified* with the difference between anti-reductionism and reductionism (the latter supposed to be the paradigm of "hard science"), and if there are many ways order can come from noise and emergent self-organisation can arise, complexity is just the name for a taxonomy of the corresponding set of forms anti-reductionism can take. Thus it cannot select any one of them.

The trouble is that such identification is wrong. Reductionism, either as a method or as a philosophical thesis, admits more formulations. If their differences are bypassed, it ends up with being a vague red herring for anyone fond of emphasising the anti-reductionistic character of the complexity turn.[20] But then we are led to ask: what does "reduction" mean? Isn't it plain that complexity provides evidence for antireductionism? As far as the question about meaning stays unanswered, this evidence is weak; even were it not, the interest of complexity would not depend on that evidence in an essential way. It is a matter of fact that by now the word "reductionist" sounds as an attribute expressing an implicitly negative value to the ears of many social scientists. Most of them manifest an "anti-reductionist" stance: they are ready to admit massive effects of "coherence" that *can't* be localised.

Many of the great past achievements of science, however, were obtained by (de-) composition, analysis, division, of factors at work which could be considered instances of a "reductive" approach. Now, in order to claim a theory T' is *reducible* to theory T, both supposed to be axiomatically presented, two conditions have to be satisfied: that any primitive T'-notion is definable by means of T-notions and that any T'-statement can be inferred by T-principles, after the required definitions are provided and put at work. In general, the theories we are talking about when we deal with complexity are not given an axiomatic setting, thus the definability issue is unclear. Consequently, the anti-reductionistic stance is unclear too.

For argument's sake, let's suppose this all has been made clear. Then the issue depends on what is reduced to what. Even if we take "mechanics" as the ideal of scientific method *before* complexity, the above argument is unaffected: a precise formal language for dynamical systems is presupposed by any understanding of both determinism and indeterminism, and linearity and non-linearity as well; if

[20]In mathematics, any category of tangent spaces fibered over a *connected* base space is an example of a whole with inseparable parts. In physics, the notion of field is not reducible to the behaviour of test particles and, turning to Quantum Mechanics, David Bohm's notion of an "implicate order" is an attempt to answer questions similar to those posed by the long-range coherence social scientists study.

non-linear and indeterministic mechanics is supposed to be a hallmark of the new view, it is no less "mechanical" – which sounds much less pleasing for anti-reductionists.[21]

It seems the target of many charges levelled by the anti-reductionist camp is rather *compositionality* – a notion with traits widely shared by what is "mechanical". Alas, compositionality is no less vaguely used, if not misunderstood, with insufficient care to avoid ambiguity. One frequently reads that complexity provides evidence against the "old" claim that *the whole is the sum of its parts* (Aristotle denied that and his view seems no less "old"). Suppose the only parts of system S are x and y, and $x \ast y$ is their combination. If the "old" claim means that, for any relevant quantity f, $f(x \ast y) = f(x) \bullet f(y)$, it just says that the quantity f distributes over the components (algebraically, f is a homomorphism). Were it even intended as a universal principle valid for every S in nature and for every (say real-valued) quantity, it would depend on what we take as its *parts*. It's easy to find counterexamples to the "old" claim only if the same equation holds transitively for the subparts (the components of each part) all the way down to the ultimate components, supposed to be individual entities (say, the "atoms" of S). But such transitivity is an additional hypothesis and the ultimate components of a system are not necessarily individual entities: the architecture of a system can also turn out to be modular and compositionality hold for modules only, with each module possibly defined in a non-compositional way, and vice versa.[22]

Modular architecture corresponds to what Edgar Morin named "restricted complexity", see Morin (2005). Morin took it as certain that this notion is insufficient for social sciences, and no less insufficient for that ecologically conscious intervention he advocated in society and nature. For this aim he rather proposed "general complexity". But such a conviction risks to make "general complexity" unfalsifiable, as far as it should account for *any* intervention (otherwise it would loose its generality). Were it unfalsifiable, no intervention on society could be rationally preferred to another by appealing to "general complexity". Thus, it's suitable to recall there are many versions of compositionality, intermediate between atomistic reduction and global transcendence and by consequence the non-compositional character of complexity is not uniquely identified.

[21]Additional care is recommended in using "mechanical" as if it had the same meaning in mathematics and physics. In mathematics it refers to a computable procedure and it is just in this sense we talk of a Turing "machine". Though the two meanings have interesting connections, see Moore (1990), they are not equivalent. Note also that there are indeterministicTuring machines and that the Church-Turing Thesis about recursive functions as exhausting the domain of what is informally said to be "computable" is a hypothesis, rather than a theorem. Moreover, there are many degrees of computability which do not match with the different kinds of mechanics in physics. Hence, the recognition of such a double spectrum of notions calls for more, not less, rigour in talking of the complexity of a system.

[22]For a detailed analysis of the range of meanings compositionality can have and their different implications, see Peruzzi (2005). Thus, even if *complex* is contrary to *independent* rather than to *simple*, the gain is low.

This is not just an abstract worry. Models of the mind offer various examples of the range of kinds of architecture that could not be ignored in making clear what we mean by complexity. The information processing model which cognitivism adopted as its paradigm is both computational and modular. On the other hand, holistic models of cognition have been provided with no reference to complexity. Jerry Fodor proposed a model of the mind as a system which hosts both modularity (of information through input systems) and holism (of belief-states).[23]

After all, any understanding of complexity is a compositional system of statements and the same applies to any argument in favour of complexity. Otherwise we could only understand each single step of the argument but not the whole argument or grasp it without being able to articulate its logical structure. There are brain pathologies which present us with unfortunate instances of this situation – hopefully not the case with our arguments! But, if the complexity of a system is not recognisable by only an external observer, what an internal observer says is self-referential. (Here the opposition of OR and SFR takes a more than methodological shape.)

Someone might note that there must be a modular architecture to block paradox, for if the mental resources whereby we tailor arguments to claim that the hosting system is complex were complex in their turn, the explanation of anything would be a miracle, and any appeal to miracles is not scientific. Someone else might say that if any argument is explained away by means of the language of dynamical systems, the resulting account would be an instance of the fallacy named "affirming the consequent". Thus, as we get to complexity models of mind and society it becomes unclear what distinguishes a vicious circle from a virtuous one. If the latter has an "emergent" feature which the former lacks, we are under an obligation to clarify the notion of emergence.

From Cluster to Tree

As already mentioned, the common picture of complexity puts this notion at the centre of a cluster of other notions and topics "associated" with each other through that nucleating centre, and if they are linked together, this is essentially due to complexity. Such cluster includes: *Attractors, Bifurcations, Catastrophes, Chaos, Cellular automata, Fractals, Genetic algorithms, Neural networks, Self-organisation, Holism*. Much of recent literature oriented towards applications of complexity in the domain of social sciences conforms to the common picture.

[23]Note that if the "size" of modules is smaller and the coherence of conceptual patterns (as emergent properties) constrains holistic aspects to local structure, there is space for dynamical models instantiating a kind of intermediate complexity. A dynamicist view of the mind should also be mentioned as a source of further models, as those worked out under the heading "mind as motion", as proposed in van Gelder and Port (1995) and in some of the essays collected in Peruzzi (2004).

Granted there are different ways of making the intuitive notion precise, even if complexity were inevitably fuzzy it wouldn't follow its analysis has to be fuzzy. The cluster view legalises fuzziness and since this view gains credit, the risk of confusion grows to epidemic levels. Doubtless, each item in the above list hosts many concepts and methods a subset of which matters for complexity and the converse also holds. Different subsets, however, extract different features in terms of which different clusters take form; so the cluster is rather a cluster of such clusters. Qualifications are needed and one is particularly required in connection with chaos theory.

For, within the spectrum of kinds of dynamical systems, complexity is located between two extremes: chaos on one side and a "frozen" order on the other. Such order could correspond to a sort of Pure Being, with no real Becoming, something different from what we see in nature, mind and society (quantum vacuum is far from nothing). As for chaos, non-linearity is essential to chaotic behaviour if the system is finite dimensional. But even though there is a kind of complexity in chaos – one which is history-free -, no huge number of components or interrelations is needed. Further, complexity is not committed to determinism, while "classical" chaos is, and whereas the evolution of a chaotic system is bounded by attractors, the emergence of macro-order in a complex system is generally supposed to be independent of them. A chaotic system by itself does not imply emergence of stable configurations, while complex systems of interest in the study of living beings and their societies appeal to emergent properties.[24]

The cluster of notions and topics associated with complexity suggests manifold contexts of application of the central notion, so that for any two items in the cluster there is one of these contexts in which they are linked together. Contexts of this sort are already known, e.g. fluid dynamics, cellular morphogenesis, psychology of perception. In each case, the emergence of patterns is described in dynamical terms. In particular, for what concerns the models of mind, a topic of great interest is the emergence of *qualia*. The resulting network of contexts is taken to show the general (and abstract) features of complexity are endowed of empirical content and helpful to overcome the forms of dualism we have seen in section "Bypassing the dual language".

The list of items in the cluster can be expanded and the network becomes more and more refined as soon as we examine different hierarchies of complexity, for example, by considering *logical* complexity and *computational* complexity. Neither of them were the focus of the books mentioned in section "Back to origins", but they successively became subjects widely investigated.

As for logical complexity, classical first-order logic is an obvious reference because of the prenex normal form theorem that can be proved for any formula within this logic. Passing to higher-order logic, the hierarchy of definability classes of formulae can be built and the growing logical complexity reveals deep properties

[24]It is pertinent to point out that emergence can also be massively order-destructive, rather than the sort of (pacified) Hegelian *Aufhebung* many take it to be.

of the universe of sets. The classification of formulae in terms of their degree of logical complexity is not possible for every logic, but for some of them other methods can be exploited, such as the game-theoretic ones introduced by Hintikka (1973).

In the light of the relationships between logic and set theory, one should also mention the area of *combinatorics*, related to the Ramsey Theorem, to enrich the cluster and also as a bridge to *computational complexity*, which is, roughly speaking, a quantity relative to the internal *space* and *time* of a machine; it is defined in terms of the number of steps needed to solve a problem (this number assumes discrete *time*) and the amount of memory used in the computation (corresponding to a discrete *space*). Such a measure of complexity refers to the amount of work and the time needed, for a computer to perform a task. Beyond P- and NP-problems, increase of complexity can lead to "intractable" problems. Related to a measure of information, what is identified as "algorithmic complexity" is focused on the idea that the complexity of a string (say, a message) can be defined as the length of the shortest (binary) program which generates that string as output. This line of research was introduced by Andrei Kolmogoroff, and then further developed by Gregory Chaitin and others, see Chaitin (1987).

There is a basic reason why reference to this area of research is important for the present remarks on the cluster view: independently of the various kinds of algorithms and measures considered, computational complexity is a *quantitative* notion. Thus what matters is not just the distinction between systems that are or are not complex, but their "degree of complexity",[25] and since there are many ways to measure this degree, the confidence in the "centred" character of the cluster is of little help as soon as vague associative links are replaced by precise logical relations. To prevent misunderstanding: the multiplicity of formal notions of complexity does not signal failure, as researchers can choose the resources suitable to manage the kind of data they work on. The problem with algorithmic measures of complexity lies rather in their stability under translation from one language to another and, as is well known, what is a "simple" string in one language can be a "complex" one in another.[26]

The amount of resources needed to solve a problem also points to the "subjective" aspect of complexity. In problem solving, it's common to talk of a problem as "complex" in a sense which does not necessarily match with the one in which we

[25]For a comparison between different measures of complexity, see Mitchell (2009) chap. 7.

[26]This problem brings back to mind the so called "paradox of analysis". Given a pair of expressions such that the second offers an analysis of the meaning of the first, the effectiveness of this analysis lies in the way it compresses or expands information; but if it is really effective, it cannot be fully faithful, and if it is fully faithful the two expressions are notational variations – within one and the same language. An axiomatic approach proves useful to bypass this "paradox", but for what concerns a first-order language, one should not forget that one and the same set *T* of axioms admits non-equivalent models and that an empirical model of two non-equivalent sets of axioms, *T* and *T'*, can be such that it satisfies *T* if and only if it satisfies *T'*. In case *T* is a theory of "complexity"…

say a system is complex. We perceive a connection between the two senses, in terms of the way the data directly compose with each other and the number of steps to perform in order to find a (the) solution. In both cases *speed* matters. By looking at the "problem space" as a set of possible states ordered by logical dependence, we investigate which strategy is the quickest one in finding the path from the start-state to the end-state (the solution to be provided, if it exists within the given space). If the set of these states is taken as a system, an additional question is *how* can speed be relevant to distinguish between subjective and objective complexity. To this aim, let's consider two cases.

Case1. Suppose the problem requires an arithmetical computation and you can observe, in runtime, the flowchart of a Turing machine while it performs the computation, say on two large positive integers with thousands of numerals in decimal notation. Now suppose you have no previous idea of what is going on. Unless the procedure is set in sufficiently slow motion, you can't escape the impression of an extremely "complex" task whereas, objectively, it is not so, and if you look at a multi-tape Turing machine performing many computations at the same time on many pairs of numbers (which ensures a time-saving performance), even slow motion can be unhelpful.

Case 2. As a complementary example, consider the solution given by the child Gauss to a classroom problem: compute the sum of all numbers from 1 to 100. While most children in that classroom presumably proceeded step by step, finding the result of each addition 1 + 2, 3 + 3, 6 + 4, ..., which takes enough of classroom time, Gauss found an immediate solution: $n(n + 1)/2$, with $n = 100$. Such a highly compressed representation of the data, with its "simplification" of the problem, was obtained through a *global rearrangement of the data*.

Both examples involve quite simple tasks. Now replace a domain of numbers and computations with one of agents and their (say, economical) behaviour: the problem may be one of understanding what is going on with each agent (as in Case 1) or what is the outcome of their collective behaviour (as in Case 2), possibly represented by the value of a function of many variables to be approximated by means of data relative to a suitable sample of the population. If we proceed as in Case 1, complexity is reduced by reducing speed. If we proceed as in Case 2, the reduction of complexity passes through the increased speed granted by suitable compression of data.

Current talk of a system as "complex" holds that complexity is not subjective (as in Case 1) and that any objective compression of data (as in Case 2) is *impossible*, whereas this should be proved rather than assumed. Such a proof is a demanding meta-theoretical task, which can only be accomplished after axiomatisation. Moreover, the very identification of parameters in terms of which to represent the dynamics of a complex system implies that a form of compression of data is *possible*. A clarification of the reasons for thinking complexity to be consistent with at least some kinds of informational compression is needed in any case.

In social sciences, the subjective, or "epistemic", side of complexity is no less important than the objective, or "ontological", one, and subjective complexity is an ingredient of objective complexity of social systems.

The suggestion that such different notions of complexity as those found in logic, computer science and physics can be smoothly reconciled meets with an obstacle: it is that *dynamical* complexity cannot be a function of syntax. Even a first-degree equation can give rise to a complex dynamics: a well-known case is the quadratic recurrence equation, $x_{n+1} = r\, x_n\, (1 - x_n)$,with r a positive constant and $0 \leq x_n \leq 1$. This equation offered a basic example of non-linearity (just compare what happens for values of r less than 1 and values immediately greater than 1) and the corresponding map turned out to be in good agreement with empirically observed types of economical growth.

Indeed, the "context" is essential to the application of multi-faceted concepts, but, for a theoretical framework which is intended to be transversal to the boundaries of different disciplines, a clear separation between invariant properties and slots to be filled with contextually-supplied parameters is required. If the hypothesis that a cross-domain relation exists between emergence, self-organisation, non-linearity and other facets of complexity is characteristic of that framework, a theoretical commitment follows from the posit of such existence: we are entitled to know what exactly this commitment is. If the "cluster" character of our notion of complexity is irreducible ("complexity is complex"), it prevents such knowledge in principle; and however more prudent and generous may it look than the idea of a conceptual core (not to be confused with a mere intersection), the assumption that the cluster of concepts is *irreducible* to one concept is no less a commitment.

In order to bite the cluster-concept bullet, one might appeal to metalogical theorems about the limitations of axiomatic formal systems. In fact, such appeal was made and it resulted in defining complexity as "the property of a real world system that is manifest in the inability of one formalism being adequate to capture all its properties". Alas, as soon as such a definition were part of a properly formalised theory of complexity (which does not exist so far), the self-referential character of the definition would produce a paradox. Thus the definition excludes the possibility of any such theory or perhaps the proposal only intends to emphasise that the join of many theories, each of partial character, can only remain partial. If so, it seems complexity is supposed to have the same nature of incompleteness, as proved by Gödel for any extension of elementary arithmetic; then, in view of the obstacle mentioned above, it does no more than carry the suggestion that complexity of a system is essentially relative to an observer internal to the system, and in this sense, complexity is "objectively subjective". But this does not seem to be the idea suggested by the seminal works mentioned above.[27]

An alternative approach characterising complexity in positive terms rather than through a *via negativa*, goes back to the Canadian biologist Robert Rosen and

[27]The relation between complexity and incompleteness would deserve further, more careful, examination.

significantly his work was also recognised as of great importance by David Byrne, whose research is a reference for applications of complexity to social sciences, see Byrne (1998). Here is how Rosen introduced the notion:

> a simple system is one to which a notion of state can be assigned once and for all, or more generally, one in which Aristotelean causal categories can be independently segregated from one another. Any system for which such a description cannot be provided I will call *complex*. Thus, in a complex system, the causal categories become intertwined in such a way that no dualistic language of state plus dynamic laws can completely describe it. Complex systems must then process mathematical images different from, and irreducible to, the generalized dynamic systems which have been considered universal. (Rosen (1987) p. 324)

This idea can be articulated: (1) complex systems have emergent properties, (2) emergent properties show up in phase jumps, i.e., non-smooth changes that lead to a new state, absent from the previous state space of the system; hence, it's the *state space* itself which globally changes and the probability associated with some state-transitions changes too. The conditions expressed by (1) and (2) are insufficient (stability of the new state relatively to the environment is not mentioned) but they may be taken as jointly necessary to characterise systems which comprise nonlinearly interacting components and exhibit qualitatively new behaviour relative to their parts, see Westerhoff (2000).

Further qualifications are required, because such a proposal may affect the spectrum of complexity having chaotic systems at one extreme, and also because many use "emergent" as synonymous with "non-compositional" (with reference to properties of a system). This use agrees with the above supposition according to which compositionality has a unique meaning, namely atomistic reduction, where "atoms" are pointlike, or elementary, components. As already noted in section "The complexity of science and the range of compositionality", this supposition is questionable. Therefore it's time to consider the notion of emergence in its own.

Emergence

Emergent properties are macro-features manifested by a system that are not only different from its micro-features but also such that single elements of the system (as a set) cannot possess them. Consider the question:

How can local interactions between a large number of small-scale components produce a large-scale, global, structure which does/did not exist at small-scale?

What is so produced is first read as "emergence of order". Its vast phenomenology – suffice it to mention the variety of emergent properties in different ecosystems – urges us to work out a unifying notion, to be precisely captured within a theory . . . as yet a work in progress. So, if emergence is a characteristic feature of a complex system, the lack of a theory of emergence is one of the reasons a theory of complexity is lacking. Since the authors who laid the ground for the study of complexity already thought of emergence as an instance of "order from noise", essentially calling for the use of statistical methods, the previous question can be reformulated:

How does the random behaviour of a large number of micro-components lead to collectively coherent behaviour?

We usually talk of emergent properties as associated with an integrated whole, and take them as those which make the "whole" irreducible to its "parts". In this sense, emergence is not so much a product of size as of self-organisation (the case of fractals would require a separate discussion). But it would be wrong to suppose emergence came to the fore with the seminal works mentioned in section "Back to origins", as the notion of emergence has a history of its own.

Emergentist stances were argued for already in the early twentieth century. The Belarusian polymath Alexander Bogdanov introduced a new term, "tectology"to mean the investigation of the build-up of self-organising systems in nature and society: the publication of his three volumes treatise, *Tektologia* (in Russian) started in 1912 and ended in 1927. Meantime, the basic line of argument in support of emergence was developed by two British philosophers, Samuel Alexander (*Space, Time and Deity*, 1920) and Charles D. Broad (*The Mind and its Place in Nature*, 1925), by the British psychologist Conwy Lloyd Morgan (*Emergent Evolution*, 1923) – who also advocated for OR! – and by the German philosopher Nicolai Hartmann, who proposed a stratified ontology of self-sustaining systems of growing complexity (*Das Problem des geistegen Seins*, 1933), and his emergentist view strongly influenced Arnold Gehlen's anthropology.

Piaget's "genetic epistemology" is also an emergentist theory, as it presents cognitive development as layered in four global stages, each with its own properties irreducible to those present at previous stages. In biology an emergentist stance was proposed by Ernst Mayr,[28] and Rosen's emergentist view of evolution led him to claims that do not perfectly match with the received view of Darwinism.

In spite of the interest of these stances, emergence was not a mainstream topic of twentieth century philosophy of science. When it came to the fore[29] in recent decades, it was discussed chiefly in relation to the so-called "mind-body" problem, as emergence cuts across the opposition between reductionism (cognitive states reduce to neural architecture) and dualism (minds and their specific properties transcend any physico-chemical structure). In fact, the references to an emergentist standpoint in the philosophical literature on mind forget that the notion of

[28]"Systems almost always have the peculiarity that the characteristics of the whole cannot (even in theory) be deduced from the most complete knowledge of the components, taken separately or in other partial combinations. This appearance of new characteristics in wholes has been designated as *emergence*", Mayr (1982), p. 63.

[29]This recovery was stimulated by the growth of interest for Hartmann's stratified ontology. The merit of this recovery mainly goes to Roberto Poli, see Poli (2012). Some analytic philosophers connected Hartmann's stratification with the concept of "supervenience" in philosophy of mind, but such a connection is misleading.

emergence was at its root motivated by issues of biology of much wider scope than the issues inherent to the emergence of mental states in the human brain. As for a similar opposition, one also involving emergent properties, recurrent in the history of social sciences, namely that between heteronomic and auto-nomic views, suffice it to recall the metaphor of the state as an *organism* from Plato to Hobbes, or the famous claim by Èmile Durkeim: "L'individu écarté, il ne reste que la société" and finally the controversial perspective of sociobiology as presented by Edward Wilson.

The metaphor of a society as a living being of its own lies behind the term "organisation" and in this regard, rather than a fatally cursory account of such a manifold of perspectives, a parenthesis on language is in point. Briefly, the ancestral pattern of "organisation" may also turn to be a habit inherent to human mind, but science should not be in thrall to mental habits. In our case, resort to ordinary language is dangerous *as far as* (a) the metaphorical nature of "organic" expressions used to refer to society goes generally unnoticed and (b) papers and books on emergence in complex systems mean *literally* that a set, however "organised", of living systems (e.g., human beings) *is* a living being (say, a meta-human one). Everybody realises a swarm of bees is not a bee and a pack of wolves is not a wolf, but curiously, when we come to describe the behaviour of swarms and packs, we talk of them *as if* they were animals on their own. When a set of persons is considered to form a collective unity, such as a party or a state, the same occurs and common talk reveals an even stronger ontological commitment we fail to realise (and we do not feel the need to realise).

Metaphor patterns are oriented: one can represent public debt as a disease but not disease as a public debt. But if the complexity framework is adopted and given a holistic meaning, we can turn the issue upside down. Namely, we are justified to avoid any worry about the attribution of one and the same predicate to single organisms or their species-specific societies. It's not much that organic metaphors get literal legitimacy, since systemic features are equally instantiated in biology and sociology, as that they are turned upside down once the *real* organism is The System, not the individuals. Talking of emergence is opposite to this. Thus, there is sense in which a widespread interpretation of complexity is inconsistent with emergence.[30]

Moreover, those promoting an emergentist view are not necessarily endorsing a holistic thesis. But since holism features as one of the components of the "cluster view" of complexity, a clarification of the relationship between emergence and

[30]I don't intend to deny that emergence, together with nonlinearity, is one of the essential motivations for the appeal to complexity in social sciences – in particular, emergence allows to avoid the dichotomy between dualism and reductionism. Though the impact of an emergentist perspective on foundations of social sciences is yet to be evaluated, it has to be consistent with the nested hierarchy of systems from cells to conscious minds, with top-down feedbacks. I introduced the term "entwined naturalism" to label such a view, which keeps track of both bottom-up and top-down causality, see Peruzzi (1994).

holism is needed. Before turning to this relationship, however, the strategy of objections raised against emergentism should be dealt with first.

This strategy is well-known, as it was repeatedly used in support of theoretical and/or ontological reduction of high-level concepts/entities referred to in science B to low-level ones used in science A. A clearcut formulation of the main argument is due to a philosopher of mind, Jaegwon Kim, and it's aimed at proving that psychology (as B) is reducible to neuroscience (as A), as far as mental states are correlated by psycho-physical laws to brain states. Though the argument obviously refers to specific features of the brain, the underlying strategy should apply to any other pair of sciences as our A's and B's.

Here is how the argument goes: if the B-properties supposed to be "emergent" belong to the domain of science A, they are reducible to A-components of an A-system plus their dynamics, otherwise emergence is nothing but dualism in disguise. Applied to properties of a social system, this means that no essentially top-down causality exists from an institution to a citizen and from the market to a company, *apart from* interactions between single agents. It is the same set of laws which rule the A-level, which is always at work, no matter how difficult in practice may be the description of the A-dynamics, due to the huge number of bodies and their interconnections involved. Otherwise, we have to declare our dualism, rather than trying to sell it as a sophisticated unification achieved by means of the "emergence" label.

Once the place of B = psychology and A = neuroscience is taken by another suitable pair of B-science and A-science, the structure of the argument is preserved, and may also act as an OR-inspired "therapy", against the syndrome associated with taking, for *any* kind of system, metaphors of personification as literal. Thus, if emergence is confined to "supervenient" properties, it can be dispensed with, as supervenience only relies on *covariance*: if there is a change in the whole there is a change in some part of its (material) support, and vice versa.[31]

The problem is that even the simplest social behaviours such as those examined by proxemics, relative to small groups of agents in a confined space, display the existence of interaction patterns that cannot be explained by means of point-to-point causal chains. Different constraints play a similar role for different kinds of systems: from the formation of Benard's cells to visual gestalts manifested in cases such as the Müller-Lyels illusion.

The admission that "at present we don't know how to provide a bottom-up explanation of them, but we shall have one in the future" expresses an old conviction similar to that behind Poincaré's claim that "Chance is only the measure of our ignorance" – a claim by the father of chaos theory! The essential use of statistical considerations in thermodynamics turned what was "ignored" into what had to be "ignored" in order to achieve scientific explanation. The idea that, were more

[31]There is large debate on this issue filled with any kind of subtleties. Hilary Putnam brilliantly questioned Kim's argument, see Putnam (1999), but since Putnam's objections do not support emergentism, their examination would lead us off topic.

information about each single component of a given system accessible, emergence would be dispensable, collides with basic motivations behind research on complexity.

On this concern an important contribution to conceptual house-cleaning was provided 10 years ago by a group of biologists at Amsterdam. Their research, focused on biochemical networks, was endowed with general epistemological relevance, especially for the "defects" of mechanical models and the "transcendence", so to say, of the whole with respect to the parts. So the authors sum up the guiding idea:

> Emergence in biology must be compatible with the thought that all explanations of systemic properties are mechanistic explanations and with their sufficiency. Explanations of systemic properties are always in terms of the properties of the parts within the system. Nonetheless, systemic properties can still be emergent. If the properties of the components within the system cannot be predicted, even in principle, from the behavior of the system's parts within simpler wholes then there also will be systemic properties which cannot be predicted, even in principle, on basis of the behavior of these parts. (Boogerd, Bruggeman et al. (2005) p. 131)

In fact, the same paper argues for an emergentist perspective which is compatible with mechanical models – in the wider sense of dynamical systems, see Bechtel, Richardson (1993) – and argues that, as unpredictability is instantiated at the level of system components, compositionality is not prevented at the global level. In relation to the above paradigmatic argument against emergence, the case made in this paper helps us to understand that emergence cannot be confused with *supervenience*. The latter is a notion which relies only on covariance, but covariance provides no *dynamical* explanation of emergent properties. It remains timeless as any purely logical correlation, whereas emergence involves a process in time, one that turns out to be divergent for some values of the parameters and possibly leads to different equilibria, corresponding to different global properties.[32]

In view of this all, how tight is the connection between emergence and complexity? After Rosen's remarks, what we have just seen suggests more than the inclusion of the former in the latter. Not that emergence, as a byproduct of self-regulation is sufficient (even if in conjunction with nonlinearity) for complexity, but at least we have a workable notion for most of the purposes for which social scientists make use of complexity. Admittedly, it seems an excessively restricted version of the notion, but as a rule of thumb it is better to start from a narrower but more precise concept and then to worry about its maximal extension, which at present remains unclear, than vice versa. Here OR is more helpful (to understand complexity) than SFR: were this rule of thumb more widely followed, the debate would have been less prone to invasion by distracting rhetoric.

[32]Early twentieth century's emergentism had a metaphysical import which hindered efforts to establish the scientific credentials of a theory of emergence.

Local/Global $\neq$ Micro/Macro

The variety of types of micro/macro relationships is of obvious interest for the study of multilayered systems in connection with massively interactive systems with emergent features. Among the factors to consider, one is a general and precise account of the relationships between local and global structure of dynamical systems, where the distinction between local and global is primarily *topological*. Such an account has long been at hand in physics. But the efforts made to properly exploit such distinction in models of emergence have been confined to specific topics.

In fact, attention to the local/global distinction and its mathematical aspects is missing from the above confident claims for the reducibility of the whole to its parts just as it is from most frequented arguments in support of complexity, which rely on irreducibility, so that the contrast is instantly polarised into an opposition of "holistic" vs "atomistic" views. Both those who accept and those who reject such arguments generally bypass the local/global distinction, and in particular, if the confusion of emergence with supervenience is adopted, complexity means holism and holism means dualism in disguise. As a result, the study of emergence in complex systems is prevented from making profitable use of a large amount of mathematical knowledge.

The distinction between local properties of a space, i.e., in a sufficiently small neighbourhood of a point, and global properties was brought to the fore by Bernhard Riemann and has since become crucial for differential geometry, topology and general relativity, before acquiring also an abstract meaning for logic, through category theory. The problem is how to link the local to the global in a *general* setting – Euclidean geometry was of no help on this regard, as in it local structure coincides with global structure. Saying that different properties show up at different levels is equally of no help. Riemann discovered how to face this problem in geometry, but the problem is not confined to geometry at all. We have to devise mathematical tools to deal with it, through concepts flexible enough to cover the most different kinds of systems – which is nothing beyond classical scientific method and nothing depending on holism. For an example, it is sufficient to consider the *orientability* of a space. A Möbius band, though locally orientable, is not a globally orientable surface. If the difference is not of metric nature, should we say that being non-orientable is a holistic property? Or that global structure *emerges* out of the local one? Such a simple example acts as a reminder of the role of topology in dealing with the "parts" and the "whole" when we talk of complexity.

Sometimes, complexity is presented as what fills the gap between micro- and macro-analysis, particularly when we are concerned with social dynamics. No doubt it reduces the gap, but the vertical dimension along which the micro- and the macro-structure of a given system are located (in the relevant scale of orders of size) should not be confused with the horizontal dimension corresponding to the difference between local and global structure. The local, relative to the whole system, does not reduce to the micro, and the macro is not exclusive of global structure.

If it is plain the two pairs have to be linked, it's not plain how they are linked in each case: a task more difficult than to assign priority to the macro on the micro, or vice versa, once and for all. We can find counterexamples to the claim that the microstructure of any system fully determines (bottom-up) its macrostructure, exactly as we can find counterexamples to the claim that the macro determines (top-down) the micro. We can find analogous counterexamples for the local/global pair, but the two sets of counterexamples do not match. Thus the twofold phenomenology demands more precision, in contrast with the popular view of complexity as a fuzzy overlapping of high-level structure and long-range co-ordination, prevailing over a low-level, small-range, actions of "individual" components.

Research on complexity drew to our attention a variety of many-layered systems, In general, however, the layers are not confined to two, namely, the micro-structure of ultimate components and the macro-structure of the whole. The conceptual and theoretical resources developed so far cannot risk to be confined to the variety of ways the micro and the macro link, and taking the local/global pair into account contributes to avoid this risk.[33] There is another risk, however, and it must be dealt with differently: the risk of getting more sophisticated but no less abstract models than cellular automata or artificial neural networks.

Beyond exploiting such general resources in the investigation of social systems, there is something *specific* within them that has to be taken into account, if we just consider the affordances provided by a group to its members as well as ethological patterns which only show up at a *local* level, before passing to large-scale economical structure and institutional architecture. Any appeal to complexity thus needs further qualification and, since virtual tests of theoretical models are no less theoretical, it is field research that remains decisive to assess the scientific status of complexity in the social sciences.

"Large scale analysis", "systemic architecture" and "global level" mean nothing if treated as "Open Sesame" all-purpose keywords, endowed with explanatory power by themselves. What needs to be taken into account, the *specific content*, requires methods of data-processing no less than of selection of data. This worry was well expressed by Niklas Luhmann, when he observed that if we miss the particularity of what is human, sociology is gone. But complexity refers to systems (even when discretised in an algorithmic state-transition table) independently of their "hardware". Therefore, if *the very notion of system* had to be revised in passing from physics to biology to social sciences, such claims would be at odds with any stable use of complexity. Here we see how the cluster view plays *against* the introduction of complexity in social sciences.

[33] My personal opinion is that the resulting two dimensional analysis (micro/macro, local/global) is a key to that "unity of science" proposed by logical empiricists, and it is more than a merely formal key because it is made of notions directly theoretical, rather than meta-theoretical as those which became the tool-chest of philosophy of science: in a nutshell, the language of dynamics comes first, logical analysis of language follows.

The difficulty rather lies in previous use of the notion of system *within* sociology: for, it is clear that Marx, Durkheim, Parsons and Luhmann (to mention just a few names) assigned the notion a different meaning. One advantage of a "systemic" approach focused on complexity consists of allowing a *precise comparison* between such different perspectives.[34]

Someone might object that a radically different notion of system is at work, namely the one provided by the set of relations which defines each element of a system in terms of its role (as "contrast value"). Those who share this objection place complexity within the mainstream of early twentieth century structuralism since Saussure – a losing move as structuralism has well-known problems already in linguistics. Rebaptising structure with a more fashionable name, say "network model" does not get out of troubles. No network model has ever explained syntactic structure and much the less the emergence of any syntax. The same kind of problem is present in other fields too in which structuralist views have flourished, from mathematics to psychology. Then social scientists who closed they heart with relational structure but now realise it does not pay at they expected, will adopt a cluster view of system, to save an aura of potential significance beyond the too crude notion of "dynamical system". Such a step back faces the same obstacles the cluster view of complexity faces.

Understanding massive simultaneous interaction of a great number of agents is a difficult task. True, data can be mined through statistical methods, but care has always to be taken in their use, and the arguments leading to a particular interpretation of data should be as crisp as possible. The explanation of mist is not a mist of explanations. If nonlinear process, self-regulation and its "equilibria" prove to be essential to such aims, the implicit assumptions should be made explicit and our language should be free from any ambiguity, particularly from that concerning the very notion of system. This puts us right back in front of the issue of the neglected relationships between micro/macro and local/global dynamics, now as specifically instantiated in social systems.

Multidimensionality + Unpredictability = ?

Among the aspects which make the appeal to complexity attractive in social sciences, one is multidimensionality, intended as more than reference to different factors in mutual interaction to be considered: what is meant is different *kinds* of factors.

[34]Just as biophysics and theoretical biology need much more mathematics than biologists are usually requested to know, a complex systems approach to social sciences demands a similar familiarity with mathematical tools, which would mean drastic changes in the syllabus for the next generation of social scientists. I doubt this demand will be met, if not for what concerns already established trends in "modelling", thus stopping before the foundational analysis proper.

Models appealing to multidimensionality are already present in the natural sciences. In particular, Hilbert spaces, a standard tool in quantum mechanics, are of infinite dimension. No reason for a methodological divide between natural and social sciences is detectable in the mere consideration of a very large number of independent variables being required to identify the state of a (social) system. In social dynamics, multidimensionality seems to be intended to refer not to a small set of properties/quantities at a large set of points (individuals), but to a large set of non-homogeneous relational features between even a small set of points. Thus the situation is different because each dimension is now associated to a feature inhomogeneous with the others. This sense of multidimensionality is more in line with SFR than with OR. No matter we take it as indispensable, this difference with multidimensionality in physics suggests three basic issues, to deal with in order to prevent misleading consequences.

First, is multidimensionality, by itself, necessarily related to a metric? Second, does dimensional reduction prevent preservation of relevant information and explanation? Third, is there a direct correlation between increase in dimension and increase in complexity?

The answers to these three questions are three No's. First, as dimension is a purely topological property of a space, the difference in dimension of one space from another does not refer to any notion of "distance". This goes back to Brouwer – and Brouwer's Theorem implies that no homeomorphism is possible between two spaces of different dimension. Second, the piecewise projection of a space onto a subspace of lower dimension can be *essentially* defective or not: it depends on how the projection of each "part" of the space matches with the projection of other "parts". Linear perspective is a case in point and the famous satyrical drawing by Hogarth, illustrating false perspective, is a telling example of a piecewise local projection not matching with a global projection. Third, a high-dimensional state space can have low complexity while a low-dimensional state space can be very complex. A case in point is offered by Penrose tiling of the plane.

These three negative answers, in their turn, should not be equivocated: in social sciences, as in any other scientific area of research, when *n* independent dynamical parameters are insufficient to understand what's going on, it's reasonable to add further parameters and take account of the contribution each of them provides. If what is searched for is a metric, the dimensional increase may make the difference between an inadequate and an adequate metric. This is once again in line with SFR, whereas a principle of informational economy suggests a reduction of dimensions to the minimum required for explanation, and this reduction is clearly in line with OR.

Evidence that such a reduction *can* be sufficiently faithful comes from the very efficiency of drawings, maps and diagrams in 2D, and not just for artistic aims: were it essentially faulty, science, particularly as it developed in the 17th century, would not exist. Contrary evidence is associated with aspects of complexity already occurring in low dimension – and in only adding one dimension, there are different thresholds for different kinds of systems and for different notions of complexity. The most immediate example is a double rod pendulum, with his chaotic behaviour.

As for logical complexity, the essential jump in classical logic is from unary predicates to binary predicates: monadic logic is decidable, full first order logic (with n-ary relations, for $n \geq 2$) is not. In classical mechanics, the essential jump for gravitational systems was proved by Poincaré to be from 2 to 3, in considering the "three bodies problem": the proof that the resulting system lacks stability means that determinism does not imply predictability, and as is well known Poincaré's result, with no reference to random variables, is considered the first step into chaos theory.

While that discovery may only have had a *philosophical* impact after Lorentz, it nevertheless initiated, early in the twentieth century, the study of complexity (and topological dynamics). This took place before quantum entanglement, leading to the rejection of the view that positivism and mechanistic models are inherently linked. For the young logical empiricists who intended to renew positivism by combining mathematical logic with Poincaré's conventionalism could also exploit that very result of Poincaré; but then (a) the projected combination would not yield a view in which unpredictability is tied with indeterminacy, (b) an ontology based on dynamical systems, once recognised as the present day counterpart of rational mechanics, would then tell a different story from the one which became standard in philosophy of science, (c) the choice between the two razors would appear as conventional, whereas it was supposed to be anything but conventional.

Half a century later, philosophy of science has become an academically distinct field, made possible by the Viennese synthesis of positivism and logical analysis. Language became the main subject of twentieth century philosophy, approachable from two radically divergent viewpoints, one based on mathematics and the other on ordinary language. Logical empiricism pivoted on an ideal set of "rules" of scientific rationality, while the opposite viewpoint focused on the pragmatics of language, with no granted transfer of "rules" from one context to another. How is one to approach the language of the social sciences? To a considerable extent, they are expressed in ordinary language, though suitably "regimented" with respect of common use of speakers – which is, in principle, among the subjects investigated by the social sciences. The language of complexity marks the transition to a formal framework which goes well beyond such opposing views.

Also the epistemological holism advanced by the leading American philosopher Quine caused the two viewpoints to converge, in a way that gets rid of the "reductionist" dogma of positivism, but without mentioning complexity. And yet complexity, already in the seminal works mentioned in section "Back to origins", marked not only a turn relative to science but also one relative to philosophy, in either the form of the analytic mainstream or the form of "continental" philosophy (of which postmodernism is part). As for the convergence of both forms to holism, complexity had motivations of a different sort from "analytic" holism and, as already argued, its *specific* import should not be confused with a generic holistic view. Nonetheless, analytic and continental philosophy *seemed* to find in complexity the keyword to shift holism from a metalinguistic notion to a linguistic one. So it *seemed* to many, forgetting that a consistent view of nature, society and language in particular, in terms of a hierarchy of dynamical systems, is radically different from both the analytic and the continental viewpoints (however idealised this opposition may be). While noteworthy attempts at a complexity-based philosophy of science have been made,[35] the task of showing that a *precise* notion of complexity solves *precise* issues in mainstream philosophy of science has never been accomplished. As often happens, many jumped on the complexity wagon while ignoring the task of clarifying the notion (as too time-consuming and demanding).

But Poincaré was also one of the fathers of topology and, in the study of dynamical systems, topological aspects are crucial. For instance, the characterisation of an attractor in the state space of any given system is topological, thus independent of metrics. This matches our intuition that something of a qualitative nature is what mostly matters – but must be made precise by quantitative language.[36]

Once more, an epistemological account of complexity benefits from exactness and from precise definition. This is also the case with the notion of determinism, which is used in different senses, all related to causality. There is a clear *mathematical* definition of a deterministic system and it is associated with a well-known fact: any ordinary differential equation has a unique solution if *the* initial state is given. Alas, the information characterising the initial state of a *physical* system is far from being empirically at hand and, even in such a case, the notion of cause is absent from any equation: no physical law, as expressed in equational form, mentions what causes what. So, when we investigate a class of dynamical systems corresponding to an equation, we can vary the initial state, or we can vary the equation while keeping the initial state as fixed, and in each case the causal interpretation, if it has objective content, is confined to the metalanguage; otherwise we should admit it has subjective content only. Therefore, *as far as* complexity is conceived of in terms of causal processes, something is missing. Social scientists wishing to make use of the notion are expected to be conscious of this gap.

A confirmation of this comes from the combination of topological and statistical methods that seems so promising in social dynamics, particularly if we are interested in showing how complexity models can be of use in political decision-making. If models that do not take complexity into account lead to wrong pre-

[35]See the general picture proposed to unify the "reference" essays collected by Bocchi and Ceruti (1985) or "systems biology" as argued in Boogerd et al. (2007).

[36]Topology is also a source of a special notion of complexity, which deserves mention here, for it draws attention on an issue which I consider of great epistemological meaning: *is it possible that one and only one dimension is associated with the very existence of a kind of structure observed in reality?* The answer is positive: knots exist in 3D and only in 3D. Moreover, since braids and knots are strictly related (a knot is obtained from a braid when the extremal points of any single cord in the braid are glued together), the very idea of "knotted structure" frequently implicit in talking about complexity would reveal a commitment to macro-spatial experience. This is a basic example of complexity, *essentially* related to a space of low dimension, and there is a precise way to measure its "degree". Algebraic topologists found a way (actually, more than one) to compute the degree of a knot.

dictions, it remains unclear which predictions can be made by models that rely on a notion associated with unpredictability.[37]

Granted that certainty in predictions is that old positivists' dream, we can't dispense with making predictions. Once we live within a jungle of complex systems, is our prediction-making just an evolutionary placebo? The very existence of an attractor provides a constraint on possible predictions about a given system, so that only some statistical projections relative to a given distribution of data are legitimate: complexity is not coextensive with total unpredictability. Similarly, the Uncertainty Principle in quantum mechanics did not prevent physicists from making predictions, but these were of a different kind than those previously considered and, even more telling, the very limits to prediction became the source of new understanding which explains effects that otherwise would not even be possible (such as the tunnel effect). With thermodynamics and quantum theory, statistics had already proved to be *essential* to the scientific method and it didn't prevent predictions at all; if social complexity can only be described by means of statistics, there have to be constraints which allow for an analogous window of at least a particular kind of predictions, depending on the topology of the state space of the given dynamical system, but then the popular emphasis on the general unpredictability of a complex system is in need of qualification, exactly as we have argued for other aspects of complexity the outset.

Against Vagueness and Rhetoric

In the last 30 years bombastic statements about "the complexity turn" have been an expression of faith as well as a source of skeptical reactions. In addition, the persistent lack of unification was obscured by a pragmatic attitude: researchers interested in applying the new "tools" in their particular domain of interest simply took what they could profitably exploit from one or another area of research on complexity, setting theoretical questions aside.

No doubt the same thing occurred in the past with other "tools" too, such as the techniques provided by the Calculus to physicists and engineers in the eighteenth century. But its applications were accompanied by successful efforts to provide it with a rigorous foundation. The present situation seems rather to signal that the study of complexity needs no similar effort but also is no warrant of a foundation to come. The concurrent explosion of publications and their accelerated circulation

[37] Since a few years I am collaborating to the QOL project in Tuscany, and thanks to the great work by such an expert statistician as Filomena Maggino, this research group provides the Regional Council with a feedback on political decisions. This experience led me to two considerations. First, the use of synthetic indicators is a source of, say, "suggestions" only in conjunction with ethical hypotheses and long-term projections of *qualitative* nature. Second, neither such hypotheses nor such projections warrant the existence of an attractor and, in case more than one could exist, we are unable so far to determine which attractor satisfies the optimal balance point between the multidimensional factors to be considered for QOL.

had both positive and negative effects. Among the latter, one can mention the increase of vagueness and rhetoric fed by worldwide accessible advertising. As a result, apart from highly "technical" studies within a pre-established framework such as cellular automata, talk of complexity is verging on a sort of pop philosophy, even looser than it was 50 years ago.[38]

"Complexity" is a noun, and a noun is not an assertion. It refers to a subject and is not accompanied by any theory, just as "gravity" and "electricity" do not come with a theory of their own. Indeed, it is the development of a theory which allows one to define precisely a subject of scientific investigation (and also make a quantity out of a quality), rather than the opposite, but when we are in search of a theory, we have to be able to identify, progressively, its subject – consider, for example, the notion of heat before and after the development of thermodynamics. Care for definition acts as a sieve for a cluster of ideas: though insufficient by itself, it traces the way to a theory (or at least a family of theories) and will be refined through the same process involving related concepts – consider the back-and-forth increase of measure-precision achieved by refining length units with dilation coefficients and temperature units by thus refined length units. In the case of complexity, had we even replaced an associative cluster with a core view and a selective definition, appeal to a noun would remain neither true nor false, let alone explanatory. Thus we should expect a progressively stricter fusion of the manifold features investigated by different methods – the mark of a general theory.

Such a general theory calls for a (finite) set of clearly stated principles, and in order to be of scientific relevance, the principles must have explanatory/predictive power, otherwise we revert to the present state of the art, namely, a phenomenology of manifold kinds of complexity. Such principles have only been found for some particular aspects of complexity. The problem is that different aspects are covered by different principles, and even for one and the same aspect there are different approaches. It would be futile to make use of one aspect in the context of a problem and of another one in a different context, while referring to them as instances of one and the same notion, as far as no (though partial) translation between the different principles is formalised and their being mutually consistent in a model is not explained. In the absence of a consistency proof, a minimal policy should be to avoid this kind of breezy practice and to state explicitly such a consistency *assumption* anytime it is exploited. This policy seems to be rarely observed.

Before worrying about consistency, if the very idea of a set of "common principles" underlying the phenomenology of complexity is justified, principles so far expressed in different mathematical ways have to admit a common format. This means to provide a "framework" theory proper, to be formulated in a more

---

[38]This, as long as philosophy is intended as the land in which everyone, from ordinary people to top scientists, can say anything about anything. Since I belong to that small community of researchers who subscribe the project of a scientific philosophy (no to be confused with philosophy of science), I would say the above convergence is not welcome at all; since the audience by which it is welcome is large and its growth rate is still positive, the risk of losing the baby with the bathwater is higher. Of course, it is a rough estimate.

comprehensive language than the one used in each specific area of research. Until both this theory and its language remain as a hinted "future" as in a stock exchange, it is only a *hope* that such a phenomenology will reveal its unity sooner or later. We can well share this hope. But hope is no proof; and a cluster of languages, with their associated methods and a however interesting list of examples in which the elements of the cluster are pairwise linked, remains no theory.

A recommendation of prudence, not to take what has been achieved so far as necessarily convergent, is not hostile to the idea; it simply contrasts the pragmatic stance behind the cluster view of complexity, for it skips the foundational task, and in particular the search for a set of basic complexity-parameters, in terms of which, by fine-tuning, the different topics in the cluster could be covered, and a fortiori a language in which to express parametric principles. The present reflections, however, were not aimed at providing such parameters. Our task here was to make clear, first, that social scientists are confronted with a list of notions which, while associated with each other in subtle ways, do not constitute a theory and if they make use of different notions, the comparison of their results is at risk; second, that the hypotheses relative to which kind of complexity a model refers must become explicit; third, to use a metaphor for brevity, that in order to study a "cloud" we can't rely on a "cloud" of notions.

Such a threefold task is made all the more urgent because of the many shapes the appeal to complexity assumes in areas such as sociology, education, psychology and economics. Moreover, it's a task made harder by the widespread belief that holism is *the* philosophy of complexity: among the traits of popular speech about complexity, one of the most frequent is the emphasis on holistic features, where obviously the *holon* (the whole) to take into account varies with the context.

Consider the claim (1) that a new set of tools is needed to deal with "complex" phenomena so far unexplained in terms of reducibility to the ultimate components of that kind of system. We could easily subscribe to it, but it does no imply the claim (2) that explanation of "complex" phenomena is necessarily holistic. Rhetoric prevents realising the slippery-slope from (1) to (2), which is a fallacious inference, and allows people to think that explanation magically flows from the holistic view. The proverbial baby will be thrown out with the bathwater if such fallacies, hampering any careful consideration of the philosophical meaning of complexity, pass unnoticed. Already epistemological holism, notwithstanding its battery of subtleties, was affected by a similar slippery slope argument, but now it is transformed into a metaphysics of nature and society.

Holistic rhetoric is not the only danger. According to Byrne, many of the projects aimed at introducing complexity in the social sciences arose through the influence of a particular philosophical idea, namely, the existence of a deep connection between complexity and postmodernism. Although the belief that a connection exists may be effective even if the connection does not exist, the point is that, were such a connection real, caution, care, rigour, would be impossibile to realize since postmodernism is an even more cloudy cluster of badly identified items. Not only would a scientist be entitled to be suspicious of such interest in complexity, but also the gain in understanding we expect from complexity would be lost. Who could imagine

Bateson's ideas, or the "new alliance" proposed by Prigogine, as converging on postmodernism? How can Poincaré be viewed as a postmodern mathematician?

Maybe the slippery slope from claim (1) to claim (2) only marks confidence in the potential of a new kind of models. Confidence is needed to achieve results and answering a question is stimulated by confidence that one can in fact answer it: "Allez en avant, la foi vous viendra", D'Alembert claimed, confronted with doubts about the Calculus. As already noted, confidence is no argument, and the possibility of an explanation is not an explanation; yet here, we could say, the word goes to social scientists. But the question can't be dismissed so quickly, since philosophical ideas enter the design of scientific models, and are at work in orienting research and interpreting the results of observations. It's preferable to cope with the uncomfortable task of making them explicit and precise rather than resorting to an intuitive use of implicit but unclear ideas. For this reason, a few additional remarks on holism are in order, particularly in relation to the view of complexity proposed by Edgar Morin, see Morin (1990, 2005), for Morin indeed argued against the confusion of complexity with holism, but in a way which is different from the cautious line recommended here.

If the Lack of Epistemology Does Not Pay, Ready-Made Epistemology Does Not Pay Either

In the last decades education has become a subject of growing interest for complexity. The ways of looking at education in terms of complex systems generally relies, however, on an intuitive notion, which is used as if it were captured in a formal setting, whereas, as we have seen, there are non-equivalent settings which comprise a set of different "models" linked by the intuitive notion. Here too we face the vagueness which damages the subject and stands in the way of a solid acknowledgment of its importance. Suffice it to quote the suggestive words from a widely accessed source for educational methodology: "Change is ubiquitous, and stability and certainty are non-concepts in complexity theory. Educational research can be viewed through the lens of complexity theory, replacing positivism with complexity theory".

Apart from being easily expected that looking at x through the lens of A rather than B, the B-view of x is replaced with the A-view, the implicit anti-positivistic stance is taken as if logical empiricism never existed. This is anything but an isolated incident: as we noted, much of the current literature on the complexity turn easily dispenses with the methodological standards of philosophy of science and particularly with the *logical analysis* of scientific language; but if we envisage a growth of educational sciences by means of complexity, such analysis is required. Emphasis on complexity typically goes with disregard for the analytic tradition. Not to be equivocal, the present remarks do not suggest any nostalgia for the supposedly golden age in which logic and methodology were thought to be that general problem-solver. They simply suggest that familiarity with a metatheoretical habit, the one to which such analysis can also educate the educators, would save talk about complexity from confusion, e.g., between theory and model, language and meta-language, an implicit definition and a criterion of empirical tests.

Now, much of the attraction exerted by complexity on educational studies comes from the perspective advanced by Edgar Morin. Three points argued in the previous pages were already advocated by Morin, namely, that

1. complexity is other than unqualified holism,
2. recursive thinking is compatible with complexity,
3. the lack of a general theory does not prevent profitable applications of complexity to social systems.

Though Morin is not responsible for the ways his ideas have been used, other traits of Morin's perspective favoured misinterpretation of these points and turned out to feed that misleading mix of intuitive ideas and formal models which resist fusion. To check this diagnosis, let's go back to such a program essay by Morin as "From the concept of system to the paradigm of complexity", in which it's explicitly claimed that "In order to make sense of the concept of system, we must postulate a new, non-holistic principle of knowledge". Immediately following, we learn that "This will be possible, however, only if we conceive of systems [. . .] in terms of a paradigm", that is, as a set of basic, essentially relational, patterns "between a restricted number of master notions", where these "master notions" are supposed to generate our overall view of nature, see Morin (1992).

This formulation, as well as similar ones in other works by Morin, could be willingly accepted. But, apart from vagueness which is of no help in identifying such a grand project, it takes little time to realise that one and the same set of master notions can remain in different relations, consistently with the basic patterns, and as these patterns are specified, what we get is a good old, Lego-style, *foundational* theory, axiomatically presented. Which cannot be! For, Morin also claimed that the aim of complexity is to pass over *any theory* of systems, while keeping a systemic view of theories. Thus a contradiction is reached.

To such a verdict one cannot reply that it's just the adoption of a view inspired by holism that Morin proposes, since he takes explicit distance from holism. The problem is that the holism which is the target of his objections is the most naive one, by now common in pop culture. But, if one intends to raise an epistemological objection to holism, it's fair not to aim at its most dogmatic, or lazy, formulation, but rather at the best argued one, which is due to Quine. Alas, Quine can hardly be regarded as a postmodern pragmatist and his arguments are not sketches of conceptual maps but rather logical inferences; they require meticulous analysis and if they turn out to be indebted to a slippery-slope argument, one should prove that the appeal to complexity is free from it. We find no such discussion of Quine's holism by Morin, so it seems that Quine's arguments (his "Two dogmas of empiricism" is among the most quoted and commented papers of twentieth century's philosophy) do not deserve the extremely careful attention philosophers of language and philosophers of science have devoted to them.

Now, the spectrum of possible versions of holism can be defined in terms of logical complexity, see Peruzzi (1993) and Quinean holism is only one of them. Among many other kinds, one, named "local holism", avoids the above mentioned confusion of *micro* with *local* and may be profitably used in the investigation of social systems too, provided the class of dynamical systems with locally holistic features is identified.[39] Thus: in which sense is complexity holistic? As complexity can arise within the genesis of a modular (non-holistic) system as well as in the interaction between its modules, the answer is not easy. Moreover, with reference to "the society of mind", a case for modular architecture of input systems is well-known and it's neither atomistic nor holistic (in its strongest sense). Hence the present epistemological worries are not off the point.

We can subscribe to Pascal's remark "I consider it as impossible to know the parts without knowing the whole as to know the whole without knowing the parts", were it not that the smell of a vicious circle in it is not nullified by the wisdom of his words. But we can't claim that it is sufficient to declare such a part-whole vicious circle as a virtuous one and ascribe it generative power, since this would transform the claim of complexity to scientific revolution into magic. The philosophical view underlying Morin's manifesto seems thus to be somewhat similar to the Neoplatonic view according to which, first, the basic categories of reality and thought are pairs of concepts in mutual tension, analogously related to other basic pairs, and, second, the tension between Being and Becoming gives rise to every system in nature by means of an emanative process of energy stemming from the whole-embracing One.

Indeed, a large scale circularity was already ascribed to nature by the contamination of Neoplatonism with Stoic elements, not to mention that a dialectics of nature was later proposed by Schelling, and in a materialistic version by Engels. Curiously, it seems that those who in our time emphasise complexity as a central notion to view reality in a new way, ignore or prefer not to mention old footprints of dialectics. Mathematical details aside, were the arguments for complexity of the same kind, we would get just an updated version of those ideas, now expressed in the language of dynamical systems.

Moving on to Morin's appreciation of recursive thinking as a key to understanding systems, no doubt recursion involves iterated self-application, but the two notions should not be confused: an endomorphism of a given algebra X, e.g. a structure-preserving map from X to X, can also be treated as an element of X (provided suitable closure conditions are satisfied), but such a map is not necessarily *recursive*. Likewise, the implicit assumption that any recursion generates something "emergent" cannot be maintained, and we cannot assume the cases in which this is true as an explanation, on pain of creating *another* vicious circle.

Finally, Morin charged the mathematical notion of "structure" as being an "atrophied concept" of organisation, as it ignores process and only looks for static order. This charge is two-sided. On one side, it is wrong in failing to recognise that

---

[39] As far as I know, this is an open problem.

the mathematical notion of structure is presupposed by the very equations in terms of which we describe a dynamical system. Are the field of complex numbers, the topology of a state space, the algebra generated by any set of propositions, all instances of "atrophied" organisation? On the other side, it is correct if we assume the set-theoretic notion of structure; but the category-theoretic notion of structure takes the variability of "cohesive" sets into account, see Lawvere (1994).

The Future

Which kind of role can complexity play in the future of social sciences? This is the kind of attractive question that receives many answers. Tipically, the answers are based on faith, hope, confidence. Should they take also the complexity of world-wide research system into account, they ought to be kept in the waiting room. But to refuse the challenge in the question seems to indicate a fear to make a commitment. This paper suggests a *conditioned* answer.

If an explanation of emergence and self-organisation in massively relational systems is indispensable for a unified scientific image of nature and society, and if no complexity-free image offers this explanation, the study of complex systems is decisive. In fact, this study already provides a set of general dynamical patterns (involving massive interactions, discontinuous state-transitions, attractors) across the most different domains, from a crack in the ice to a crack in the stock market, from aperiodic tiling to city planning and growth, together with a set of general notions in terms of which to frame a collection of *Gedanken-experiment* models. In the social sciences, such models allow the progressively narrowing of the window of explanations and thus to intervene, depending on our objectives, in a more rational way, or, if you prefer, in a less irrational one – for instance before taking a politico-economic decision.

The variety of case-studies and methods has proved useful in enlarging the scientific basis for research on complexity. But if we don't proceed beyond a collection of models for different kinds of dynamics, described by different sets of concepts, the claim that they all are instances of one and the same thing – named "complexity" – risks being no less metaphysical than Aristotle's appeal to "Substance".

It would not be the first time, in the history of science, that, before reaching a theory proper, research passes through a lengthy phenomenological phase. Nor would it be the first time a notion passes through "prismatic" decomposition – with one term replaced by many, which do not refer to different species of one and the same genus. In the case of complexity, the plea for semantic care as argued in the previous sections, however, is not a plea for retreating from the original project: it is rather an exercise in that conceptual "house cleaning" which is needed to make the advancement of the project possible.

As a matter of fact, the search for concepts unifying different areas of research starts with a set of models formulated in a pidgin idiom, and at some stage the very ambiguity of concepts proves to be fertile. The process of selection takes time. In the case of complexity, such selection faces two kinds of difficulty: theoretical and methodological. The definition of complexity and the most general conditions for emergence are of the first kind, the amalgamation of statistical and non-statistical aspects are of the second.

Granted that there is evidence in the social sciences supporting the claim that, in order to explain the connection between micro- and macro-dynamics, appeal to complex systems is indispensable, their indispensability implies no cult of complexity. The whole set of worries expressed here, starting from all implicit assumptions about a notion to be made explicit before its use, show as they are intended, namely, to avoid the danger inherent in the vagueness in which complexity is often wrapped. Care in the framing of definitions is needed in order to formulate a working theory and SFR-users can but benefit from it. Let us remind ourselves of what Morin rightly claimed, when he said that "complexity" names the problem, rather than the solution.

Passing from a cluster view of complexity to a core view, and from a variety of models to a theory, as recommended above, is, however, not enough. For, it would not reveal which specific patterns are actually at work within an actual, real, social system, just as the theory of metric spaces does not tell physicists which metrics to adopt in dealing with a set of observables. Social science is not mathematics, whereas, if complexity is independent of the hardware, i.e., the "nature" of system-components, it is (at best) a mathematical notion. Complexity intends to be *more* than that. As far as this *more* merely assumes the form of a collection of possible "models", it would be wiser to avoid any risky generalisation and remain with the (already difficult) task of providing a taxonomy as wide and detailed as possible. But a collection of different, however sophisticated, mathematical models for different kinds of phenomena does not match with the original idea. To retain that idea, the search for a unifying theory, axiomatically presented and endowed with empirical content is to be nurtured by overcoming any comfortable ambiguity.[40]

References

Baianu, I. C. (Ed.). (2011). *Complex systems theory and biodynamics*, cogprints.org

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of 1/*f* noise. *Physical Review Letters, 59*(4), 381–384.

Bateson, G. (1972). *Steps to an ecology of mind*. San Francisco: Chandler.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as research strategies*. Princeton: Princeton University Press.

[40] I am grateful to Filomena Maggino for stimulating remarks on some topics dealt with in this paper. I also thank John Bell and Mike Wright for suggestions in order to improve the English form.

Bertuglia, C. S., & Vaio, F. (2011). *Complessità e modelli*. Torino: Bollati Boringhieri.
Bocchi, G., & Ceruti, M. (Eds.). (1985). *La sfida della complessità*. Milano: Feltrinelli.
Boogerd, F. C., Bruggeman, F. J., Richardson, R. C., Stephan, A., & Westerhoff, H. V. (2005). Emergence and its place in nature: A case study of biochemical networks. *Synthese, 145*, 131–164.
Boogerd, F. C., Bruggeman, F. J., Hofmeyr, J.-H. S., & Westerhoff, H. V. (Eds.). (2007). *Systems biology: Philosophical foundations*. Amsterdam: Elsevier.
Bunge, M. (1979). *A treatise on basic philosophy: A world of systems* (Vol. 4). Dordrecht: Reidel.
Byrne, D. (1998). *Complexity theory and the social sciences*. London: Routledge.
Chaitin, G. (1987). *Algorithmic information theory*. Cambridge: Cambridge University Press.
Ehresman, C., & Vanbreemersch, J. C. (1987). Hierarchical evolutive systems: A mathematical model for complex systems. *Bulletin of Mathametical Biology, 49/1*, 13–50.
Gleick, J. (1987). *Chaos: Making a new science*. New York: Viking Penguin.
Haken, H. (1977). *Synergetik: eine Einführung*. Berlin: Springer.
Hintikka. (1973). *Logic, language games and information*. Oxford: Clarendon Press.
Holland, J. (1995). *Hidden order: How adaptation builds complexity*. New York: Basic Books.
Horgan, J. (1995). From complexity to perplexity. *Scientific American, 272*(6), 104–110.
Horgan, J. (1996). *The end of science: Facing the limits of knowledge in the twilight of the scientific age*. New York: Addison-Wesley.
Lawvere, F. W. (1994). Cohesive Toposes and Cantor's lauter Einsen. *Philosophia Mathematica, 2*, 5–15.
Lawvere, F. W., & Schanuel, S. (eds). (1986). *Categories in continuum physics*. Springer lecture notes in nathematics 1174. Berlin: Springer.
Lorenz, E. (1972). Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas? AAAS Report, Decbemer 29.
Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition*. Amsterdam: Kluwer.
Mitchell, M. (1996). Complexity and the future of science, Santa Fe Institute.
Mitchell, M. (2009). *Complexity: A guided tour*. New York: Oxford University Press.
Moore, C. (1990). Unpredictability and undecidability in dynamical systems, *Physical Review Letters*, 64, n.20 pp. 2354-2357.
Morin, E. (1990). *Introduction à la pensée complexe*. Paris: Edition du Seuil.
Morin, E. (1992). The concept of system and the paradigm of complexity. In M. Maruyama (Ed.), *Context and complexity* (pp. 125–138). New York: Springer.
Morin, E. (2005). Restricted complexity, general complexity. In C. Gershenson, D. Aerts, & B. Edmonds (Eds.), *Worldviews, science and Us: Philosophy and complexity* (pp. 1–25). Singapore: World Scientific.
Peruzzi, A. (1993). Holism: The polarized spectrum. *Grazer Philosophische Studien, 46*, 231–282.
Peruzzi, A. (1994). From Kant to Entwined Naturalism, *Annali del Dipartimento di Filosofia*, IX [1993], Università di Firenze, Olschki, Firenze 1994, pp. 225–334. https://unifi.academia.edu/AlbertoPeruzzi
Peruzzi, A. (Ed.). (2004). *Mind and causality*. Amsterdam: Benjamins.
Peruzzi, A. (2005). Compositionality up to parameters. *Protosociology, 21*, 41–66.
Peruzzi, A. (2006). The meaning of category theory for 21th century's philosophy. *Axiomathes, 16*, 425–460.
Petitot, J., Varela, F., Pachoud, B., & Roy, J.-M. (Eds.). (1999). *Naturalizing phenomenology*. Stanford: Stanford University Press.
Poli, R. (2012). Nicolai Hartmann, *Stanford Enciclopedia of Philosophy*, plato.stanford.edu/entries/nicolai-hartmann
Prigogine, I., & Nicolis, G. (1977). *Self-organization in non-equilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley.
Prigogine, I., & Stengers, I. (1978). *La Nouvelle Alliance: metamorphose de la science*. Paris: Gallimard.
Putnam, H. (1999). *The threefold chord*. New York: Columbia University Press.

Rosen, R. (1987). Some epistemological issues in physics and biology. In B. Healy & D. Peat (Eds.), *Quantum implications* (pp. 314–327). London: Routledge.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin and fabrication of life*. New York: Columbia University Press.

Thom, R. (1972). *Stabilité Structurelle et Morphogenèse. Essai d'une théorie générale des modèles*. New York: Benjamin.

van Gelder, T., & Port, R. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition*. Boston: MIT Press.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

von Bertalanffy, L. (1968). *General systems theory*. New York: Braziller.

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

Waddington, C.H. (1968–1972). *Towards a theoretical biology*, 4 vols., Edinburgh: Edinburgh University Press.

Westerhoff, H. V. (2000). Live perspectives of biochemistry. *The Biochemist, 22*, 43–47.

Wolfram, S. (1984). Cellular automata as models of complexity. *Nature, 311*, 419–424.

Wolfram, S. (2002). *A new kind of science*. Champaign: Wolfram Media.

Chapter 2
Building Knowledge. Between Measure and Meaning: A Phenomenological Approach

Rocco Sacconaghi

A Methodological Premise

In this chapter I intend to propose a *phenomenological approach* to the problem of the synthesis of analytical data, of which social indicators are a subset. *Phenomenology* is a "theoretical practice" (Husserl 1959) that attempts to make a rigorous description of experience in its totality (which is, as such, subjective *and* intersubjective). In pursuing this end, phenomenology is aimed at understanding its fundamental structures (Husserl 1959, 1976; Scheler 1986).

At the basis of this philosophical-scientific perspective (Husserl 1950, 1952; Merleau-Ponty 1945) lies the idea that every formation of sense is rooted in (inter-) subjective experience.[1]

In contrast with all other scientific disciplines, phenomenology does not entail the adoption of a point of view other than the subjective one: its aim is to establish itself in a perspective which belongs to our "original" personal standpoint, not to find a point of reference that is external to it. In this sense one can speak of "first-person perspective", as opposed to the "third-person perspective" adopted by specific sciences, though in different forms and with varying degrees of awareness.

[1]The connection with the *subject* gives an essential contribution to the establishment of sense of any given *object* – just like the connection with the world belongs to the nature of every subject. It is only by assuming the connection which what we call "experience" consists in as original and constitutive that we can avoid both a radical mind-body dualism and the equal and opposite reductions of one of the two poles into the other. These two perspectives would lead respectively to the absolutization or the negation of subjective will, as well as to the corresponding negation or absolutization of objective reality. Both a radical dualism and an absolutization of one of the two poles would give rise to unresolvable theoretical and practical problems.

R. Sacconaghi (✉)
Studium Theologicum Jerosolymitanum, Jerusalem, Israel
e-mail: r.sacconaghi@gmail.com

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_2

Phenomenology does not however present itself as a radical alternative to specific sciences, nor can it on the other hand settle for standing parallel to and never intersecting with them, ignoring their methods and results in order to safeguard its own (presumed) autonomy. On the contrary, phenomenology strives to establish a fruitful dialogue with specific scientific practices by offering a "spontaneous convergence" (Bergson 1938; Merleau-Ponty 1948).

On the other hand, stating that adopting a "further" – in other words, "external" – point of view beyond the (inter-)subjective personal one is not necessary does not mean that the philosophical-scientific discourse is reduced to common sense. Rather, through phenomenology, what is *implicit* in common sense, what is given but not grasped, becomes *explicit*. To accomplish this work of making things explicit, one needs to remain within experience itself (that is, within the field of common sense) so as to grasp its every element and thematise its fundamental (transcendental) form (Merleau-Ponty 1960). This form is usually ignored inasmuch as it is taken for granted as being the constitutive relation between subject and object, or in other words precisely what makes experience what it is. The phenomenological approach thus consists in looking at the object neither objectifying it – and therefore absolutizing it – nor "explaining" it as a subjective projection – and therefore dissolving it (Di Martino 2013).

The aim of this chapter is to illustrate how the phenomenological approach can contribute to an effective interpretation of the relation between *heterogeneous* elements, moving from a *list* (which is already a particular kind of possible *order*) to a *synthesis* without causing an undue *homogenization* of the elements themselves.

Through a phenomenological description, an unstructured mass of data can be "synthesized" in a unique way. The resulting synthesis is in fact *revealed*, not *produced*; it is a synthesis that is co-originary to its single elements and not extrinsic to them (Husserl 1966).

The synthesis attained by the phenomenological description does not imply the addition of subjective interpretation to mere objective, loose data (such as an arbitrary narrative), nor is it the reduction of single elements to an "objective" measure which would entail the dissolution of their qualitative differences. Both these operations involve a unity which is external to the data itself and which does not spring from the actual, integral experience we have of it.

Instead, phenomenology proposes to bring to light an organic relation already existing between distinct elements. It is an organic relation that is already implied in the "things themselves" (Husserl 1987) – all the while being still implicit precisely for this reason.

Synthesis and Concreteness

The Need for a Synthesis

A definition of synthesis is necessary in order to proceed. To this end, it can be useful to ask where the need for a synthesis originates. Why do we need to synthesize analytical data?

This is a question that lends itself to a variety of answers, as the realization of a synthesis responds to a number of different needs. However, we already find ourselves in the act of operating a synthesis of the multiple by determining that all possible answers can be traced back to two closely intertwined orders of needs: cognitive ones (among which "theoretical" needs are but one of the possible forms) and practical ones. Dealing with a mass of data, the spontaneous tendency to synthesize it derives from the need to be acquainted with what lies before us and to know how one can use it, as well as what one can obtain from it. So, why is there a tendency (conscious or not as the case may be) to seek after a synthesis in attempting to respond to this need?

It is due to the fact that knowledge and use both entail some form of *reductio ad unum*: since I am "one", until I have before me something that can be arranged into a unity – in other words, into a set of things that may be identified as *one thing*, however complex it may be – I will not be able to enter into an active relation with it (a relation which is as such both cognitive and practical). Every time we enter into a relation with something, we bring about a synthesis that arranges a multiple into a unity, in a variety of different forms. The fact that one might be dealing at once with multiple "things" poses an additional problem that does not concern the present argument; in any case it has to do with multiple, simultaneous relations with "individual things". When one is faced with a mass of data obtained through an analysis destined for a practical application, the need for a synthesis thus emerges in response to the following question: "What shall we do with this data?", or in other words, "How do we use it?".

In other terms, the synthesis responds to a need for *concreteness* in the relation with things. Concreteness is indeed precisely what is realized in a cognitive-practical relation with reality in its complexity, a relation that is both integral and effective. Without some sort of synthesis, there would be an unbridgeable gap between subject and complex reality. The existence of this gap is usually attributed to the separation between knowledge and action, on the basis of the interpretation of the relation between these two dimensions in terms of ideal and real – thus normally leading to considering knowledge as being abstract and to equating concreteness with practical action. Nonetheless, the need for concreteness emerges clearly first of all on a cognitive level: there are indeed cognitive acts that are more suitable than others for introducing us to the relation with reality. These are the cognitive acts that grant us access to certain specific realities, revealing their internal connections and articulations. Inasmuch as a synthesis (which is primarily a cognitive act) introduces us to reality in its concreteness, it can be called concrete itself – in the sense that it is *concretizing* act.[2]

Synthesis and Representation

Any type of action always implies a synthesis, even if this often occurs implicitly and even unknowingly: the act will entail a synthetic interpretation which is "invisible" to the subject of the action, though no less "operative". A synthetic mediation that is conscious (and made explicit) facilitates the subject considerably in its act of practical application. On closer consideration, what we're dealing with is the extension of a dynamic that is common for human beings to the point of always being overlooked; it is the act of *naming* things, which is none other than the first form of synthesis of the multiple. When something is named, a qualitative leap in the cognitive-practical relation with that thing occurs. To possess (in a metaphorical sense) the name of something is at once a preliminary condition and an effect of a potentially full possession – both cognitive and practical – of that thing.

When the synthetic act is conscious, it always tends to make itself explicit, and in turn the making – explicit of a synthetic act offers an effective support for the permanence of the awareness itself. In particular, when faced with very complex synthetic unities (e.g. a situation made up of numerous factors, a project, a memory, a difficult calculation etc.) one tends to avail oneself of a further form of synthetic mediation, which makes it possible to be understood in an explicit way. This further form is the *translation* of the multiple into a *system of signs*, which gives life to a representation. A poem, an algorithm, a diagram, a map: these are all examples of a representation.

In its multiple forms, a representation is a very powerful tool, as has been widely documented by a number of different thinkers and scholars (Husserl 1959; Derrida 1962; Havelock 1986; Di Martino 1998). I will focus on highlighting one of its functions, which is that of allowing one to (partially) transcend the spatio-temporal limits that subject and object constitutively find themselves in. More precisely, by translating a complex unity into signs placed on a physical support, the borders of the subject-object relation are dilated, inaugurating cognitive and practical possibilities that were unprecedented and even unimaginable until a short while before (as clearly emerges in our age radically marked by information technology (IT) – and by the use of the internet in particular – which not only contains various forms of writing, but which in turn is in itself a further representation – one which is so innovative that it redefines the meaning of the term itself). For instance, the

[2]One should on the other hand enquire as to whether even more profoundly than the tendency to oppose theory and practice (the former being abstract and the latter concrete) there may be an even more sketchy paradigm, one that equates the concrete with what is material and the abstract (which in this case tends to be synonymous with "inexistent" and "illusory") with what is not.

representation makes it possible to fix what would otherwise be destined to disperse itself temporally and spatially in a point in space and time, making it *potentially* reactivable and therefore once again knowable, thus supporting the limited memory of the subject – as well as communicable, thus favouring the communication between distant subjects (Husserl 1959; Derrida 1962).

Representation can be seen as a paradigmatic form of synthesis: it is the development of a preliminary synthetic hypothesis that and in turn it reinforces the synthesis by *objectifying* it. A representation thus enables a synthesis to fully realize its function of cognitive-practical mediation: a represented synthesis serves as a *model* of complex reality with which one is in relation to, as a "frame" of reality. A model "incarnates" the synthesis of a specific complex reality, thus simplifying the role of mediator that is required from the synthesis itself; indeed, a model offers a stable image of the possible unity of the complex reality in question, thereby making it easier to know and to transform.

Abstraction vs. Abstractedness

However, precisely in this paradigmatic form of representation, the synthesis reveals a potentially problematic (or even self-contradictory) feature, namely the necessary passage through some form of *abstraction*. Even if it is not objectified by means of a representation, the synthesis nonetheless implies this feature, though in "latent" form, while with the representation it becomes patent.

Indeed, every representation is as such an abstraction in both senses of the term. A representation is the transposition of one reality into a different reality, which allows us to partially reproduce the former within the latter. This operation implies an abstraction in the first sense of the term, so in other words as the act of abstracting; extracting, isolating, and extrapolating one or more aspects of an element of reality from the organic totality of connections in which it is originally inserted. Secondly, abstraction is not only the act of abstracting but also the product of the act itself, a "reality *sui generis*" that is introduced into an order that precedes it and inevitably modifies it. One could even say that an abstraction, which always begins with the conception of a possibility (imagination), is only realized by means of a representation; it can only be actualized as a sign. In this sense, an abstraction is in turn a thing, the creation of which is functional to the exposition of a particular aspect (or series of aspects) of another thing. A model, as we have described it above, is an abstraction.

At this point it is necessary to inquire whether this passage through an abstraction implied in the objectification of a synthesis does not produce a short circuit in the synthetic act, in itself aimed at concreteness. In other words: doesn't the actualization of an abstraction preclude the very possibility of a synthesis?

To answer this question we first need to recall that the abstraction should be seen as a passage, a means. It needs to be functional to the actualization of a synthesis, i.e., functional to the establishment of an effective mediation between the subject and a complex reality, which in turn is aimed at founding an integral cognitive-practical relation with the latter. Though it implies an abstraction, the representation can give a significant contribution to a concrete relation with certain realities. This function emerges, for instance, in those very particular representations of signs that are found in works of art, such as certain films and novels, or in some paintings and sculptures; these are all representations that introduce us to reality in a privileged way, revealing it in its richness and variety, that is, in its concreteness (Merleau-Ponty 1948, 2002; Maddalena 2015). These are the representations that achieve in an eminent and paradigmatic way what all representations should accomplish in some way, namely to make reality emerge more distinctly, thus enabling us to engage in a more integral relation with it.

On the other hand, it is precisely in this great potential found within the representation of signs that the possibility of *abstractedness*, which consists in an absolutization of abstraction, is rooted. Abstractedness is in fact the opposite of concreteness, and it is obtained where the synthesis is exhausted in the act of abstraction. Taking the work of art as an example, we can begin to see the possibility that is implied in the power of the representation; just as a great novel can enhance the relation with a certain reality, it can also isolate us in that "other", parallel reality, which should be functional to the exposure of the concrete reality. When a synthesis uses an "objective abstraction", as in the case of a model, and instead of going beyond it, it terminates in the abstraction itself, the complex reality is replaced by a fictitious unity; this is what abstractedness consists of. Now, with the absolutization of the abstraction, the synthesis denies itself.

There are two cases of abstraction in which this self-denial of the synthesis tends to realize itself more easily, or in other words two forms of abstraction that make the abstractness more difficult to avoid; namely the two types of non-phenomenological syntheses that I illustrated in the methodological premise. In both cases, the synthesis tends to coincide with the abstraction, realizing itself by means of a reference to something that is external to experience and from which the unification of analytical data can be derived.

The first case seems to have little relevance to the general problem that is being discussed in this chapter, i.e., the synthesis of analytical data. It is the peculiar representation found in the *narration*, in the narrative account and transfiguration of events. As long as it consumes itself within the individual perspective of the narrator, a narration can indeed represent the exact opposite of the concrete synthesis that enables us to establish a cognitive-practical relation with reality, causing us instead to lose our way in the meandering labyrinth of the narrator's mind. A refined and brilliant example of this self-denying synthesis is James Joyce's renowned *Ulysses* – while keeping in mind that the experience of artistic fruition itself, as much as it may be extreme and paradoxical on a formal level, always conceals a synthesis that is authentic and therefore concretizing. The artistic act in fact presents a unique ability to synthesize the complexity of reality.

Where the narration loses its artistic dimension and is reduced to pure arbitrary subjectiveness, the synthesis is exhausted and dissolves in a complexity on a secondary level. We are left with a mass of particular images, which can no longer be rearranged into a unity. In other terms, in this type of absolutization of the abstraction, the synthesis carries out a partial intervention on the data that is gathered and offered by the analysis: it is a partial reorganization of the set of analytical data, which however does not fully complete its trajectory. It is an unfinished synthesis, that is, a synthesis that in a certain sense remains analytical: the only unitary point of reference is the interpreting subject itself, who produces fictitious connections among things (in this sense, every lie can be seen as a particular application of this unfinished synthesis). The outcome of this operation is a lack of a proper unification of the individual data.

A second type of abstraction which tends to deteriorate into abstractedness is that of a synthesis which is realized by means of a reference point which is external to a purely objective experience (in this sense it appears as the opposite of the purely subjective narrative, though it shares its reference to an element which is external to experience). The purely objective element that the analytical data is to be referred to is what we call unit of measure. The problematic outcome of this second form of abstraction is the complete dissolution of the original complexity of the reality in question, the absolute homogenisation of its individual elements. In this case, the analytical data is unified in the sense that it is attributed to a single dimension. This, in simpler words, is what measure is.

In contrast to the first, this second type of abstraction which is exposed to the risk of abstractedness displays a clear connection with this chapter's theme: the synthesis of social indicators can in fact be conceived in terms of pure measure. Precisely in the consideration of this possible application, however, the powerful instrument of measuring exhibits some internal problems.

Measurement of Subjective Values

Value as a Subjective Relation

There is a particular case of measure-synthesis in which the abstraction's tendency to absolutisation clearly emerges, and that is *money*. By offering a phenomenological interpretation of it we can better understand the difference between the abstraction itself (which is functional to concreteness) and its absolutisation (which is the cause of abstractedness), and at the same time identify the step in which the latter is surpassed.

From a phenomenological point of view, money appears to be a synthetic unity that is realized as a measure in a very peculiar way. It can be described as the establishment of a measure-synthesis in which everything can potentially be included – or even as a representation through which everything can be "re-written". The operation implied in this powerful measuring technology that is money is in a certain sense *specular* compared to the operation carried out by all the other measuring practices. While in these the condition of possibility of the synthesis is the exclusion of (or the separation from) *the subjective*, in the money-measure synthesis the inclusion *in the subjective* – on which a second degree objectification of the subjective dimension is based – presents itself as necessary.

Normally, the act of measuring occurs through the mediation of an objectification (an objectifying abstraction); in order to arrange a certain reality inside a measure, one needs to separate it from its connection to the subject, as Galileo Galilei wrote, in a move that would mark the birth of modern science (Husserl 1959). The separation from the subjective dimension (regardless of whether one considers the nature of the subject-object connection as accidental or as essential to the nature of the object itself) is the necessary condition for the operation of abstraction that is characteristic of the act of measuring, i.e., the rearrangement of a multi-dimensional reality into a single dimension. If a reference to the subject remains, it plays out as a reference to what is objectifiable in the subject – merely psychical, or physical-psychical, according to the ontological conception that underlies the inquiry (Husserl 1952). The fact that some maintain that the reference can also be to the subject in its integrality depends on the fact that they view the subject as integrally objectifiable; once again, the operation that makes measuring possible here is an objectification, although it is used as a foundation for the general ontological view at play. It is therefore a basic option, not a functional one, though the nature of the procedure remains unvaried.

When a science that includes measuring procedures has a series of strictly subjective-personal dynamics as its object of inquiry – as for instance occurs in some branches of psychology or social statistics – one performs the operation described above. The subjective-personal dynamic is translated into purely objective terms, "projecting" on a *one*-dimensional level what is in itself multidimensional. The figure obtained from such a measurement can then be integrated by adding the results obtained from parallel measuring procedures. One can thus attempt to regain the complexity of what was originally multidimensional, but the individual aspects of the newly obtained complexity – besides remaining somewhat irregular – are in any case an expression of the objectifying and "mono-dimensionalising" operation described above (Merleau-Ponty 1948, 1960).

The proper realization of the money-measure is a specular operation to the one I have just illustrated. Instead of carrying out a *direct* objectification (i.e. an identification of the objective in itself, which might imply a partial or total reification of the subjectivity) as a necessary preliminary condition for the measure, the money-measure realises an *indirect* objectification (which I will call second-degree objectification), inasmuch as it is preceded and made possible by a sort of "*subjectification*". We will call "subjectification" the explicit thematisation of a peculiar form of relation that reality in its *entirety* entertains with the subjectivity, or in other words the interpretation of reality in its entirety *inasmuch as* it is connected with the intersubjectivity according to a certain relation.

This relation – thematised and made explicit (established/recognized[3]) – is the relation of *value*, a term that here calls for the adjective "subjective/personal"; it is in fact first of all and originally in relation with a subject that a certain entity reveals a value. It is in the context of a relation in which (at least) one of the two poles is a subject that a value can be established.

From a phenomenological point of view, in fact, "value" is referred to the relation (which can take on a variety of forms) describable in terms of an *action that is beneficial – that produces benefits* – and that is "exercised" by a certain objective reality upon the (inter-)subjectivity. A certain reality thus has a value inasmuch as it is able to exercise a certain "action", or in other words to produce some positive effects. This definition of value is phenomenological in that it is descriptive of the subject-reality connection. In this definition, the (human) subject takes on the role of the receiver, but he or she remains the subject of the action (in this sense, the *personification* through which one attributes to an objective reality the ability to exercise an action is a conscious, explicit and functional metaphor) – that is one who acts, who exercises an act. It is in fact starting from a subject that a certain element acquires a value.

As a subjective relation, Every value takes in and mirrors the complex nature of subjectivity. We can identify four dualities that make up the complexity of the subject.

(a) *Mind-psyche/body* polarity. The subject presents two distinct "ontological dimensions": the physical and the psychical (and/or spiritual, and/or relating to consciousness). Regardless of the scientific or metaphysical "solution" one might want to provide this problem, in a phenomenological perspective (that is, in experience) it presents itself as a problem, because neither pole can be reduced to the other, and at the same time both poles appear to be bound in a constitutive relation. As subjects, we are determined by material conditions which we can define as such precisely because through consciousness we "overcome" them somehow by objectifying them. On the other hand, although we transcend them the moment we thematise them, our limits remain just that: non transcendible (Merleau-Ponty 1945, 1948, 1960). We can identify the following basic statements about the phenomenon of "consciousness" inasmuch as it appears to be irreducible to mere corporeality:

– The ability to *comprehend*. In particular, the ability to *put oneself in someone else's shoes*; this dynamic underlies the development of the subjective consciousness and it can be expressed figuratively as an "exit from oneself" through/aimed at the adoption of someone else's point of view, be it affective or cognitive. It is evidence of an overcoming of pure spatiality seen as

[3]As it sometimes presents itself as an intuition of an existing reality (as in the case of "natural" values that are somehow permanent or at least recurring), sometimes as an institution/production of a "new" reality (as in the case of values that present themselves as such based on *convention*). In actual fact, this subjectification occurs in terms of an interpretation that presents both these characters (descriptive and interpretative), though each time in varying degrees and modalities.

isolation of a point in space within itself, outside any context (a concept that in this case would not be possible). In fact, the act of putting oneself in someone else's shoes implies[4] an experience of space considered as *contemporaneity*, as the simultaneous co-presence of multiple subjects (and therefore, of different points of view of the same reality). This is what the experience of the *present* as the "place and moment of self-awareness" is founded on.

– *Memory* (considered as the permanence of past events in current consciousness) and *imagination* (considered as the ability to produce in current consciousness the image of events in forms that have never been experimented before – an ability on which a typically human desire, a desire that is irreducible to a merely physiological need, is founded) attest the overcoming of pure temporality, understood as the present instant that is consumed in itself, independent from the before and after or – which is the same thing – as the incessant negation of the previous instant which simultaneously negates itself, inaugurating the instant that follows after it.

(b) *Individual identity/intersubjective identity* polarity. The subject is irreducible to the collective inasmuch as it is defined by a belonging that is original to it, so to speak. It is not, in fact, a complete subject that can at a later time or accidentally engage in a relation with other subjects; the subject is, as such, in its essence, in relation with other subjects. At the same time, this collective can never consume in itself the eminently individual nature of the subject. Subject and community are co-original (Merleau-Ponty 1945, 2002; Costa 2010).

(c) *Freedom/necessity* polarity: the subject is both free *and* bound by necessary, "external" laws. In what sense can we understand this "non-absolute freedom" (Merleau-Ponty 1945; Husserl 1952)? Again, the intent here is to provide a phenomenological description, and not a metaphysical one, which would offer a solution – or a dissolution – of this issue. In this sense, we accept freedom as a phenomenon, simply because it offers itself in experience – and in experience freedom appears as such and at the same time it appears limited. The concept itself of "non-absolute freedom" obviously poses enormous metaphysical-ontological issues which we will not be discussing in this chapter. The "non-absolute freedom" that characterizes the subject is explained in two fundamental ways:

– The subject is free first of all in the sense that one of its dimensions (the fundamental one, for which a subject is constituted as such) persists in an identical way (and in this sense we can speak of "subjective identity") even if this occurs in the incessant and inevitable spatial-temporal metamorphosis in

[4]We will not go into the issue of the nature of this relation of implication; it can in fact be explained as an institution by the identification of the consciousness of the present, or vice-versa as an opening to the possibility of the identification starting from the consciousness of spatiality as contemporaneity; or again, as co-originality of the two phenomena of consciousness.

which it is immersed with its corporeality. In this first sense, freedom is "freedom from"; the subjective identity has its own ontological consistency that is independent from external conditions.

– Secondly, the subject is free in the sense that its own personal identity changes – in this sense one speaks of a development of identity –, but this occurs not due to causes which are external to the subject itself, but rather according to its own freedom. In this second sense, freedom is "freedom to": the *creative* ability of the subject is ultimately both the paradigm and the maximum expression of this modality of freedom. It is not my intention to propose a view of freedom that establishes itself in an absolute way, being one's own freedom involved with freedoms belonging to others (cfr. polarity "b"), and being as such limited. But the initiative originating from others will never cause the dispossession of a subject's initiative; rather, it is an incentive to act that the individual freedom moves from. Naturally, in this creative development of a personality, even the corporeal dimension plays an essential role, albeit an indirect one; any physical influence on the subject is mediated by subjective freedom, even though this often occurs in an almost imperceptible way.

Both these modalities of freedom (stable identity and development-creativity) attest to the existence of limits in individual freedom (space-time and the freedom of others), which however do not negate freedom as such. On closer inspection, these are another way of expressing the original and insurmountable nature of the first two polarities: (a) mind/body and (b) individuality/intersubjectivity.

(d) *Actuality/potentiality* polarity: the subject is immersed in a condition which is at once *actual* and open to the *possible*. This fourth polarity derives from the three polarities illustrated above. It not only applies strictly to the subject, but to the entire field of its experience; it expresses the two fundamental modalities of reality as it is experienced by the subject, in other words of the two fundamental modalities of experience. In the subject, however, this fourth polarity is, so to speak, transfigured/transformed by the mediation of the third polarity: freedom/necessity (which, in turn, synthesises and expresses the first two). In the perspective of subjective freedom, the dimension of the possible is, in fact, even more distinctly defined in its being irreducible to the actual, and vice-versa; the possible breaks free from the limits imposed by the actual, and the latter can be seen further as the mere realisation of a previous possible. In other terms, freedom introduces the dimension of a gap in the passage from one to the other (from the potential to the actual and from the actual to the potential), maintaining however the polarity as such.

As they intertwine and overlap, these four polarities characterise subjectivity as such, and are reflected in every value; as an essentially subjective relation, in fact, value retranslates in itself the essential characteristics of subjectivity, thus assuming its complexity as well. The definition of value, at this point, can be integrated by applying the four polarities: a value is a beneficial action, whether it be actual or potential, free or necessary (meaning that it acts upon/through the subjective free will or irrespective of it), exercised by a real element on one or multiple subjects on a psychical and/or corporeal level.

Measure: Between Abstraction and Abstractedness

The objectification belonging to the synthetic representation that occurs with money is applied to the preliminary subjectification of reality; in this sense it is a second-degree objectification, or indirect objectification – mediated as it is by a preliminary, potential *inclusion* of everything in the subjective dimension.

This second-degree objectification is realised as a *measure of value*; this way a subjective relation is objectified and somehow substantialised, and its intrinsic complexity is synthesised in the sense that it is homogenised. Through the reference to a unit of measure, the multidimensionality that defines every value is translated in terms of a mono-dimensionality. This mono-dimensional representation allows for a commensurability among otherwise incommensurable values.

However, subjectivity persists as an implicit basis in the second-degree objectification, since the measure of value does not cancel but rather translates in terms of measure the complexity of the subject that is reflected in every value inasmuch as it is a subjective relation. The money-measure, in other terms, is "sensitive" to the four polarities. All the elements that determine the complexity of subjectivity, in fact, concur to determine the measurement of the values relative to it; in this sense, the money-measure tends to "adapt" to the tensions at play within the polarities and to resubmit them in itself. In applying itself to the polarities, the money-measure absorbs them in itself and offers a representation or a model of them. The measure-synthesis belonging to money is thus a sort of *mono-dimensional complexity*.

The *market* is at once the condition and the modality of actualisation – as well as of expression – of this mono-dimensional representation of a complexity that remains such. The market is in fact founded on the corporeal and psychical, individual and intersubjective, free and necessary, potential and actual nature of every value (and of the subjects for which the values are constituted as such), offering a representation of it that is at once complex *and* mono-dimensional. Like every representation, the measure-synthesis of the value that is developed and expressed as a complex and mono-dimensional representation in the market is in itself an abstraction, in the twofold sense – analysed above – of the term. It abstracts a particular aspect of the total reality and produces with it another "reality".

Being abstract, the measure-synthesis of value is an incomplete and therefore analytical synthesis; it cannot therefore present itself as a conclusive synthesis that is concrete in itself. It is thus an *intermediate synthesis*. However, this synthesis reveals itself to be extremely pervasive, inasmuch as it is potentially applicable to everything; everything in fact presents a value (potential or actual). It is potentially functional to a concretisation of the relation that the subject establishes with reality. By offering a mono-dimensional representation of the complex reality of the value, the measure-synthesis that is realised with money allows us to extend the limits of the subjective perception of the value. But it is precisely in its potential (as a second-degree objectification that absorbs the complexity of the subjectivity in itself) that the tendency towards its own absolutisation is hidden; the more the model is able to reproduce the complexity of the phenomenon that it represents, the more it will tend to conceal its nature as an abstraction and tacitly present itself as the whole truth about the represented phenomenon.

The moment an incomplete (and thus still analytical) synthesis presents itself as a conclusive synthesis, instead of leading to concreteness it produces a form of *abstractedness*. The abstractedness that is applied in a systematic way to the interpretation of the total reality constitutes an *ideological narration*. Thus a convergence occurs between the two forms of abstractedness that were briefly presented in section "Abstraction vs. Abstractedness": the absolutisation of objectivity merges with the absolutisation of subjectivity – meaning the production of a narration that starts necessarily from real experience but breaks free from this and consequently obscures it rather than enabling a better understanding of it.

The abstractedness, however, is only one *possible* development of the abstraction; in what conditions can the abstraction belonging to the money-measure be functional to a concrete knowledge?

Measure and Meaning: From Abstraction to Concreteness

The Anthropological-Semantic Synthesis

The first step towards concretisation consists in reintegrating the measure-synthesis in a larger picture, aimed at an *anthropological-semantic synthesis*. The anthropological-semantic synthesis is obtained by (re)assuming the subjective-human point of view, starting from which it is possible to (re)read the data gathered through a procedure that is founded on the adoption of an impersonal point of view (measure) (Merleau-Ponty 1945, 1960; Polanyi 1958). The overcoming of the abstraction that is thus made possible can also be expressed as a passage from an objectification of the subjective (preceded by the subjectification of the objective reality) to a first form of concretisation that is realised through an anthropologization. Anthropologization (which reintegrates things into a fully human logic) should be seen as the opposite of *anthropomorphisation* (which is the undue attribution of a human *form* to a non-human identity) that is implicit in *abstractedness*. In fact, the moment one translates a subjective relation in purely objective terms, an undue attribution of a fictitious subjective trait to an object is inevitably produced.

The synthetic representation that is obtained through anthropologization is *semantic* because it is a representation whose *signs* do not indicate a *measure* (which is in turn a sign) but rather a *meaning* – inasmuch as they display the relation between reality and the subject in its integrality, i.e. in a fully human sense. In its strict sense, the value is only one of the possible forms of the subject-reality relation – and the measure is only one of the possible representations of this particular form of relation. If instead "value" is understood in a broader sense, it tends to coincide with "meaning"; in this case we will say that the anthropological-semantic synthesis represents values as such (integrally), that is in their quality, and not as measurable things (unilaterally, in only one of their aspects) or, in other words, in their quantity.

While the money-measure carries out the objectification of a subjective relation (value) by means of the reference to an objective unity (which is external to the subject), the anthropological-semantic synthesis brings the result that is thus obtained back to the subjective unity – operating thus a sort of re-subjectification, a return to the subjective experience. The reference to the subjective unity (anthropological-semantic synthesis) allows us to maintain the multidimensionality of the subjective complexity, which is instead lost with the realisation of the reference to the objective unity (measure-synthesis).

This inquiry, aimed at rewriting the measure of value in terms of (human) meaning, can avail oneself of the mediation of one or more anthropological models – such as the one illustrated in section "Value as a subjective relation" – in which the subjective experience is represented in its complex, multidimensional articulation. Before briefly explaining in what sense an anthropological model can enable the development of an anthropological-semantic synthesis, it would be useful to list four points.

1. In case the model is not elaborated, the anthropological-semantic synthesis is realised anyway in the act itself of interpreting and utilising the measure-synthesis for practical ends; in other words, the rewrite in terms of the meaning of the measure of subjective values is realised independently from both the awareness one has of it and the use of models that allow it to be made explicit. There is always a certain anthropological conception at play at the basis of an interpretation of measurement, though it is not necessarily coherent in itself.
2. It is also useful to note how this anthropological conception does not play a decisive role only at the end of the measurement (as inevitably occurs with interpretation and possible application), but rather from the very beginning of the procedure – it is in fact already at play when reality is interpreted as a value (in its strict sense), preceding the measurement.
3. By making this an aware interpretative operation, however, it is possible to improve its effectiveness. In fact, to make the anthropological conception at play explicit by means of a model constitutes a further support in this sense. Knowing the criteria with which one is evaluating something improves the evaluation and simplifies the identification of possible errors.
4. Finally, the anthropological model responds to an additional, particular need: it bears the traces of the anthropological conceptions which are dominant in a certain group, in other words of the particular *culture* of this group – a culture in which the polarities remain identical but in widely varied forms.

Now, to show the need for the integration of the money-measure synthesis in the anthropological-semantic synthesis, we will discuss three fundamental anthropological phenomena that embody the multidimensional complexity of the four polarities: *desire*, *work* and *sharing*. In each of these three phenomena, the four fundamental subjective polarities can be observed. However, each of these phenomena also has one characterising polarity. In the phenomenon of *desire* (the subjective perception of being attracted-by and driven-towards something), the actuality/possibility polarity emerges as dominant; in the phenomenon of *work* (a subjective action finalised at the transformation of a certain reality), the freedom/necessity polarity emerges as dominant; finally, in the phenomenon of *sharing* (an action – whether it be gratuitous or not – of one or more subjects aimed at pooling a certain reality), the individuality/intersubjectivity polarity emerges as dominant. In addition, each of the phenomena is constitutively defined by the primary polarity of corporeality/mind-psyche; all three phenomena, in fact, bear traces of the corporeal and psychical dimensions intertwining as something unsurmountable.

These fundamental phenomena clearly show that the complexity of the subject reveals itself as *paradoxical* (Merleau-Ponty 1948). *Desire*, for instance, is at once the condition for the experience of satisfaction and the experience of dissatisfaction; without desire neither satisfaction nor dissatisfaction can be conceived. The issue has a considerable impact in the sociological sphere; for example, a group of people that is on average satisfied can be evaluated in an opposite way in its ability/disposition to progress in its cultural or technological development. The same thing, on the other hand, should be said about a group of people who is instead on average dissatisfied. A similar argument can be made on the phenomena of "work" and "sharing".

Now, there is no subjective value that does not have some sort of direct or indirect relation with at least one of these three fundamental anthropological phenomena: in this sense, no value is in itself separable in an absolute way from the multidimensional complexity that is expressed – for example – in a desire, in a type of work, and in an act of sharing. The value can be separated from this multidimensional complexity only by means of an abstraction aimed at revealing measure, which is just one of its particular aspects.

It is thus necessary to relate the data offered by the measure-synthesis to the anthropological model so that this data can "speak to us", or in other words – outside of the metaphor – it can mean something on a "human level", in a cognitive and practical sense. As long as the measurement of value is not assumed in the anthropological model (or at least inadvertedly interpreted, without the mediation of a model), it cannot offer any coherent indication from an operative point of view or in other words, it cannot offer an indication of a practical use. In this sense, it would remain a pure abstraction. When such an abstraction is reintegrated into the overall picture, its cognitive gain is not lost – on the contrary, the measurement of value thus completes its trajectory. One could say that the measurement of a subjective value can in turn assume an effective value only when (whether this occurs consciously or not) it is relativised and rewritten in terms of meaning.

The anthropological-semantic synthesis, in fact, does not exclude the measure-synthesis, nor does it remain extrinsic to it by realising a parallelism. Instead, while respecting the specificity and the relative epistemological independence of the measure-synthesis, the anthropological-semantic synthesis includes the former in itself and acquires the information contained in it. The passage through the abstract representation of measure does not constitute a quantitative-numeric parenthesis in a purely qualitative-poetic discussion, but rather a (sometimes necessary) contribution to the comprehension of an inexhaustible phenomenon such as human intersubjectivity.

The relative epistemological legitimacy of the measure of value consists in the fact that every value presents a *tendency to be measured* – a sort of "vocation to measurement". While not being "exhaustible" in it, every value is in itself open to being measured. A value is a (beneficial) effect on our experience and, as such, it needs to be observable *also* from a measuring "point of view". As an "action that is exercised on the subject", a value is in fact revealed even in the dimensions of corporeality, intersubjectivity, necessity and actuality; in this sense, it has a legitimately measurable "side". It would be useful, however, to consider this projection (expression) of value in mono-dimensional terms for what it is: a functional abstraction.

To this end, the anthropological-semantic inquiry does not put the measurable dimensions of the abovementioned polarities in opposition to their non-measurable, "opposite" dimensions (mind/psyche, individual identity, freedom and possibility), but rather it assumes the polarities in their integrality, considering them as original and thematising their reciprocally intertwining dynamic. In other words, it considers experience as corporeal *and* psychical, individual *and* intersubjective, free *and* necessitated, actual *and* open to the possible.

The anthropological-semantic synthesis makes the money-measure abstraction functional to the concretisation of the subject-reality relation, letting itself be guided by the question: "How does this measure-synthesis play in our understanding of the meaning, i.e. the integral relationship between a person and reality?"

This way, it allows us to identify the value of the measure itself – a value that in turn is not exhaustible in an additional measurement. We thus reach a new form of narration, opposite to the *ideological* one; the anthropological-semantic synthesis unfolds as a *non-metaphoric* account (an account that possibly also contains the measure-synthesis) – or, which is the same thing, a *phenomenological* narration – an almost paradoxical expression to indicate a scientific description whose active subjects (i.e. entities capable of actions) are exclusively personal subjects in their full concreteness and integrality.

Only within a narration of this kind, by bringing all the gathered elements back to (inter-)subjective experience, can one define the relations among certain realities and their relative values; as the relations with the subject emerge, the relations (even in terms of proportions, measurements) among things are revealed.

The Original Synthesis

This anthropological-semantic synthesis is none other than the development (in an intensive-qualitative and an extensive-quantitative sense) of an original, pre-reflexive perception of experience as a set of meanings – a perception that constitutes the root of conscious experience, preceding the distinction between feeling and judgement. Such an original pre-reflexive perception is the primary synthesis (Merleau-Ponty 1960), which is not realised as a rearrangement into unity of single analytical elements but a primordial establishment of the connection between subject and reality – a connection that also presides over the superior syntheses (Merleau-Ponty 1945).

The last step towards a concrete knowledge consists in the comparison between the results of the anthropological-semantic synthesis and this primary synthesis – thus making the latter explicit (Merleau-Ponty 1960). It is an operation to be carried out by both the author of the anthropological-semantic synthesis and the reader.

On the primary synthesis level, value is perceived in reference to the subject in its unity and integrality – that subjective unity that determines the need itself for a synthesis as a way to access (cognitive and practical) concreteness in the relation with objective reality. As complex as it may be, the subject in fact experiences itself as one in itself: on this level of the subject's original unity, values are understood as meanings. Values are perceived first of all through a reference to the subject in its integrality (a comparison which the primary synthesis consists in), and only on the basis of this can they be compared *to each other* through the reference to a unity that is external to the subject and that produces a secondary synthesis.

Because of this primordial unity of the subject, the mono-dimensional order of measure and the paradoxical multi-dimensionality of experience revealed by phenomenology prove themselves to be mutually complementary, rather than mutually exclusive. Experience in its original stage includes everything in a synthetic, and not homogenizing way. However the foundational order of the different syntheses must be followed in order to produce concrete knowledge. The order of the syntheses is not purely chronological, but also transcendental (Merleau-Ponty 1945).

The intuition/institution of a value, preceding any type of measurement (the process of subjectification develops and makes explicit this pre-reflexive intuition/institution, as well as acting as a mediator between it and measurement) is realised on the basis of the primary synthesis.

The primary synthesis, however, is not just at the beginning of knowledge: we must refer to it and make it explicit in order to (a) build an anthropological model; (b) reread semantically the measure-synthesis through this model; (c) interpret the anthropological-semantic synthesis itself and bring it to its full concreteness – a concreteness that provides information that is both comprehensive (i.e. including the single elements of a complex situation) and comprehensible (i.e. offering a somewhat unitary image of these single elements, revealing the relational logic that binds them together). Finally, the comparison with this primary synthesis is also the necessary condition to make the results of scientific inquiries communicable and comprehensible outside the limited circles of experts.

Our access to the full concretising synthesis is made possible by the fact that there is already a concrete synthesis at the start of the knowledge process: to this primary synthesis one needs to compare all the further syntheses and the analytical data.

The twofold meaning of the title of this chapter is thus revealed: “building knowledge” indicates both the process of building knowledge (by means of abstract syntheses/representations) and a knowledge that builds (concretising synthesis), in other words, a knowledge that allows us to enter into a relation with reality in all of its complexity.

References

Bergson, H. (1938). *La pensée et le mouvant*. Paris: P.U.F., Paris.

Costa, V. (2010). *Fenomenologia dell'intersoggettività. Empatia, socialità, cultura*. Rome: Carocci.

Derrida, J. (1962). *Introduction à l'Origine de la Géométrie de Husserl*. Paris: Presses universitaires de France.

Di Martino, C. (1998). *Il medium e le pratiche*. Milan: Jaca Book.

Di Martino, C. (Ed.). (2013). *Attualità della fenomenologia*. Soveria Mannelli: Rubbettino.

Havelock, E. A. (1986). *The muse learns to write: Reflections on orality and literacy from antiquity to the present*. New Haven: Yale University Press.

Husserl, E. (1950). *Cartesianische Meditationen und Pariser Vorträge*. Den Haag: Nijhoff.

Husserl, E. (1952). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch*. Den Haag: Nijhoff.

Husserl, E. (1959). *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*. Den Haag: Nijhoff.

Husserl, E. (1966). *Analysen zur passiven Synthesis*. Den Haag: Nijhoff.

Husserl, E. (1976). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie*. Den Haag: Nijhoff.

Husserl, E. (1987). *Aufsätze und Vorträge (1911–1921)*. Dordrecht: Nijhoff.

Maddalena, G. (2015). *The philosophy of gesture. Completing pragmatists' incomplete revolution*. Montreal: McGill Queen University Press.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Gallimard.

Merleau-Ponty, M. (1948). *Sens et non-sens*. Paris: Nagel.

Merleau-Ponty, M. (1960). *Signes*. Paris: Gallimard.

Merleau-Ponty, M. (2002). *Causeries*. Paris: Seuil.

Polanyi, M. (1958). *Personal knowledge. Towards a post-critical philosophy*. London: Routledge & Kegan Paul.

Scheler, M. 1986. *Phänomenologie und Erkenntnistheorie*. In *Schriften aus dem Nachlass*, Band I, *Zur Ethik und Erkenntnislehre*. Bern/Munich: Franke.

Part II
Methodological Issues

Chapter 3
Socio-economic Statistics for a Complex World: Perspectives and Challenges in the Big Data Era

Marco Fattore

Introduction

This chapter addresses a topic which is gaining increasing interest in socio-economic statistics and that will play a central role in what in the future could possibly be called "information-based policy-making". The topic is that of big data and data science and of their potential effects on next future socio-economic statistics (Landefeld 2014). Although no relevant applications have been produced yet at "official level", the use of big data and the applications of data science methodologies are in fact opening new avenues to the way socio-economic statisticians may extract information from different data sources and provide it to decision-makers. It is surely not easy to write a chapter on this theme. The topic, "big data and data science", is in fact a broad concept and cannot be considered as a scientific discipline yet, though it is attracting research efforts from many different sectors and many people are contributing to its development. It can be addressed from many points of view and different aspects (technological, methodological, epistemological. . .) could be underlined, giving different alternative pictures of the argument. Given the aim of the book, here we simply outline some basic concepts pertaining to big data, to clarify why socio-economic statisticians should be interested in this area, to help them realize its potentialities and criticalities and to stress the conceptual differences with respect to "traditional" statistical analysis. The chapter is somehow different from others in this book. It is non-technical and is based on reflections and experiences of the Author, who has been involved in didactical activities and in real projects pertaining to big data analysis and data science. As a consequence, the text may seem more "subjective" than other contributions in the volume. This is true and partly unavoidable: the attempt is to

M. Fattore (✉)
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
e-mail: marco.fattore@unimib.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_3

collect and share what I could learn on the topic in the last years, motivating why I think big data can open new horizons to applied statistics, in the socio-economic field.

The chapter is organized in two distinct and apparently independent parts. The first section presents and discusses some new challenges that socio-economic statistics has to face, in order to provide meaningful, useful and valid information to policy-makers, in increasing complex and dynamical socio-economic systems. These challenges have different nature; some are technical, others are more conceptual, but all of them refers to the need of measuring, assessing and representing socio-economic contexts in more appropriate and satisfactory ways. The second section goes through the topic of big data and data science, viewed as the basis for developing "informational" processes. Different issues will be touched upon, without entering into details, to provide a general picture of this broad world. The link between the two parts of the chapter is easily explained: if exploited through proper data science methodologies, big data can (or will be able to) provide new roads for socio-economic statistics to face the challenges posed by the complexity of socio-economic systems. However, this requires some change of attitude, since big data are quite different from simply "large databases" and their exploitation cannot be addressed just relying on traditional analytical paradigms.

What does "big data" mean? which is the potential effect of the exploitation of "big data" for socio-economic statistics and indicators construction? can big data empower our capability to build synthetic and meaningful views on complex and dynamical socio-economic systems? which approaches and methodologies are required to exploit big data? which is the role of technology? These are some of the questions we will consider, as a way to motivate the reader to venture into this somehow new area of the statistical discipline.

Some Challenges for Socio-economic Statistics in Complex Socio-economic Systems

We begin this paragraph mentioning a debate occurred on Italian television, pertaining to pension reform in Italy. The dispute involved an "old" economist, author of the main reform of the Italian pension system, and a "young" self-employed woman. The debate was about the pension system and the chance for young people to get a satisfactory pension in the future. Given the structure of Italian pension system, which is historically designed for an economy based on permanent jobs, the young worker stressed the difficulties and risks, for people whose working path comprises flexible sequences of different contracts, with possible unemployment periods between two consecutive jobs. As an answer, the economist reaffirmed the global economic and financial equilibrium of the system, capable to guarantee pensions to future generations. We do not enter into the technical discussion, nor we take sides of one discussant against the other. What makes the debate interesting for our purposes is that it essentially represents a conflict between a "structurally differentiated and dynamical" point of view and an "aggregated" point of view. In fact, the pension system as a whole may well be in equilibrium, but different classes of workers can be intrinsically treated differently, at the extent that two workers paying the same global amount of contributions over their working lives, may finally benefit very different pension levels. Not to mention that the contribution load charged on self-employed workers may have consequences on their investment, purchasing and savings power, with possible effects at macro-economic and social level in the long run. In practice, on average the pension systems could even work, but its equity and real effectiveness can be argued.

This example introduces well the basic challenge for current and future socio-economic statistics. In a complex and structurally dynamical socio-economic system, any "on average" description of socio-economic facts is intrinsically ineffective and inadequate to understand and to try to govern societal dynamics. Aggregate pictures are inherently unfaithful representations of socio-economic reality and cannot be used for policy-making and policy-evaluation. Another example helps to clarify this statement. One of the most relevant topic in socio-economic statistics is the study of income distribution, monetary inequality and poverty. Poverty is usually described in terms of Head Count Ratio (fraction of poor people over the total population) and in terms of Gap (average distance, in monetary units, of the poor from the poverty threshold). As the "beyond GDP" perspective on well-being spreads across economists and policy-makers, the point of view on poverty changes as well. People can be deprived in different ways and for different reasons and policy-makers have to contrast different "kinds of poverty" each of which requires different policies. Aggregate deprivation measures get replaced by deprivation patterns: people occupy the "space of deprivations" (which is not simply the "space of income") in a distributed way, ruling out the possibility to neatly separate groups in a simple poor/non-poor perspective. Metaphorically, precise poverty measures must be flanked by "poverty geographies", that provide structural views of deprivation.

It must be clear that we are not arguing against precise measurements and indicators in themselves. Socio-economic indicator systems are a fundamental basis for any sound description of socio-economic facts and for policy-making and we will discuss why and how big data can be useful in this direction. However, we want to stress what we consider a fundamental point: providing synthetic views of socio-economic phenomena need not mean computing aggregated indicators. The description of complex socio-economic phenomena must somehow preserve complexity. Complexity can be simplified and reduced, but it cannot be erased, without depriving final statistical products of their information content and hence of their value for both social scientists and policy-makers. Perhaps, the most appropriate term to capture what we mean by "synthetic information", in this context, is "stylization", i.e. the capacity of a statistical object to reproduce the fundamental structure and shape of the phenomena under investigation. Depending upon the object of study, such a stylization may require a simple indicator, a dashboard, a map or other graphical devices... and suitable statistical processes capable to feed them with meaningful information. In any case, the "informational" value of statistical processes and outputs depends upon their capacity to achieve such structural stylizations. The traditional architecture of socio-economic indicators must be enlarged, supplemented and complemented with other statistical objects, so as to achieve a more comprehensive and actionable picture of the socio-economic system.

Broadly speaking, there are at least three typical perspectives that contribute to complexity, in socio-economic statistics:

1. Temporal perspective. One of the main features of current socio-economic systems is their structural dynamics. As a consequence, the importance of promptly intercepting the signals of structural changes prevails on the attempt to precisely forecast macro-indicators. This is of a fundamental importance in view of information-based policy-making, based on a continuous system of information and feedbacks flows.
2. Spatial perspective. Socio-economic phenomena may show high variability also at territorial level (consider, for example, the different living conditions in different areas of the same town). So, not only policy-makers need to adapt their actions over time, but also they have to tune them differently at territorial level. Socio-economic statistics must then provide disaggregated information, capable to effectively "resolve" spatial differences.
3. "Structural" perspective. With this expression, we refer to the existence of a variety of features that identify social groups, other than classical socio-graphic attributes. In other words, statistical populations must be investigated and analyzed involving highly multidimensional indicator systems. To give an example, consider the analysis of poverty and the definition of related policies during the recent economic crisis. Roots of poverty may differ from individual to individual. A highly educated and trained person losing job due to a company crisis may suffer temporary deprivation, but his condition is very different from that of a non-qualified worker, driven by the crisis to the borders of the job market. Past job experiences, schooling level, health status, familiar conditions, social integration... are just some of the aspects to be considered in order to identify homogeneous social groups and to define specific poverty contrasting policies. Socio-economic statistics should then be able to provide such an articulated picture of society.

To improve existing indicators and to supplement them with new kinds of information and representations, the first issue is to search for adequate data sources. Schematically, these can be identified as follows:

1. Official statistics, produced by official institutions and statistical bureaus. They are essential and provide the reference statistical framework, but currently they usually do not have the necessary timeliness and spatial granularity.
2. Administrative data. They provide very useful information, at individual level, even if they may also suffer of quality and accessibility problems (see later in the

chapter). However, in the future they could be a fundamental source of information for socio-economic statistics.

3. Data from the digital society. This is a broad concept comprising an extremely large number of possible sources, ranging from web and social networks, to data from telecommunication companies, large retailers' networks, mobile applications and, in the next future, from the so-called internet of things. These sources, which are usually referred to as "big data" can be considered as a "space of data", or a "data ecosystem" we contribute to produce daily and we are embedded in. Potentially, they allow to "observe" detailed socio-economic behaviors almost in real-time, but at the same time their exploitation involves very complex problems at methodological and technological level.

Combining together all of these sources provides socio-economic statistics with the greatest amount of data in human history. So the challenge becomes how to turn this data flows in valuable information. This raises many issues, some of which will be touched upon in the second part of the chapter. It is however incontestable that the use of big and administrative data paves the way to a potential revolution in the way socio-economic statistics produces information and in the way policy-makers can address the government of socio-economic systems, the definition of policies and the assessment of their effects. There are at least two ways big data can be revolutionary. The first is, in a sense, technical. Big data can provide additional information to better estimate official indicators and to improve their quality. An example is the so-called "nowcasting", where heterogeneous information from the digital world and/or from administrative data can be used to determine economic figures (e.g. GDP), shortening the production of official statistics that are usually made available with delay. In this case, alternative sources feed a classical indicators system, improving its quality and information power. The second way big data may have impact on the production of socio-economic statistics is, instead, much deeper and touches the structure of the socio-economic information architecture. On the one hand, new indicators and new kind of synthetic information may be produced (as in the example of poverty geographies) enlarging the "information space" available to social scientists and policy-makers. On the other hand, since big data can realize a system of continuous feedbacks among agents and between them and decision-makers, the same availability of information becomes itself a political leverage, in that it may affect agents' behaviors. For example, as information on the quality of single schools or single hospitals produced by users spreads across social networks, it de facto realizes an evaluation system that modifies users' choices. Any agent's act becomes an "informational act", in that it introduces an information in the system. Cleaning, organizing and making available such information in an organic and structured way and in (almost) real-time could be a new goal for socio-economic statistics, in complex environments, where the need for adaptation prevails on the possibility of forecasting. Although this scenario may seem little realistic at present, some first signs of this tendency can be already found (consider the number of restaurants exhibiting their votes on Tripadvisor and the increasingly spreading habit to check it before choosing locations) and surely this represents a relevant part of the future of socio-economic statistics.

We synthesize the above discussion quoting from a recent and interesting paper by Letouzé and Jutting (2015):

> Big Data needs to be seen as an entirely new ecosystem comprising new data, new tools and methods, and new actors motivated by their own incentives, and should stir serious strategic rethinking and rewiring on the part of the official statistical community. [. . .] The key point that cannot be stressed enough is this: Big Data is not just big data. It is about qualitatively new kinds of data regarding people's behaviors and beliefs, new kinds of tools, and new kinds of actors. If any discussion of the applications and implications of Big Data for official statistics (and development) is to be meaningful, Big Data must be approached and conceived of as an ecosystem – a complex system – made up of the data, the tools and methods, and the actors. Asking whether and how Big Data may or should affect official statistics is not only about whether and how official statisticians should or should not use Big Data to produce official statistics. It is about whether, why and how, official statisticians and systems should deal with Big Data as an emerging complex ecosystem.

The above discussion must be intended as a way to motivate the interest of statisticians, social scientists and policy-makers towards big data and their potential use in applied statistics and indicator construction. Unfortunately, the term "big data" is often used in a too generic way and some confusion exists on it and its meaning. For this reason, in the next paragraph, we precise what big data are and which are their main features.

Big Data: More Than Large Databases

According to Letouzé and Jutting (2015), big data can be described as "data sets that are impossible to store and process using common software tools, regardless of the computing power or the physical storage at hand" or, quoting Mike Horrigan, again reported by Letouzé and Jutting, "as non-sampled data, characterized by the creation of databases from electronic sources whose primary purpose is something other than statistical inference." (Scannapieco et al. 2013; Horrigan 2013).

Hence the term "big data" refers primarily to the huge amount of data nowadays available from different sources and in different sectors and areas, both scientific and commercial. Two typical examples come from the telecommunication sector and from the social network world. Consider the amount of bits a telco company may produce and store in a year or think about the number of messages posted on social networks in a month (realize that, as this chapter is being written, the number of tweets per day on Twitter is about 500 millions). So "big data" stems first of all for "a lot of data". However, in a sense, this is not their primary feature, in a socio-economic statistical perspective. Big data have many other features and peculiarities, in addition to their "bigness", that make them very different from simply large databases and also really something new to deal with. They are usually characterized by three "Vs":

- *Volume*, referring to the amount of data available, as commented above.
- *Variety*, referring to the different sources they come from. This is a very important feature representing both a potentiality and a criticality of big data. In fact, big data usually arise from the "cooperation" of sources of different nature (e.g. classical surveys, the web, classical databases mobile applications, administrative archives...) regarding both the kind of data and the goal they have been designed for (not primarly for statistical purposes). This means that, typically, big data are "non-structured", i.e. they do not have a common structure and do not usually arise as a consistent and organic "data system". We will come back later to this fundamental issue, investigating the consequences of this matter of fact on statistical processes.
- *Velocity*, referring to the speed of data flows big data often come from. In fact, big data are not only obtained from integrating large statistical archives; more often they are "living" and even real time flows. To make an example, consider data coming from accesses on a web site, pertaining to users' behavior, interests, used devices (desktop, mobile...) and collected on a continuous basis.

Volume, variety and velocity may seem just technical features, however, they have relevant consequences on the potential value of big data flows and on the approach and methodologies that must be followed and employed in order to exploit them. More important, they have great impacts on the paradigm of traditional data analysis, that may well prove inadequate to extract value from them. This is not saying that traditional statistical processes must be disregarded as ineffective; the point is that big data not only are a new data source for old statistical analyses; they also enhance new kind of analyses and information extraction. This requires also a change of attitude towards data and their exploitation, as revealed by the three "Cs" that complement the three "Vs" and characterize big data from complementary and less technical perspectives (Letouzé and Jutting 2015):

- *Crumbs*, "identifying Big Data as new kinds of passively-generated individual and networked "traces of human actions picked up by digital devices". These "digital breadcrumbs" have the potential to paint a picture of some aspects of the social world with unprecedented levels of detail and shades. Their fundamental revolutionary nature is qualitative."
- *Capacities*, referring to the power of big data to provide "insights", rather than definitive views, on many socio-economic facts.
- *Community*, i.e. "Big Data must also be considered as referring to the people and groups "making use" of crumbs and capacities." This is to mean that within the huge big data ecosystem, there are different kinds of users exploiting different kinds of data, which are in turn own by different kinds of subjects (often private companies acting within different markets).

Hence, big data are both a technical concept and a "mind-set", calling for new approaches in data analysis. In the following paragraphs, we will discuss the interconnections of these aspects with socio-economic statistics, in view of information-based policy-making. We do this, quickly going through some topics.

Although we can touch them just briefly, we try to provide a comprehensive picture of the theme.

Methodological Issues in Big Data Analysis

We have already mentioned some possible practical applications of big data to socio-economic statistics. Other (possible or attempted) examples pertain to the analysis of consumptions based on data from credit cards or from retailers; the analysis of tourist or migration fluxes and patterns, using mobile phone and email data; the measurement of consumer confidence level or of subjective well-being through sentiment analysis on social networks. As usual, imagination is less fantastic than reality, so only future applications will be able to reveal the far-reaching consequences of this data revolution. There are however some typical features in big data analysis that we want to focus upon, commenting over some real experiences we have been involved in, in different application fields.

Consider the use of geo-referentiated open data about shops, schools, museums, pubs, hospitals in a given territory. These can be used not only to locate single points of interest, but also to identify the "vocation" of particular urban areas. The presence of particular mixes of shops and/or services may qualify a suburb as suitable for a family with little children, or for a businessman, or for elderly people. This kind of information, which is built upon elementary geo-referentiated records, may be used to choose where to settle, or to measure some sort of territorial quality-of-life indicator and/or to identify potential social problems deserving interventions by territorial institutions. As a different example, consider data on accesses of users to an e-commerce site. Information on the device used to access (a pad, a desktop, a smartphone...) can give insights on the age of users. Moreover, knowledge on the kind and version of the operating system installed on the device, or on the browser (quite technical details, apparently) may provide information on the attitude of users towards technology. Why not to use these infos to build some indicator on the technological level of a country?

These simple examples shade light on a typical aspect of the use of big data in the socio-economic field. The way big data provide information is often "indirect"; they are proxies of the information of interest and are linked to it in a "lateral" way. Consequently, on the one hand, some "creativity" is needed to identify possible usage of big data and to "structure" them towards information goals. On the other hand, appropriate statistical procedures must be employed to get reliable and valid results. In general, big data cannot be seen as a "controlled" source, as discussed in the next paragraph about quality issues. Like administrative data, they may fail to satisfy some basic requirements on data quality (for example, Twitter users cannot be taken as a representative sample of population, so information extracted from Twitter must be carefully considered). In this respect, they are very useful in a descriptive perspective, to build pictures of socio-economic phenomena, or as a way to anticipate changes and to extract signals in advance. They are probably less useful, at least at present, as a source for precise statistical modelling and causal inference. However, when the goal is to govern complex systems, the ability to identify the "directions of change" is perhaps more important than to measure indicators precisely. This is somehow a change of paradigm that should not be considered as a limitation imposed by scarce quality; rather it is a manifestation of the complexity of the data generation processes, i.e. of the underlying socio-economic life, that reflect into data complexity.

The nature of big data has profound consequences on the statistical methodologies needed in extracting information out of them. The variety of data sources (and the variety of questions that can be addressed, as well) make the whole world of statistical procedures and models potentially useful, so it is not easy to highlight specific techniques as the most important. There are, however, some typical statistical tools that are becoming a standard part of the data science process for big data exploitation. These are mainly (but not exclusively) borrowed from the machine learning field and prove useful in managing data systems with hundreds or even thousands of features, as typical with big data. In the effort for reducing complexity, highly-dimensional datasets usually must be treated along two directions. First, attempting to link individual features to some outcomes of interest. This is, for example, what happens when one wants to investigate the relationship existing between subjective well-being and objective achievements, as reported in official surveys like EU-Silc. Even if this kind of surveys cannot be properly classified as big data, as to the number of records, the variety of investigated dimensions make them a complex source anyway. Moreover, data are mainly scored on ordinal scales, with possibly different number of degrees, while some may be numeric. In these situations, classical regression models may be scarcely effective, being designed for simpler data structures. Other tools prove more effective, namely classification or regression algorithms, like random forests and boosting. We briefly introduce them hereafter, to give an idea of the spirit big data can be addressed by.

Random Forests have been developed and proposed by Breiman (2001) as a way to overcome the limitations of classification/regression trees, which may be affected by overfitting and scarcely robust to generalization. In a random forest, a sequence of classification/regression trees is built, basically after the classical splitting algorithm. However, trees differ each other since:

1. each tree is built on a different bootstrap sample extracted from the original population;
2. at each split, a fixed number of features (covariates) is extracted and the optimal split is searched for within them.

As a consequence, each tree is a "weak" classifier, i.e. it has poor predictive performances, and its predictions are "weakly correlated" (informally speaking) to those of the other trees. In practice, when a statistical unit must be predicted based on its feature profile, it is processed and predicted by each tree in the forest. The final prediction is then obtained by a majority rule (or by averaging, in the case of regression forest). As paradoxical as it may seem at first, the whole set of trees, i.e. the "forest", turns out to be a good classifier, even if single trees are not (clearly, its performance depends upon the intrinsic informative power of the data). Otherwise stated, a set of weak classifiers may result into a good classifier. More important, the forest is less affected by overfitting, so it generalizes better to new data.

A different approach to the same goal is pursued by "boosting" algorithms (Freund and Schapire 1996; Schapire et al. 1998). These are a way to improve the performances of a given classifier (i.e. to "boost" its performances), by producing a sequence of classifiers specialized in predicting over different segments of the feature space. In practice, new classifiers are built specializing them on cases misclassified by previous classifiers.

The research on these kinds of learning algorithms is going on and the debate on them is quite fired. It is not the aim of this contribution to enter into these details. The reason why these (and similar other tools) are of interest in big data analysis is primarily for their robustness and capability to analyze complex data structures. They produce "black-boxes" which may be very useful for classification or prediction; however, they are less useful for interpretation goals. In fact, although these algorithms provide also information on the importance of single features in explaining the outcomes, they are not designed for modeling purposes. This is the price to pay for algorithms capable to deal with large numbers of covariates, even collinear. Notwithstanding this, machine learning procedures may be very effective in identifying the structure of complex phenomena, providing essential pictures of societal facts.

The other typical perspective of big data analysis is that of pattern recognition and "clustering", i.e. of the identification of homogeneous subgroups of statistical units. Classical cluster analysis usually tries to identify a few neat clusters, i.e. a few homogeneous and well-separated subgroups of subjects, to partition the whole population. This approach may be inappropriate when dealing with complex datasets. Societal phenomena are inherently multidimensional, often vague and fuzzy, and a too rigid approach to clustering may well be at odds with their nature. The point has been already raised in the first part of the chapter: patterns take the place of neat clusters; correspondingly, statistical tools oriented to pattern extraction must be preferred to classical clustering procedures. Among algorithms of this kind, we mention one of the most famous and applied, the Self-organizing map (SOM) algorithm, developed by Kohonen (Kohonen 2001).

Technically, SOMs are a kind of neural networks, designed to achieve non-linear dimensional reduction of the data. In practice, they can also be seen as an extension of the k-means algorithm. As the k-means procedure, also SOMs try to identify a set of centroids (here called "codebooks") around which to aggregate statistical units. The main, and crucial difference with the k-means algorithm is that, in SOMs, codebooks are placed on a regular spatial grid in such a way that neighboring codebooks tend to be similar. This is achieved through an iterative algorithm, leading to a "self-organization" of the map: as the algorithm runs, the map progressively structures itself, modifying codebooks until an ordered pattern is achieved (if possible, given the data). The final result is a visual map, in which clusters (i.e. codebooks) are placed in such a way to produce a "geography" of the input feature space, by means of a tessellation of the planar grid. Metaphorically, the regular grid acts as the sensor of a digital camera: as this reproduces a "map" of the intensity and color of the input light, so the SOM reproduces a planar, but non-linear, image of the input multidimensional density distribution. A part from technicalities, the reason why SOMs are of great interest in the context of big data is that they allow patterns in the population to be identified. In fact, while SOMs can also be used to find out a few clusters, in a k-means spirit, they are very often computed based on grids of hundreds or even thousands of codebooks. These maps are sometimes called ESOMs, i.e. Emergent SOMs, to stress that their goal is not to consider any codebook as a stand alone cluster, but to identify meaningful areas and patterns within the map. Projecting statistical units onto the map, it is possible to "see" the population within the geography, really getting a stylized image of the data. Comparing the projections of different subgroups of units on the same map, and/or observing projections over time, one can also get relevant information on the structural differences among subpopulations and on structural dynamics. SOMs are currently used in socio-economic analysis, for example in the study of well-being and deprivation (Lucchini et al. 2014; Lucchini and Assi 2012).

The machine learning and statistical procedures described insofar are just a "pinch" of a huge statistical toolbox. They have been mentioned since they are among the most used techniques, but also since they are prototypical of how big data are addressed. A SOM may not be properly termed an "indicator", in the usual sense this word is used for. But in fact it is. It is clearly non-numerical, but it expresses a latent trait as indicators do. Since the trait is inherently complex (for example, a pattern) the indicator cannot be simple and must possess an internal structure capable to reproduce it. This is what a self-organized map does.

Abstracting from the examples and the tools just mentioned, we remark that it is the conceptual approach to big data analysis that marks a difference with respect to traditional statistical processes. Rather than the search for "specific information hidden in the data", exploiting big data basically involves a "structuring process" in which meaningful information emerges out of a quantity of "bits". Certainly, this is not saying that more traditional statistical approaches are useless. As already remarked, based on the goal of the analysis and on the kind of data, the most appropriate set of statistical methodologies and tools must be employed, let they be more traditional or more innovative or a mix of both.

We conclude this paragraph, commenting on another aspect of big data that has been already mentioned. Big data often come as continuous data flows. This allows for statistical analysis to be embedded into statistical processes, implemented and engineered in such a way that a continuous flow of information gets delivered to final users. The idea of "engineered statistical processes", or "informational processes", is captivating and we think it will be one of the forms big data analysis will assume in the future. This, however, calls for some caveat. As the informational power of big data and their influence on socio-economic life grow, also potentially negative effects of bad information may grow as well, given the leverage effect of modern technologies. This must enforce the attention paid to the problem of data quality and to keep under control the statistical processes implemented on big data, through careful and continuous inspections of quality and performance indicators.

Big Data, Poor Quality?

We briefly touch a somehow technical but important topic in big data exploitation. A large amount of data need not mean a large quantity of information. In this respect, dealing with big data raises the same problems on data quality that statisticians have been dealing with in their daily work since ever. However, the volume and the variety of big data make the quality problem even more critical than usual. First of all, any data source converging in a big data system must have its own proper quality profile. Unfortunately, in many cases the data generating processes are not under the control of the analyst, so that quality may often be assessed only ex-post, and not "built into the data". As an example, we mention a real case pertaining to the use of open data provided by Italian municipalities or other territorial institutions. The goal of the study was to identify relevant touristic areas of the regional territory, using open data pertaining to museums, ancient churches, ancient buildings and houses and the like. After collecting open data, it was clear that:

- The description of the geo-referentiated entities was not given in a common standard. Different terms, different kind of geographical coordinates, ambiguous and inconsistent classifications made impossible to get a satisfactory and usable dataset, without a big manual effort, to make classifications uniform.
- Some records were affected by missing values, to be manually filled using alternative sources.
- Some territorial areas were not covered by the open source data, limiting the scope of the information extractable from the datasets.

These kinds of problems are prototypical of what can happen in the daily practice with big data. Statisticians know well that consistent classifications are the basis of any data analysis process; however, when administrative data are at hand, like those in the cited example, even such an elementary issue cannot be taken for granted.

In other cases, reliability and validity may be problematic. For example, in sentiment analysis (a topic that is increasingly attracting the interest of political and administrative institutions) the evolution of terminology and language can make estimation processes unstable, even over short spans of time. To make another example, from observing individuals' buying behavior through fidelity cards, or on e-commerce sites, one could get information on the economic status of subjects, social groups, territorial areas... However, the large amount of inter-related features affecting buying behavior may lead to poor validity of statistical measures, which may reflect latent traits different from what presumed. Other fundamental problems may also come from the scarce representativity of samples obtained from uncontrolled sources that, consequently, may be very useful to get insights, but poor for inferential purposes.

All of these quality issues can be exacerbated when different data sources get integrated in a "cooperating data system", whose quality depends both on single sources and on the "combinations" of their quality profiles. As a result, the assessment and the control of data quality, and thus of data value, proves to be really a critical step in big data analysis.

Technological Issues

Big data analysis involves the technological problem of managing large amount of data. Given the complexity and the volume of such data sources, technology does play a critical role, in order to enhance data treatment and information extraction. For sake of completeness, we provide some hints on the technological aspects involved in the exploitation of big data.

There are at least two main aspects, where technology is critical:

1. *Data storage*. Big data systems may comprise billions of records with thousands of features (covariates) that cannot be easily stored in classical relational databases. The volume of data, moreover, is increasing, making the storage process even more critical.
2. *Data access and analysis*. More important in view of analytical processes, big data volume poses problems when data must be accessed and analyzed, particularly if (almost) real-time applications must be feeded with information extracted from big data systems. This is leading to the use of new database models, overcoming the SQL paradigm in favor of NoSQL systems, that guarantee better performances in indexing and information retrieval. However, NoSQL systems are often not oriented to statistical analysis and data often must be extracted and properly reorganized, prior to the analytical process.

Clearly, the impact of technology depends upon the amount of data to integrate, manage and treat. In some cases, when one deals with a few hundred thousand records, there are no particular technological problems to face. Nevertheless, it must be carefully considered that in big data analysis, the "data preparation step" is inherently more critical than in traditional statistical activities. Many sources, data quality issues, increasing data volumes, performance constraints... all of these aspects must be taken into account in the design of the process, realizing that changing data architectures, afterwards, may be a complex and expensive task. Even if such technological aspects may be considered of secondary importance form a statistical point of view, they are indeed crucial. In the case of big data, statistics is enabled by technological systems, whose performances and characteristics may affect heavily the quality of the final statistical production.

Data Visualization

Before concluding the chapter, we want to touch upon a final aspect that is sometimes considered as one of the key leverage to turn big data into useful information. The topic is that of "data visualization" which, together with the concept of "story-telling", has become one of the keywords of data science. In fact, when big data – especially coming from web sources – first became available, visual designers and people interested in social communication began to develop forms of visual story-telling, as an evolution of simpler infographics. Stated informally, their goal was to use visual grammars to make data intelligible to final users, through more or less sophisticated and often interactive visual representations. In the following, we want to briefly discuss the role of visualization in the context of big data and data science, criticizing the centrality it has been assigned in the last 10 years, but also stressing its role and potentialities in the construction of effective informational processes oriented to decision-making.

First of all, data visualization is a very broad theme, ranging from scientific visualization, to infographics, from simple graphs (scatterplots, pie charts...) to complex representations of multidimensional data and dashboards, often employing visual metaphors and sophisticated visual codes. Here we are not addressing technical topics, nor we will enter into the discussion on the relationship among data types, visual grammars and final users. After noticing that the technical apparatus of visual designers can be of great utility within informational processes, we want to discuss which is the correct role of visualization in the context of big data.

The main goal of informational processes in the socio-economic field is to provide policy-makers with sound and meaningful representations of socio-economic facts. Complexity of the socio-economic system prevents effective representations to be given in terms of a few univariate indicators. Evolving patterns, complex interconnections among covariates and intrinsic fuzziness cannot be communicated to policy-makers effectively, without drawing upon suitable visualization techniques and tools. In this respect, visualizations can be compared to "car tires", in that they "transmit" the power of information to the final user. Many tools exist, even open source (e.g. R and its visual packages, Gephi – for network and graph representations, Raw, to mention a few resources accessible through the web), that allow statisticians to get good visualizations, without being specialized in the visual discipline. However, the role of visualization has also been misunderstood. In fact, visual design often tends to apply its tools, in an effort to make information directly emerge from data. In other words, visual design tools are directly "connected" to databases, to generate visualizations for story-telling purposes. In so doing, however, it is precisely the story to tell that is missing. In fact, the "story" cannot emerge without interconnecting statistical variables using appropriate statistical tools. So, paradoxically, data visualization reduces to a mapping between two formal codes: the data on the one hand, and graphical grammars on the other. Statistical variables are linked to graphical features (colors, line length or thickness...), in a sort of recoding of original bits into “graphical” bits. Unfortunately, bits remain bits, even if under a different form, and no clear picture is likely to emerge. In between data sources and visualization, it is the statistical layer which plays the fundamental role of information extraction. It is properly this information that must be visually represented, using infovis tools. Once “the story” has been extracted, its representation may assume different forms, depending upon the nature of the information and the needs of the user. Final users want simple and interpretable visualizations. Graphical representations should then minimize the cognitive effort of the user, avoiding too hard “decoding” activities. Often, the key to achieve this is not only to choose the right visual object (a suitable kind of graph, for example), but to extract from the data the right synthetic and meaningful information. So, even in view of effective visualizations, the main effort must always be concentrated on the statistical layer, which produces the information to be communicated.

Conclusions

In previous paragraphs, we have discussed the possible role of big data in next future socio-economic statistics, with the aim to raise interest towards the topic. Big data, considered as an ecosystem of technological, statistical and communication processes, represent a revolution for socio-economic statistics and policy-making. Indeed, there are many issues to address and solve, before big data can exploit all of their potential, but their importance can hardly be overrated. Exploiting big data requires some change of attitudes by socio-economic statisticians, particularly since big data are basically non-structured, i.e. they are not a priori designed and organized for inferential and knowledge purposes and to answer specific questions. Statisticians must also acquire new competencies, pertaining to analytical tools, to technologies and to visualization and communication, so as to be able to design and develop informational processes to feed the activities of social scientists and policy-makers. Traditional indicators cannot suffice to provide faithful representations of complex socio-economic systems. On the one hand, they must be improved, in terms of spatial and time “resolution”; on the other hand, they must be supplemented with other “indicators”, i.e. with new statistical objects, conveying synthetic, yet non-aggregate, and structural information. In both cases, big data can be a key factor, providing the right inputs to improve the statistical production of socio-economic indicators. The leverage effect of communication technologies makes information a “critical good”, in that it may have (and usually has) deep impacts on the social system, producing collective phenomena and affecting people’s choices. So data and information quality becomes of public relevance, given its potential consequences on socio-economic and, ultimately, democratic processes. In this perspective, socio-economic statistics is going to become a “part of the socio-economic system” and not just a “discourse on the socio-economic system”.

References

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Freund, Y., & Schapire, R.E. (1996). Experiments with a new boosting algorithm. In *Proceedings of the thirteenth international conference on machine learning*, (pp. 148–156). Burlington: Morgan Kaufmann.

Horrigon, M.W. (2013), "Big data: A perspective from the BLS". *AMStatNews*, http://magazine.amstat.org/blog/2013/01/01/sci-policy-jan2013/

Kohonen, T. (2001). *Self-organizing maps*. New York: Springer.

Landefeld, S. (2014). "Uses of big data for official statistics: Privacy, incentives, statistical challenges, and other issues", International conference on big data for official statistics Beijing, China, 28–30 Oct 2014, United nations global working group on big data for official statistics Beijing, China, 31 Oct 2014, available at http://unstats.un.org

Letoruzé, E., & Jutting, J. (2015) "Official statistics, big data and human development", Data Pop alliance – White paper series, available at www.odi.org

Lucchini, M., & Assi, J. (2012). Mapping patterns of multiple deprivation and well being using self-organizing maps: An application to Swiss household panel data. *Social Indicators Research, 112*(1), 129–149.

Lucchini, M., Butti, C., Assi, J., Spini, D., & Bernardi, L. (2014). Multidimensional deprivation in contemporary Switzerland across social groups and time. *Sociological Research Online, 19*(1), 3.

Scannapieco, M., Virgillito, A., & Zardetto, D. (2013) "Placing big data in official statistics: A big challenge?", 2013 New techniques and technologies for statistics conference, Brussels.

Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics, 26*(5), 1651–1686.

Chapter 4
Developing Indicators and Managing the Complexity

Filomena Maggino

Developing Indicators

A Normative Exercise

Developing indicators is considered an exercise associated with the measurement process.

Actually, in order to start any measurement process, a crucial guiding principle is identified. In particular, three different approaches can be distinguished:

- *fundamental process*: the measurement is *fundamental* when it is not derived from other measurements. Fundamental measures are estimated directly from observations (such as length and volume);
- *derived process*: the measurement is *derived* when it is indirectly derived by means of other measures (like density, velocity);
- *defined process*: in this case the measurement is achieved as a consequence of a definition confirmed through the relationship observed between observations and the concept to be measured. Almost all the measures developed in social sciences belong to this group. The measures developed in this context are defined *indicators*.

That is because measuring in social sciences field requires:

- a robust conceptual definition
- a consistent collection of observations
- a consequent analysis of the relationship between observations and defined concepts.

It is important to start from the meaning of the word "indicator".

Even though the terms *indicator* and *index* are often used in an interchangeable way, actually they have different origin and meaning:

F. Maggino (✉)
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy
e-mail: filomena.maggino@uniroma1.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_4

– *Index* comes from the Latin word "index", which means "anything that is useful to indicate". In statistics, it represents historically a very generic term applied with multiple meanings.
– *Indicators* comes from the late Latin word "indicator" that means "who or what indicates"; in statistics, it represents a more recent term indicating indirect measures of economic or social phenomena not directly measurable. In this perspective, an indicator is not simple crude statistical information but represents a measure organically connected to a conceptual model aimed at describing different aspects of reality.

In other words, a generic index value becomes an "indicator", when its definition and measurement occur in the ambit of a conceptual model and is connected to a defined aim (indicators as *purposeful statistics*; Horn 1993).

Developing indicators is a normative exercise per definition, since indicators are related to a conceptual definition of the phenomenon of interest and since a phenomenon can be defined in different ways. Consequently, in order to describe a phenomenon different group of indicators can be selected.

Between Objectivity and Subjectivity

The normative nature of the selection of indicators makes many scholars assert that the process contains a "dangerous" or "unacceptable" subjective component.

Actually, the subjective component cannot be excluded from the process of selecting indicators but can be conducted in a process involving many "subjectivities" towards a shared definition.

In any scientific field, a meaning does not exist without "subjectivity". This dimension helps in structuring the observed reality. The epistemological research of the last century has focused on the role of the subject in knowledge production clearly showed how pure objectivism cannot account for the knowledge process, even in scientific disciplines. This is particularly evident when observing and analysing socio-economic phenomena.

Consequently, subjectivity represents one of the dimensions inevitably involved in defining and using data in socio-economic statistics. In this perspective, data represent the "text" to be read by the researcher concentrated in looking for a sense. This process cannot be considered arbitrary, since the sense does not autonomously live "outside" and cannot be "invented" regardless of the relationship with the reality. Many times, in the name of "objectivity", technical choices are done in an arbitrary way.

Given the complexity and the nuances of socio-economic issues, data can often be considered as a (fragmented) "text" to be "read" by the researcher, in search for a "sense" and a structure in it. This "sense structuring" process is not an arbitrary one, but necessarily involves some subjectivity.

To make an example, think about the issue of defining poverty thresholds in deprivation studies, both in a monetary and in a multidimensional setting, with the consequences that different choices have in the final picture. Admittedly, in applied studies subjectivity is generally felt as an issue to be removed. Since ordinal data are intrinsically subjective, many evaluation procedures try to "correct" for this. Ironically, removing subjectivity is not an objective process and often produces arbitrary results.

Since subjectivity is unavoidable, the real methodological issue is not removing it; rather, it is building a sound statistical process, where subjective choices are clearly stated and their consequences can be clearly worked out in a formal and unambiguous way.

It is important, at this point, clarifying that term "subjective" changes its value with reference to the context in which it is used:

– are we talking about defining the phenomena?
– are we talking about components of the phenomenon?
– are we talking about defining the method of measurement and analysis?

• *"Subjective" in defining phenomena*. In this case, "subjective" is referred to the description of the reality. The process of describing the reality (conceptual framework) is always *subjective* since it is related to the researchers' view of the reality.

  Concerning this, as Michalos (1992) noticed, the models defined to observe a reality are only apparently neutral. Actually, the conceptual definition represents only a "small window" through which only some facets of the reality can be seen (*reductionism*); in this sense, the view is politically and socially distorted and will condition evaluations, choices, actions, and policies. In this sense, subjectivity expresses the unavoidable working in defining hypothesis helping in describing and understanding the reality. The researcher, through the dialogue with the working hypothesis, can change the perspective in a continuously evolving knowledge path. In any scientific field, a meaning does not exist without "subjectivity". This dimension is present also in the ambit of hard sciences, where it helps in structuring the observed reality. In this perspective, data represent the "text" to be read by the researcher in search of a sense. This process cannot be considered arbitrary, since the sense cannot be "invented" regardless of the relationship with the reality.

• *Subjective" as one of the components of the reality*. In this case the adjective refers to the kind of information which has been defined in the ambit of a conceptual framework and subsequently observed. In order to make the distinction between objective and subjective characteristics operationally clearer, we can refer to the observed unit, possessing this kind of characteristic. Consequently, we can distinguish between:

  – *objective information*, collected by observing reality
  – *subjective information*, collected only from individuals and their assertions.

The distinction can be developed into other resulting definitions:

– *objective indicator*, based upon explicit criteria, shared by external observers
– *subjective indicator*, based upon subjective evaluations and implicit criteria which can vary from individual to individual.

Subjective indicators aim at measuring and quantifying individual components involving different elements – as conscience, cognition, emotion, attitude, and opinion – that are related to contingent and mutable situations. Consequently, measuring subjective aspects needs many elements to be considered and requires an interdisciplinary approach, able to consider and understand the different levels at which each individual react to the submitted question. The different levels involve personality, values, interests, motivations, intellectual and expressive dispositions, memory, experiences, social attitudes as a member of a limited group or of a community, and so on.

• *"Subjective" in the measuring process*. The methodology adopted to observe and study a phenomenon defined in the ambit of the conceptual framework should be objective. In other words, the methodological objectivity concerns the capacity of a procedure to observe without alteration due to external factors and to be free from effects due to the observer; this notion spreads from the procedure of observation to the data analysis to the interpretation of the results.

  Sometimes in this methodological context, the dichotomy "subjective-objective" is considered equivalent to the dichotomy "qualitative-quantitative". However, the two dichotomies should be kept distinct[1]:

– *"objective – subjective"* refers to what we are going to observe
– *"quantitative – qualitative"* refers to the methodological approach applied in order to observe the previous dimensions

The Hierarchical Design

By following the classical Lazarsfeld's model (1958), indicators should be developed, through a *hierarchical design*, requiring the definition of the different subsequent components (Fig. 4.1).

The hierarchical design and its logic can be applied both at micro and macro level.

[1]The transition from the quantity to the quality paradigm implies a consistent choice of the indicators: this means, for example, adopting a "quantity" indicator like "life expectancy" as well as to a quality indicator like "healthy life expectancy."

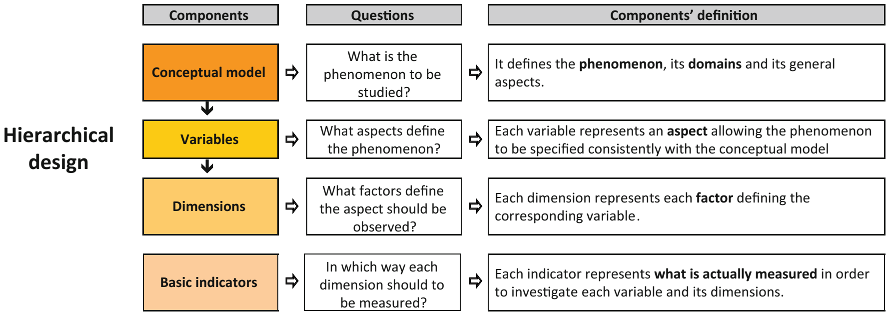


Fig. 4.1 The components of the hierarchical design

The Components of the Hierarchical Design

Conceptual Model

Defining a concept is always represents a process of abstraction, a complex stage that requires the identification and definition of theoretical constructs that involve the researcher's point of view, the applicability of the concepts, the socio-cultural, geographical, historical context. It is a demanding exercise, especially when the concept refers to a complex state, such as quality-of-life.

The process of conceptualisation allows us to identify and define:

(a) the model aimed at data construction,
(b) the spatial and temporal ambit of observation,
(c) the aggregation levels (among indicators and/or among observation units),
(d) the models allowing interpretation and evaluation.

Latent Variables and their Dimensions

Each variable represents an aspect to be observed and confers an explanatory relevance onto the corresponding defined concept. In other words, the identification of the latent variable is founded empirical statements so that the defined variable can reflect the nature of the considered phenomenon consistently with the conceptual model.

The identification of the latent variable is founded on theoretical assumptions (requiring also an analysis of the literacy review) also about its dimensionality. In fact, according to its level of complexity, the variable can be described by one or more factors. The different factors of each variable are referred as dimensions. The concept of "dimensionality" is quite complex, because its meaning is mainly and essentially theoretical. In this perspective, two different situations can be observed:

(a) *uni-dimensional* when the definition of the considered variable assumes a unique, fundamental underlying dimension;

(b) *multidimensional* when the definition of the considered variable assumes two or more underlying factors.

This identification will guide the selection of the indicators. The correspondence between the defined dimensionality and the selected indicators has to be demonstrated empirically by testing the selected model of measurement.

In dealing with concepts like wellbeing or quality of life, the consensus on what variables should be selected and on their interpretation is lower.

What it should be clarified is that comparing different realities (represented by countries or by areas inside one country) does not necessarily imply using the same variables but could require differentiated choices (Stiglitz et al. 2009; Sharpe and Salzman 2004).

In fact, variables' choice depends on shared societal values, which are functions of time and place. Consequently, transferring a wellbeing concept developed in a certain context could be misleading. With this respect, a good example is the variable "leisure time" whose definition can be different also from one individual to another.

Basic Indicators

Only in particular cases, variables can be directly measured (e.g., some simple objective information like age, sex, certificated level of education, and so on). This means that in the majority of the cases the defined variable can be measured only indirectly through observable elements which are called *indicators*[2] of the reference variable. Each basic indicator (item, in subjective measurement) represents what can be actually measured in order to investigate the corresponding variable. In other words, the **indicator** is what relates concepts to reality.

In particular, a statistical index can be considered "indicator" when (Land 1971, 1975):

- it represents a component in a model concerning a social system
- it can be measured and analysed in order to compare the situations of different groups and to observe the direction (positive or negative) of the evolution along time (time series analysis)
- it can be aggregated with other indicators or disaggregated in order to specify the model.

The lack of any logical cohesion should not to be hidden by the use and application of sophisticated procedures and methods that can deform reality through distorted results.

With reference to a latent variable, a variety of different indicators can be identified and we have to accept the idea that maybe no set of indicators exists

[2]In data analysis, indicators/items are technically defined "variables"; consequently, these are conceptually different from "latent variables".

able to perfectly capture a conceptual variable. Moreover, the same defined variable can find different indicators according to different (social, physical, etc.) contexts. For example, different countries could agree on the same definition of wellbeing but then could select different indicators from each other.

From the conceptual point of view, indicators can describe a phenomenon through different level of observations

– Indicators are said *micro* when the values refer to individuals or groups, while they are said *macro* when the values refer to communities, regions, countries, etc.,[3]
– Indicators are defined *internal* or *external*, duality sensitive to individual observation; in fact, the concepts defined at individual level can be observed at both "external" (e.g., objective living conditions) and "internal" (e.g., subjective evaluations or perceptions) level.

How Many Basic Indicators?

According to a simple and weak strategy, each latent variable is defined by a single element (*single indicator approach*). While undoubtedly thrifty and functioning, the single-indicator approach is simple and weak and actually requires the adoption of robust assumptions concerning the possibility of measuring one dimension (with reference to one domain) with just one indicator (i.e., direct correspondence between one latent variable and one indicator).

Such assumption shows some risk since each single indicator can produce a wide and considerable amount of error related to:

(a) *precision* (*reliability*), since the measurement through one single indicator is strongly affected by random error[4];
(b) *accuracy* (*validity*), since the chance that one single indicator can describe one latent complex variable is highly dubious and questionable;
(c) *relationship* with the other variables and dimensions;
(d) *capacity of discriminating* and *differentiating* among observed cases.

Consequently, the adoption of several indicators for each conceptual variable is desirable (especially in case of complex multidimensional latent variables) is desirable. This can be done by adopting the *multi-indicator approach*. This

[3]Macro does not necessarily correspond to summing up micros and the micro level does not necessarily reflect what emerges at the macro level. Quality of life is typically observed at individual level, while other concepts, like economic and social cohesion, are observed at community level. Some concepts require both levels of observation, like sustainability, which can be defined through different dimensions (capitals) and two time perspectives observed at both micro and macro level.

[4]By using multiple measures, random errors tend to compensate each other. Consequently, the measurement turns out to be more accurate. The greater the error component in one single measure, the larger the number of required measures needs to be.

approach allows the problems produced by the single indicator approach to be overcome or, at least, reduced. In fact, multiple measures allow the conceptual dimensions to be measured with more precision (multiple measures allow random errors to be compensated),[5] accuracy and discriminant capacity.

Other Conceptual Issues: The Domains

The relevant concepts, their variables and dimensions have to be assessed and observed within each *domain*. Each domain represents a segment of the reality in which the concepts should be observed and monitored. They refer to individuals, families, territorial/social areas, actually, each individual lives in all these contexts.

Typically they are housing; health, transport, environment, leisure and culture, social security, crime and safety, education, labour market, working condition, and so on.

Actually, a shared list of domains showing explicit priority does not exist, also because the list strictly depends on value judgments, valid and acceptable in a certain place or time (Noll 2004). However, many scholars noticed that many domains recur in empirical studies (Felce and Perry 1995; Nuvolati 1997; Johansson 2002; Stiglitz et al. 2009), highlighting how human conditions lead individuals to face challenges that are common all over the world and that require collective solutions. Generally, the differences concern the importance assigned to each domain.

Defining the Model of Measurement: Reflective and Formative Approaches

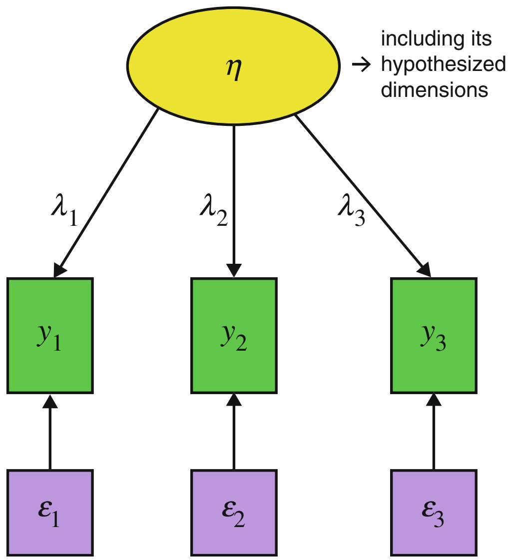
The model of measurement refers to the relationship between variables (and their dimensionality) and indicators. It can be conceived through two different conceptual approaches (Blalock 1964; Diamantopoulos and Siguaw 2006)' reflective and formative.

Models with Reflective Indicators

In this case, the indicators are assumed to be *reflective* in nature, which is function of the latent variable. This implies that any changes in the latent variable are reflected (i.e., manifested) in changes in the observed indicators.

[5]In particular, the basic indicators defined in the multi-indicator approach are considered *multiple measures*, since they are assumed to cover the conceptual dimension's variability.

Fig. 4.2 The model of measurement: the reflective approach



The reflective approach refers to the *top-down* explanatory approach and assumes (Diamantopoulos and Winklhofer 2001) that (Fig. 4.2):

– indicators are interchangeable (the removal of one of the indicators does not change the essential nature of the underlying construct),
– correlations between indicators are explained by the model of measurement,
– internal consistency is of fundamental importance: two uncorrelated indicators cannot measure the same construct,
– each indicator has an error component, in other words, the indicator's variance is explained also by the error term (ε).[6]

[6]Generally, the formal representation of models of measurement uses symbols referring to the Greek alphabet:

Greek alphabet

Capital	Small		Capital	Small		Capital	Small		Capital	Small	
Α	α	alfa	Η	η	eta	Ν	ν	nu	Τ	τ	tau
Β	β	beta	Θ	θ	theta	Ξ	ξ	csi	Υ	υ	upsilon
Γ	γ	gamma	Ι	ι	iota	Ο	o	omicron	Φ	ϕ	phi
Δ	δ	delta	Κ	κ	kappa	Π	π	pi	Χ	χ	chi
Ε	ε	epsilon	Λ	λ	lambda	Ρ	ρ	rho	Ψ	ψ	psi
Ζ	ζ	zeta	Μ	μ	mu	Σ	σ	sigma	Ω	ω	omega

The reflective formal specification implies that

$$y_i = \lambda_i^\eta + \varepsilon_i$$

where

η latent variable
$y_1, y_2, \ldots, y_n$ set of observable indicators
λ_i expected effect of η on y_i
ε_i measurement error for the i-th indicator ($i = 1, 2, \ldots, n$).

For $i \neq j$, it is assumed that.

$$COV(\eta, \varepsilon_i) = 0 \quad COV(\varepsilon_i, \varepsilon_j) = 0$$

$$E(\varepsilon_i) = 0$$

Reflective measures are assumed to be sampled from a population of measures describing the latent construct. The relationships between latent variables can be inferred only when the significant relationships between indicators and the corresponding latent variables are observed. According to this, the reliability and validity assessment can be accomplished through a statistical approach consistent with the traditional specification used in *factor models*, where an observed measure is presumed to be determined by a latent factor and a unique factor.

Models with Formative Indicators

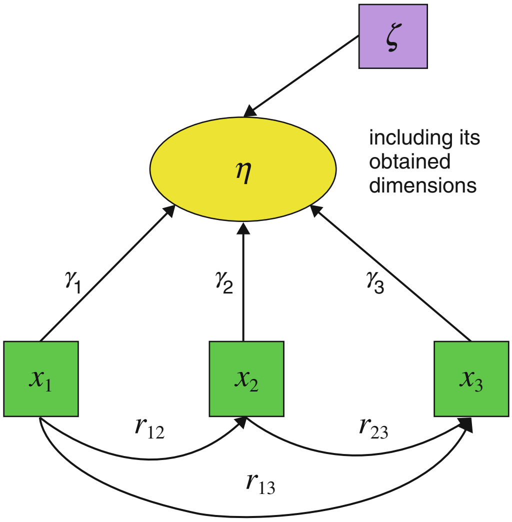
In this case, the indicators are assumed to be *formative* (or causal) in nature since they are viewed as causing – rather than being caused by – the latent variable. This means that changes in formative indicators (Blalock 1964) determine changes in the value (and meaning) of the latent variable. The formative approach refers to the *bottom-up* explanatory approach. According to this approach, a concept is assumed to be defined by (to be a function of) a group of indicators, identified in order to define it.

The main specific properties of the formative indicators (Diamantopoulos and Winklhofer 2001) can be synthesized as follows (Fig. 4.3):

– indicators are not interchangeable (omitting an indicator is omitting part of the construct),
– [eventual] correlations between indicators are not explained by the measurement model,
– internal consistency is of minimal importance: two uncorrelated indicators can both serve as meaningful indicators of the same construct,
– indicators do not have error terms; error variance is represented only as disturbance terms (ζ).

An example of this situation is the following. We could think to measure the individual socio-economic status (SES). Its definition depends on which indicators

Fig. 4.3 The model of measurement: the formative approach



are selected, by referring to different dimensions like education, income, and occupational prestige. So, the selected indicators cause (form) the latent variable SES. If an individual loses his or her job, the SES would be negatively affected. However, a change in an indicator (say, income) does not necessarily imply a similar directional change in the other indicators (say, education or occupational prestige).

In some cases, the latent variable is defined by a linear combination of indicators. In this case, the formal specification is the following:

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \ldots \gamma_n x_n + \zeta$$

where

η latent variable
x_i indicator i
γ_i the expected effect of x_i on η
ζ disturbance term with $\mathrm{COV}(x_i, \zeta) = 0$ and $E(\zeta) = 0$

The correct distinction between formative and reflective indicators is related to the correct definition of the latent variable and conceptual model and allows not only to correctly interpret the relationships between indicators but also to correctly identify the procedure aimed at aggregating basic indicators.

In choosing the model of measurement, four different situations can occur (Diamantopoulos and Siguaw 2006) as represented in Fig. 4.4.

The adoption of a wrong perspective produces two types of error:

- Type I occurs when a reflective approach has been adopted although a formative approach would have been theoretically appropriate for the construct;

Fig. 4.4 The model of measurement: possible decisions

		'Correct' theory	
		reflective	formative
Chosen perspective	reflective	*correct decision*	**Error type I**
	formative	**Error type II**	*correct decision*

- Type II occurs when a formative approach has been adopted even if the nature of the construct requires a reflective operationalisation (identification problems).

As we will see, the correct interpretation of the relationships between indicators and latent variable allows the procedure aimed at aggregating basic indicators to be correctly identified.[7]

From Basic Indicators to Systems of Indicators

The proper and accurate application of the hierarchical design leads to the consistent definition of a complex structure which does not represent a pure and simple collection of indicators. In this complex structure, each indicator measures and represents a distinct constituent of the defined phenomenon.

By considering concepts, variables, dimensions and domains, a *conceptual matrix* can be conceptualized (Fig. 4.5), in which each combination (cell) can be covered by indicators (actually, not each cell will be necessarily covered by indicators).

For its characteristics, such complex structure can be defined as a *system*. A group of indicators without functions and interconnections represents what technically is called *set of indicators*.

Characteristics of a System of Indicators

Principles Defining a System

In order to understand if we are dealing with a system, we can check the following list (Meadows 2008):

- Can we identify parts? and
- Do the parts affect each other? and

[7]In case of reflective indicators, the synthesis assessment (in terms of reliability and validity) can be accomplished through a statistical approach related to the *factor models* (scaling models), where an observed measure is presumed to be determined by a latent factor and a unique factor.

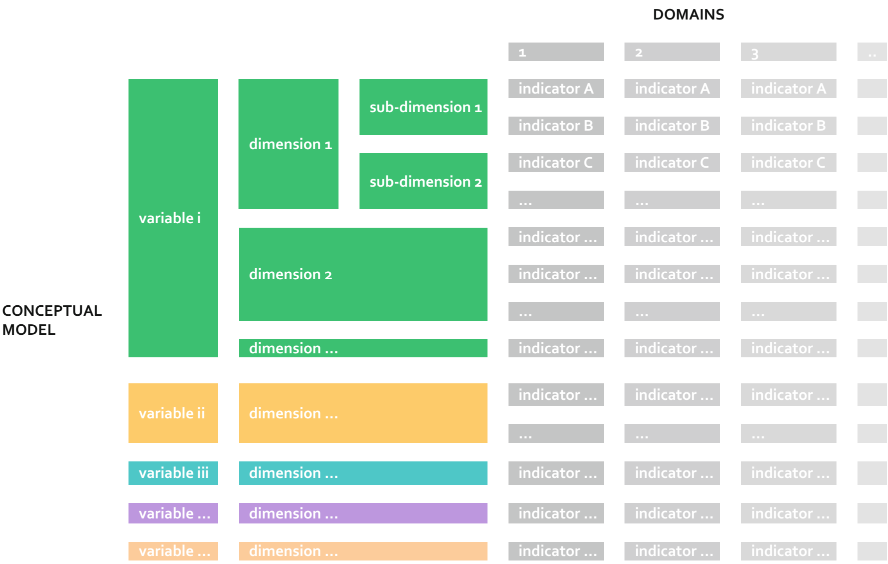


Fig. 4.5 The conceptual matrix, concepts, domains and indicators

- Do the parts together produce an effect that is different from the effect of each part on its own? and maybe
- Does the effect, the behavior over time persist in a variety of circumstances?

A system is not a simple collection of things (indicators, in our case). A system is an interconnected set of elements, organized consistently with a perspective. In other words, a system is defined by three **components**: *functions*, *elements*, and *interconnections*.

What characterizes a system is *integrity* and an active set of mechanism to maintain that integrity. A system is more than the sums of parts. "It may exhibit adaptive, dynamic, goal-seeking, self-preserving, and sometimes evolutionary behaviour". (Meadows 2008).

Systems can be embedded in systems, which are embedded in other systems and so on. Systems can be nested within systems.

Generally, the elements composing a system (the indicators, in our case) represent the easiest part to identify. Individually, they are deeply specialized. By identifying indicators even more and more specialized, the whole picture could be missed. Changes in the elements are not very effective as long as functions and purposes remain the same. Their interpretation may change if the interconnections change.

What helps in preserving the integrity of the system is looking at the interconnections, i.e. the relationships that hold the elements together. Some interconnections are physical; represent flows of signals or information (about resources or incentives); and so on. Any change in the interconnections introduces crucial and critical changes in the system especially in the interpretation of the elements.

Generally, functions and purposes are not necessarily expressed and declared explicitly by a system and are deducted by the operation of the system. Actually, they represent the most crucial determinant of the system in itself. The complexity of a system increases if we think that it is possible to identify purposes within purposes.

In a system, it is difficult (and also inconsistent) trying to establish which component is the most important. However, functions and purposes represent the most crucial determinant of the system and its functioning.

In a system of indicators monitoring, for example, the wellbeing of a nation, functions and purposes are pre-defined. Since purposes can refer to different actors of a society (individuals, organizations and so on), the definition of a system of indicators should involve different stakeholders of the society to be monitored.

Functions of a System of Indicators

Generally, *function* refers to a nonhuman system while *purpose* refers to a human one. Actually, if we refer to a system of indicators monitoring, for example, the wellbeing of a country the distinction is not absolute. However, both are related to the aim, the general motivation to have the system of indicators. Actually, a system of indicators may represent an important and valid support to

– forming and developing particular sensitiveness,
– informing and stimulating public debates,
– scientific or operative goals,
– supporting, guiding, and directing decisions and possible interventions (policies),
– evaluating impacts of different policies.[8]

[8]An example is represented by decision makers who need to know and manage a composite mosaic of information in order to define and evaluate priorities to be translated into actions. With reference to this, we can distinguish between:

– *Conceptual aims* (**goals**). In other words, goals refer to the direction to be adopted by the society. They are defined through a consensual process, by referring to cultural paradigms or normative demands, or through expert groups' pressure or public opinion movement In the *wellbeing* perspective, they are not only time and space dependent but rely on political views. Consequently, it is important to set clear and shared goals, by giving philosophical and political debate (understandable for all) more democratic space. Goals require action/intervention proposals to be defined by taking into account that the taken decisions will influence all the domains, even when no resolution is made on each of them.
– *Operative aims* (**objectives**) that represent the instruments identified in order to attain the conceptual aims. Objectives can have different temporal prospects (monthly, four-monthly, annual, bi-annual, etc.)

Consequently, the motivations urging the development of systems of indicators go from the simple description of a phenomenon to its evaluation. In particular, a system of indicators can show different functions, which can be seen in cumulative terms since each of them requires the previous one/s.

Moreover, consistently with functions and purposes of a system and its structure, different analytic approaches aimed at analysing indicators within a system can be distinguished. In general terms, the analysis within a system should (Michalos 1992):

– allow future trends to be forecasted,
– show and point out particular problems,
– help in defining priorities of policies,
– allow territorial comparisons,
– suggest new domains that need to be studied in order to define new theories and a deep knowledge of social structures and functions.

Figure 4.6 summaries different functions of a system of indicators.

Independently from the function, a system of indicators should be built up by respecting some

- formal criteria (comprehensiveness, consistency, non-redundancy, parsimoniousness)
- basic characteristics (Noll 2004):
  – *objectivity*: provided information should turn out to be equal or comparable, independently from who are the users;
  – *quantification*: provided values should be quantitative – obtained through standardized procedures and measures; this allows results to be reported with more precision and detail, and data to be analysed through complex methods;
  – *efficiency and fidelity*: methods, techniques and instruments that allowed data and results to be obtained have to be communicated and publicized,
  – *economicity*: the system has to produce simple, standardized, available and up-to-datable information;
  – *joint development*: the system has to be developed in a shared way by all the "actors".

– *Planning aims* (**actions**) that represent the specific activities identified to accomplish the objective. They can include developments and infrastructural changes in policies, institutions, in management instruments, and so on.

Each goal, objective and action has corresponding **targets**, representing those elements allowing each goal, objective and action to find measurable criteria and to define a *timetable*, and corresponding **indicators** defined in order to assess progress towards the target with goals and objectives and the accomplishment of actions.

Descriptive and explanatory functions	**Monitoring**[a]	This basic function concerns and refers to the capacity of the system to clearly define the phenomenon by describing its conditions and dynamics with reference to a certain reality (a country, an institution, etc.). When the indicators are observed over time, they allow changes (economic, social, etc.) to be detected. Consequently, through the monitoring functions it is possible to identify eventual new problems and critical points issues.
	Reporting	Social reporting represents the most important function of indicators systems (Noll, 1996). It can be defined as "the presentation of data which enable the evaluation of living conditions of the population and their change over time." A consistent and continuous monitoring allows the analysis (and the interpretation) of the existing relationships between different components of the system. Then the reporting is obtained by adding to the monitoring function the possibility to analyse and interpret the phenomenon (Noll, 1996; Berger-Schmitt & Noll, 2000).
Evaluation functions	**Forecasting**	The systematic reporting (*monitoring* + *analysis* + *interpretation*) allows trends and effects attributable to change to be observed. Understanding change allows hypothesis on future trends to be expressed. This function, representing a natural consequence of the reporting function, allows procedures unable to allocate resources and plan efficiency procedures *ex-ante*. (Cannavò, 2010)
	Accounting	The systematic availability of data, consistent with a conceptual framework, supporting decision concerning allocation and destination of resources represents a useful mean of *accounting*.[b] In particular, this function allows a system to (Cannavò, 2010) - control ex post the suitability of the defined standards and of the planned resources flows, - evaluate efficiency and correctness of the defined procedures, - test adequacy and actual attainment of results.
	Program management and performance evaluation	Systems of indicators represent valid supports to specific strategic programmes which can be evaluated with reference to their present realization (capacity to meet particular and specific purposes) and their possibility to suggest future actions. In particular, indicators should allow the following assessments: - evaluating present state (*where are we now?*) - identifying priorities and actions to be pursued (*where do we want to go?*) - evaluating adequacy (*are we taking the right path to get there?*) - evaluating progress towards goals by quantifying the strategic performances (*are we there yet? can differences be observed?*). Since these systems are constructed with reference to specific programmes, they can be hardly generalized. In this perspective, this function can play an important role in policy analysis (policy guidance and directed social change) by allowing problem definition, policy choice and evaluation of alternatives, and program monitoring (Land, 2000).
	Assessment	A system of indicators can represent a valid support to assessment procedures (certification and accountability). In this case, the goal is to certificate or judge subjects (individuals or institutions) by discriminating their performances or to infer functioning of institutions, enterprises, or systems.

Fig. 4.6 Functions of systems of indicators

[a]The word "monitor" comes from Latin *monitor –oris*, from the verb *monere*, which means *to warn, to inform, to advice*. Originally, the term was applied in industrial context in order to keep under continuous surveillance a machine, a process, or an operative structure. In the meantime, the concept of warning a context through meaningful information (data) has been spread also in different fields. Monitoring requires programming methods allowing the context to be observed and compared. Monitoring represents an interacting cycle with a reciprocal influence. The analysis can be different according to the

– context/process (critical or not critical)

– characteristics of the monitoring in terms of cycle (continuous, at high frequency or low frequency) or level of measurement (simple, complex or very complex)

			critical context / process	not critical context / process
Continuous cycle	measurement apparatus and tools	simple	earthquakes (seismographs)	earthquakes (seismographs)
		complex	intensive care unit, moving systems (airplanes)	telecommunication systems informatics networks
		very complex	financial market	astronomical monitoring
High-frequency cycle	measurement apparatus and tools	simple	air pollution (accidents)	air pollution (norms controls)
		complex	accounting	
		very complex		
Low/ mid-frequency cycle	measurement apparatus and tools	simple		opinion polls
		complex		social surveys
		very complex		social surveys

Different frequencies require different data transformation approaches allowing comparisons (normalization, interpolation, and so on).

Architecture: Structure of Observation and Analysis

As we have seen, monitoring represents the fundamental functions. Generally, this refers to the availability of indicators allowing observation of the conditions of a particular reality (e.g., a country) over time. In this perspective, each selected indicator should be grounded on data collected over time, across different geographical areas or/and groups. These require the definition of the architecture allowing the system to carry on the identified functions. The architecture involves different aspects, from the identification of the units to monitor to the procedures of measurement. (Noll 1996; Berger-Schmitt and Noll 2000).

One of the most important aspects defining the architecture concerns the different views of the monitoring:

– *time view*: the same phenomenon on the same reality is observed over time (years, months, etc.). This view requires an organization in term of cadence (rate) and continuity through which indicators are collected and updated; indicators will not necessarily have the same rate but will be updated with reference to the permanence of the measured phenomenon; moreover, the dynamics should consider also the optimal duration to understand resource diminution as well as impacts of changes.
– *territorial view*: the same phenomenon in the same moment is observed and analyzed across different territorial areas (regions, provinces, etc.). the main issue is identifying the frame adequate to the phenomenon (the best area aimed at understanding air pollution may require a different area size from the one aimed at understanding crime); in other situation, the size is established at administrative level (e.g., regions or countries) on which the decisional system (policy) is sized[9];
– *group view*: the same phenomenon in the same moment is observed in different groups (e.g., different generations, income levels). This view requires an organization in terms of sample of observed individuals.

[9]It should be taken into account that observing a wide territory does not entail that a lower level is necessarily covered. Beyond statistical representativeness, the conceptual model (in terms of dimensions and/or indicators) and the observation approach need to be reviewed and adapted in order to monitor the lower level (e.g., province, city, etc.). Seen in this perspective, the approach aimed at reaching small area estimations from representative data collected in wider areas appears questionable. Projects calibrated on smallest areas should be urged and encouraged.

Fig. 4.6 (continued) Any analytic approach aims at transforming data in order to be communicated. The relevant social impacts and political implications that can be derived rely on bodies' and organizations' ethics and fairness in managing monitoring systems (e.g., official statistics).
[b]According to OECD definition (http://stats.oecd.org/glossary/index.htm), the set of accounting procedures, internal mechanisms of control, books of account, and plan and chart of accounts that are used for administering, recording, and reporting on financial transactions is defined "accounting system". The transactions accounts include a balancing item which is used to equate the two sides of the accounts (e.g. resources and uses) and which is a meaningful measure of economic performance in itself.

Of course, the views can be combined; the combination can help in understanding the relationship between the concepts and the different components and how domains can be related to policy actions.

The indicators should show a high level of *comparability* according to the impact of differences in applied concept and measurement tools/procedures. Comparability is related also to necessity of harmonizing different data sources and levels of observation.

A particular attention should be paid in comparing different time points. *Trend-indicators* can produce *time series* that need to be carefully managed since the observed moments could reveal themselves incomparable and/or the defined indicators could reveal themselves as non-applicable after some time.[10]

Comparing different geographical areas represents also a delicate task, especially when areas consist of regions or countries. In this perspective, the comparative issue gets more complicated. In fact, the indicators should be characterized by their *consistency* with reference to concepts, and their *adequacy* with reference to the territory (country, region, province, etc.).

In order to systematize the structure of the system, a design through which data are collected has to be defined. The design can be:

– *Vertical*: data are collected from local levels (e.g. regions) in order to be systematized (aggregated) at a higher level (e.g. country). This structure allows policy goals to implemented, according to local information.
– *Horizontal*: data are collected only at one level (e.g. regional) and allow particular observational ambits (environment, education) to be monitored and compared.
– *Local*: this structure is typically designed in order to support local decisional processes. This kind of system is characterized by two levels, internal, when the indicators are aimed at monitoring the internal organization of the level, and external, when the indicators refer to parameters existing at higher levels (e.g. transportation).

Characteristics of Indicators Within a System

Indicators developed through the hierarchical process are seen in relation to each other and show a meaningful and precise position in the system consistently with the conceptual model. When seen in the ambit of a system, indicators may show different characteristics related to (i) the *perspective* through which the indicators are reporting the phenomenon to be observed, (ii) the *communication context* in

[10]Generally, trend analysis is used to observe changes and is aimed at estimating future events. Data analysis approaches can be mainly distinguished according to the adopted design; particularly, each design allows analysis at different level (Maggino and Facioni 2015).

which the indicators are used, (iii) the *interpretation* attributed to the indicators in statistical analyses, (iv) the *criteria* of their adoption, and (v) their quality.

(i) Perspectives of Observation

With reference to the perspective, an indicator can describe:

– *a status*, when it measures the phenomenon in a particular moment; it allows for cross-comparisons between different realities or *a trend*, when it measures the phenomenon in its time dynamics; it requires a defined longitudinal observational design (for example, repeated surveys on particular populations)[11];
– an *objective* or *subjective* fact (e.g., relatively subjective appraisals of housing and neighbourhoods by actual dwellers may be very different from relatively objective appraisals by "experts")[12];
– a *positive* or a *negative* aspect (e.g., in the health domain, measures of morbidity and mortality from one side and life expectancy on the other side);
– a *conglomerative* (e.g., it measures overall well-being, where increases in well-being of the best-off can offset decreases in well-being of the worst-off) or a *deprivational* perspective (e.g., it measures only the welfare of the worst-off)[13];
– a *benefit* or a *cost* (different measures of value or worth yield different overall evaluations as well as different evaluations for different people);
– an *input* (resources) or an *outcome* (results) feature (e.g., income can represent an input for a family's material wellbeing while other situation can represent results)[14]; these indicators can be also combined in order to define efficacy/efficiency indicators;
– an *impact* aspect (especially when indicators are used in evaluation term).

[11]Time of observation should not be necessarily equal for all selected indicators according with their different dynamics; in fact, some phenomena show "fast" dynamics while others show extended changing progression.

[12]Another way to look at the dichotomy objective-subjective is adopting the duality sensitive to individual observation, i.e. internal level and external level; in fact, at individual level the defined concepts should be observed at both "external" (e.g., objective living conditions, equity and sustainability of those conditions) and "internal" (e.g., subjective evaluations about the living conditions, subjective perceptions about equity and sustainability of living conditions) level.

[13]Anand and Sen (1997), arguing that the conglomerative and deprivational perspectives are not substitutes for each other, proposed a *complementary* approach (Sharpe and Salzman 2004).

[14]Classifying indicators in terms of input and outcomes aspects is difficult; in fact, some aspects could be classified at the same time (or in subsequent times) as input or output; families' lower expenses for foodstuffs could represent an output indicator related to a short-term situation but could also represent an input indicator towards a change (worsening?) in family members' health status.

(ii) Communication Context

With reference to the communication context, indicators can be classified in:

- *cold indicators*: in this case, the indicators have a high level of scientific quality and show a high level of complexity and difficulty;
- *hot indicators*: in this case, the indicators are constructed at a low level of difficulty and show a high level of understanding. It is unusual for these indicators to be used in a policy context;
- *warm indicators*: in this case, the indicators show a good balance between quality, comprehensibility, and resonance.

(iii) Interpretation

With reference to the interpretation of indicators in an analytical context, an indicator can be analyzed in

- *descriptive* terms, when an indicator is just informative of a situation,
- *evaluating/explicative* terms, when an indicator is aimed at interpreting the results of another one (e.g., a descriptive indicator);
- *predictive* terms, when an indicator helps to delineate plausible evolutional trends (development or decrement).

In order to interpret the results, the following issues should be privileged:

- Distributions: e.g., because average figures can conceal extraordinary and perhaps unacceptable variation, choices must be made about appropriate representations of distributions.
- Distance impacts: e.g., people living in one place may access facilities (hospitals, schools, theatres, museums, libraries) in many other places at varying distances from their place of residence.
- Causal relations. Before intervention, one must know what causes what, which requires relatively mainstream scientific research and, which may not yet be available.

(iv) Criteria

Criteria refer to the interpretation of the results according to a specific reference frame. This can also include particular *standard-values*, which can be defined a priori, according to the **objectives**. The reference frame allowing the present situation to be interpreted and consequently evaluated and assessed are called *benchmarks*. Generally speaking, benchmarking is the process of comparing some characteristics (cost, productivity, performance, quality of a specific process or method). Essentially, benchmarking helps in understanding where each case is in relation to a particular standard. The result often stimulates each case in making changes in order to obtain improvements.

The systematic use of benchmarks may be useful for monitoring but also for encouraging actions and improvements since it[15] allows priorities to be established, better practices to be defined, impacts to be evaluated, and awareness among the stakeholders to be aroused.

Benchmarks, interpreted in terms of reference points, can assume different shapes (Śleszyński 2012):

– reference point (or critical value),[16] representing quantitative information established thanks to scientific research or to desired norms;
– signpost arrow, indicating the comparison with reference to previous performance (betterment or worsening);
– best practice, representing a model to be followed;

In certain cases, along with general standards, differential standards can be defined with reference to different groups (e.g. for males and females). Comparisons among groups are possible according to the availability of a unique scale for the observed and standard values.

Identifying the benchmarks is strictly related to the definition of goals, which represent broad statements concerning what has to be achieved or which is the problem to be faced.

According to the different perspective adopted, different benchmarking analysis can be identified, such as

– *Process benchmarking*: the goal is identifying and observing the best practices from one or more benchmark cases.
– *Performance benchmarking*: it allows each case to assess its position by comparing results (products, services, and so on) with those of target case.
– *Output benchmarking*: it allows strengths and weaknesses to be found.
– *Strategic benchmarking*: it involves observing how others plan the activities.
– *Functional benchmarking*: it focuses on a single function in order to improve the operation of that particular function. In case of more complex functions, their disaggregation into processes allows valid comparisons to be performed.

From the analytic point of view, the main approaches allowing comparisons to be accomplished are Data Envelopment Analysis (DEA) and regression analysis.[17]

This kind of analysis is considered a strategic process, through which various aspects of the monitored system can be evaluated. This allows plans to be

[15]The use of benchmarks plays an important role in the ambit of a program development. Used in combination with the program objectives they provide the basis for program accountability.

[16]A reference point can be actually represented by a reference group (e.g., percentage of people with a high level of satisfaction with life as a whole).

[17]With regression analysis cases that performed better than average can be rewarded while cases that performed worse than average can be penalized. Such benchmarking studies are used to create yardstick comparisons, allowing outsiders to evaluate the performance of operators in an industry. A variety of advanced statistical techniques, including stochastic frontier analysis, have been utilized to identify high performers and weak performers.

AN INDICATOR SHOULD BE	clear, meaningful, consistent	in describing the conceptual models and in relating to the defined aims and objectives	ACCURACY AND VALIDITY	**METHODOLOGICAL SOUNDNESS**
	appropriate, exhaustive, pertinent	in meeting requirements underlying its construction (knowing, monitoring, evaluating, accounting, …)		
	repeatable, robust	in measuring the underlying concept with a degree of distortion as low as possible	PRECISION	
	reproducible, stable		RELIABILITY	
	transparent, ethically correct	in data collection and dissemination	OBJECTIVITY	**INTEGRITY**
	relevant, credible	in meeting users' needs	APPROPRIATENESS	**SERVICEABILITY**
	practicable, up-to-datable, thrifty	in observing through realistic efforts and costs in terms of development and data collection	PARSIMONY	
	well-timed, timely, punctual	In reporting the results with a short length of time between observation and communication	AVAILABILITY	
	periodic, regular	In observing the phenomenon over time (for example, short time between observation and data availability)		
	discriminant, disagregable,	in recording differences and disparities between units, groups, geographical areas and so on	COMPARABILITY	
	accessible, interpretable, comprehensible, simple, manageable	in being findable, accessible, useable, analyzable, and interpretable	USABILITY	**ACCESSIBILITY**

Fig. 4.7 Dimensions of quality of indicators

developed aimed at making improvements or adopting best practice, usually with the aim of increasing some aspects of performance.

However, in studying well-being of societies, the benchmark process needs to be considered with great attention and care and needs a strong a shared conceptual framework.

(v) Quality

Different issues need to be addressed in order to select and manage indicators, especially when indicators relate to a complex system and allow the accomplishment of functions like monitoring, reporting and accounting. Michalos (Sirgy et al. 2006) identified 15 different issues (not mutually exclusive and not exhaustive) related to the selection of social, economic, and environmental indicators. Choices and options selected for each issue have implications for the other issues.

One of the most claimed characteristics is the *quality of indicators*. Many international institutions, like the World Bank & Unesco (Patel et al. 2003) and Eurostat (2000a, b) have tried to identify the attributes of *quality* that indicators (and approaches aimed at their management) should possess and need to be considered in the process of developing new indicators or in selecting available indicators.

Synthetically, the dimensions of quality are methodological soundness, integrity, serviceability, and accessibility (Fig. 4.7).

Although they do not represent a dimension of quality in itself, prerequisites of quality refers to all those (institutional or not) preconditions and background conditions allowing quality of statistics.

These prerequisites cover the following elements:

(i) Legal and institutional environment, allowing conceptual framework to be defined, coordination power within and across different institutions to be framed, and data and resources to be available for statistical work
(ii) Quality awareness informing statistical work.

In other words, indicator construction is not simply a technical problem but should become part of a larger debate concerning how to construct indicators obtaining a larger legitimacy to be promoted.

When indicators are considered in terms of system, the social acceptability becomes an important issue related to the legitimacy of the body managing the system. Further dimensions of quality are completeness, embedding responsibility, verifiable through external bodies and organizations, continuously improved (improvable). This leads to the concept of *accountability*.[18]

[18]*Accountability* is a concept with several meanings. It is often used synonymously with such concepts as responsibility, answerability, enforcement, liability. In governance perspective, it has been central to discussions related to problems in both the public and private worlds.

- Political accountability. Political accountability is the accountability of the government, civil servants, and politicians to the public and to legislative bodies (congress, parliament). Generally, voters do not have any direct way of holding elected representatives to account during the term for which they have been elected. Moreover, some officials and legislators may be appointed rather than elected. Constitution, or statute, can empower a legislative body to hold their own members, the government, and government bodies to account. This can be through holding an internal or independent inquiry. The powers, procedures, and sanctions vary from country to country. The legislature may have the power to impeach the individual, remove them, or suspend them from office for a period. The accused person might also decide to resign before trial. In parliamentary systems, the government relies on the support or parliament, which gives parliament power to hold the government to account. For example, some parliaments can motion for a vote of no confidence in the government.
- Administrative accountability. It refers to internal rules and norms as well as some independent commission are mechanisms to hold civil servant within the administration of government accountable. Within department or ministry, firstly, behaviour is bounded by rules and regulations; secondly, civil servants are subordinates in a hierarchy and accountable to superiors. Apart from internal checks, some supervisory bodies accept complaints from citizens, bridging government and society to hold civil servants accountable to citizens, but not merely governmental departments.
- Market accountability. Nowadays, it is "customer-driven" and is aimed at providing convenience and various choices to citizens; ideally, this perspective should improve quality of service. The standard of assessment for accountability requires a neutral body. Government can choose among a shortlist of companies for outsourced service; within the contracting period, government can hold the company by rewriting contracts or by choosing another company.
- Constituency relations. A particular agency or the government is accountable if voices from agencies, groups, or institutions, which is outside the public sector and representing citizens' interests in a particular constituency or field, are heard. Moreover, the government is obliged to empower members of agencies with political rights to run for elections and be elected; or, appoint them into the public sector as a way to hold the government representative and ensure voices from all constituencies are included in policy-making process.

Managing Indicators: Instruction for Use

Managing indicators needs their quality to be considered and introduces at the same time a **challenge** (given by the *complexity*), a **need** (represented by the *relativization*) and a **risk** (given by the over-*reductionism*).

The key allowing for the proper identification of new measures lies in the players' (statisticians, researchers, analysts, policy makers, and so on) capacity and awareness in considering the complexity, avoiding over-reductionism and investigating relativization.

A Challenge: Complexity

Defining and selecting indicators imply several methodological issues, in order to manage the complexity referring to several issues.

1. **Multidimensionality**, which requires different aspects to be defined, not necessarily consistent to each other.
2. **Nature**, which can be
   - *objective and subjective*, which refers to two aspects of the reality integrating each other
   - *quantitative and qualitative*, which implies afterwards a consistent choice of the measures[19]
3. **Levels of observation**, which can be
   - *micro* (individuals, groups)
   - *macro* (communities, regions, countries, etc.)

   where macro does not correspond necessarily to sum of micros and micro does not necessarily reflects what emerges at macro level. *Quality of life* is typically observed at individual level, while other concepts, like *economic and social cohesion*, are observed at community level. Some concepts require both levels of observation, like *sustainability*, which can be defined through different dimensions (capitals) and two time perspectives observed at both micro and macro level:
4. **Dynamics**, referring to:
   - *Internal levels and external conditions*, in fact, at individual level the defined concepts can be observed at both "external" (e.g., objective living conditions, equity and sustainability of those conditions) and "internal" (e.g., subjective

[19]For example, "life expectancy" represents a quantitative aspect, while "healthy life expectancy" represents a more qualitative aspect.

evaluations about the living conditions, subjective perceptions about equity and sustainability of living conditions) level.

- *Trends*, which can be different from one dimension to another (linear, non-linear, chaotic, and so on).
- *Relationships between phenomena*.[20]

A Need: Making Relative

Making relative represents a crucial need in dealing with indicators and concerns issues which are at the same time conceptual and technical.

From the conceptual point of view, making indicators relative is related to their

- consistency with the defined concept
- adequacy with reference to involved territory /groups

The construction of wellbeing indicators represents a good example of the need of making a concept relative. The definition of wellbeing finds a wide agreement (combination of living conditions and subjective well-being). Its operationalization (in terms of indicators) aims at adapting such definition to the territorial domain in which the observation is made. Consequently, different areas could identify different domains and adopt different indicators in order to measure the same concept. Indicators will be compared with reference to the conceptual synthesis and not with reference to single indicators (comparing synthetic indicators).

Making relative is directly related also to the interpretation of the results described by the indicators (also in the perspective of urging better policies). This can be illustrated with another simple and simplified example: for a particular region, a high value was produced by the ratio *number of hospital beds/dimension of population*. How to interpret it? A flat interpretation could be that -----e following: a high level could reveal a region paying attention to the population's needs in terms of health. However, a question could be raised: does the high number of available beds in hospital fit a real need of that population of that territory? If so, maybe that need could be raised by particular pathologies strongly related to difficult environmental conditions.

Making relative has strong implications with reference to *comparability* of indicators. Actually, making relative and making comparable are strictly related and involve the general need of harmonizing different concepts, data sources and analyses with reference to different levels of observation (Fig. 4.8).

[20]Some aspects could be classified at the same time (or in subsequent times) as input or output; families' lower expenses for foodstuffs could represent an output indicator related to a short-term situation but could represent also an input indicator towards a change (worsening?) in family members' health status.

	over time	across territories / areas	between groups
concepts			
data			
analysis			

Fig. 4.8 Dimensions of indicators comparability

Comparing concepts, data and analyses over time, across territories and between groups should consider differences in geographical, linguistic, social, cultural, political, environmental, and administrative conditions as well as methodological conditions for data production (sampling design, questionnaire design, data collection methods, and so on).

All these issues have direct and strong implications in the statistical treatment of indicators (checking measurement invariance and equivalence). From the technical point of view, data comparability is (more or less) solved in the ambit of data normalization. The procedures of data standardization aim at giving a correct and comparable meaning to each indicator also in view of synthetic indicators creation. Selecting the more appropriate approach to standardization is not easy. Several issues have to be considered:

– data properties
– original meaning of the indicators
– values to be emphasized or penalized
– whether or not absolute value are used
– whether or not cases are compared to each other or to a reference unit
– whether or not units are evaluated across time

A simple example concerns the interpretation of the frequency distribution of one indicator. An indicator may show variations that cannot be interpreted in the same way along the measurement scale. This issue remains crucial also after normalization when the standardized indicators have to be aggregated (same increment/decrement of two indicators on the same standardized scale may not be interpreted in the same way).

A Risk: Reductionism

As we will see, the systematic identification of elementary indicators, in terms of concepts and domains, allows a downright *system of indicators* to be constructed (more complex than a simple *set of indicators*, which are not always related to a conceptual framework). In other words, the consistent application of the hierarchical design produces a complex structure.

In order to obtain a meaningful and interpretable picture, data should be managed in some way. In other words, the system of indicators may require the indicators to be reduced in order to allow more comprehensive measures. This issue is referred as *reductionism*. Reductionism cannot be avoided, since it is actually impossible pull an image and a story from a pure observation of the reality and completely grounded on it. On the other hand, it is dangerous to concentrate on just a few elements and statistically infer the sufficiency of the reduced observation from them.

Reductionism applied to indicators can find essentially two solutions: (i) reducing the number of indicators, (ii) synthesising indicators.

The former approach needs a solid conceptual support. From the statistical point of view, the only evidence supporting the exclusion of one between two indicators is interrelation. A high level of relation between two indicators allows us to consider just one of them, assuming that indicators showing high relation are actually measuring the same concept or component. However, this assumption is not necessary always true. The degree of freedom for such decisions is in the reality: the relationship between two indicators (e.g., number of firemen and amount of damages in a fire) can be high but mediated by a third one (e.g., dimension of the fire). If the nature of the third indicator changes, the relationship between the other two indicators changes or disappears, even though they will continue to describe, autonomously, the reality. If, by observing the previous high correlation, we excluded one of the two indicators, doing without one of them could deny ourselves precious pieces of the whole picture (as represented by the indicators). This means that having a solid conceptual model allowing indicators concepts' relationships to be identified and interpreted. The latter approach consists in combining the indicators in a meaningful way.

References

Anand, S., & Sen A. (1997) Concepts of human development and poverty: A multidimensional perspective. *Human Development Papers 1997*. New York: UNDP.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Berger-Schmitt, R. & Noll, H.-H. (2000). *Conceptual framework and structure of a european system of social indicators*, EuReporting Working Paper No. 9, Centre for Survey Research and Methodology (ZUMA) – Social Indicators Department, Mannheim.

Diamantopoulos, A., & Siguaw, J. A. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*(4), 263–282

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277.

Eurostat. (2000a). *Definition of quality in statistics*, Eurostat Working Group on Assessment of Quality in Statistics, Eurostat/A4/Quality/00/General/Definition, Luxembourg, April 4–5.

Eurostat. (2000b). *Standard quality report*, Eurostat Working Group on Assessment of Quality in Statistics, Eurostat/A4/Quality/00/General/Standard Report, Luxembourg, April 4–5.

Felce, D., & Perry, J. (1995). Quality of life: Its definition and measurement. *Research in Developmental Disabilities, 16*(1), 51–74.

Horn, R. V. (1993). *Statistical indicators*. Cambridge: Cambridge University Press.

Johansson, S. (2002). Conceptualizing and measuring quality of life for national policy. *Social Indicators Research, 58*, 13–32.

Land, K. C. (1971). On the definition of social indicators. *American Sociologist, 6*, 322–325.

Land, K. C. (1975). Social indicator models: an overview. In K. C. Land, & S. Spilerman (Eds.), *Social indicator models*. New York: Russel Sage Foundation.

Maggino, F., & Facioni, C. (2015). Measuring stability and change. Methodological issues in quality of life studies. *Social Indicators Research* (first online issue) http://link.springer.com/article/10.1007/s11205-016-1241-5

Meadows, D. (2008). *Thinking in systems: A primer*. White River Junction: Chelsea Green Publishing.

Michalos A. (1992). Use and abuses of social indicators In *Sinet*, n. 32.

Noll, H.-H. (1996). Social indicators and social reporting: The international experience. http://www.ccsd.ca/noll1.html

Noll, H.-H. (2004). *Social indicators and indicators systems: Tools for social monitoring and reporting*. Paper presented at OECD, World Forum "Statistics, knowledge and policy", Palermo, 10–13 November 2004.

Nuvolati, G. (1997). Uno specifico settore di applicazione degli indicatori sociali: La qualità della vita. In F. Zajczyk (Ed.), *Il mondo degli indicatori sociali, una guida alla ricerca sulla qualità della vita* (pp. 69–94). Roma: La Nuova Italia Scientifica.

Patel S., M. Hiraga, and L. Wang (World Bank) D. Drew and D. Lynd (Unesco) (2003). *A framework for assessing the quality of education statistics*, Development Data Group and Human Development Network, World Bank, Washington, DC

Sharpe, A., & Salzman, J. (2004). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Ottawa: Center for the Study of Living Standards.

Sirgy, M. J., Michalos, A. C., Ferriss, A. L., Easterlin, R. A., Patrick, D., & Pavot, W. (2006). The quality-of-life (QOL) research movement: Past, present, and future. *Social Indicators Research, 76*(3), 343–466.

Śleszyński, J. (2012). *Prospects for synthetic sustainable development indicators*, Paper presented at the conference "Quality of Life and Sustainable Development", September 20–21, Wroclaw (Poland).

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the commission on the measurement of economic performance and social progress*, Paris. http://www.stiglitz-sen-fitoussi.fr/en/index.htm

Chapter 5
Dealing with Syntheses in a System of Indicators

Filomena Maggino

As we have seen, indicators developed through the hierarchical design form a complex system. This characteristic may require approaches allowing more concise views able to summarizing the complexity. In this perspective, the guiding concept crossing all possible strategies is *synthesis*.

From the conceptual point of view, any synthesis should preserve the consistency between the single elements and their synthesis and represent a "stylization" and not a compression of the represented reality.

From the methodological point of view, synthesis concerns different aspects of the system (Maggino 2009)[1]:

- *synthesis of units* (cases, subjects, etc.), which aims at aggregating the individuals' value of one indicator (basic or synthetic) observed at micro level; this synthesis should allow the created macro units to be compared (social groups, age groups, geographic areas, time frame), with reference to the indicators of interest.
- *synthesis of basic indicators* which aims at aggregating the values referring to several indicators for each unit, micro (individual level) or macro (regional, national, group level).

One of the issues to be disentangled in order to proceed with any synthesis concerns the element which should represent the synthesis. As we will see, the synthesis could be represented by numerical values or a graphical representation.

---

[1] By looking at these aspects in terms of dimensions of the classical data matrix, we may say that the former perspective refers to the rows of a matrix (synthesizing rows), while the latter refers to the columns (synthesizing indicators).

F. Maggino (✉)
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy
e-mail: filomena.maggino@uniroma1.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_5

Synthesizing Cases Through Numerical Values

This kind of synthesis is aimed at defining *macro-units* synthesizing scores collected for many micro-units with reference to one [synthetic or not] indicator, as defined in the monitoring perspectives.

In other words, this perspective in synthesis aims at condensing values observed at micro (usually, individual) level to higher levels in order to produce new meaningful units, identified according to different kinds of scales. Generally, macro units refer to partitions which are pre-existent or pre-defined, such as geographical or administrative areas, social groups, time periods (years, decades, etc.).

Generally, this kind of synthesis is accomplished by applying very simple statistical instruments (e.g., average), which turn out to be very unsatisfying since they do not allow the phenomenon's distribution to be correctly represented. A possible (not necessarily the best) solution is to report, for example, the percentage of a subgroup or a dispersion index (standard deviation or inter-quartile range).

Aggregating scores collected at micro levels is a well-known issue in many scientific fields, like economics and informatics, where particular analytic approaches are applied (like the probabilistic aggregation analysis). In econometric fields, particular empirical methodologies have been developed, allowing the explanation of systematic individual differences (*compositional heterogeneity*) that can have important consequences in interpreting aggregated values (Stoker 1993). Other attempts aimed at weighting average values by different criteria can be identified (Kalmijn and Veenhoven 2005; Veenhoven 2005).

The kind of synthesis is generally obtained by averaging individuals' values at the level of interest (country, region, social group, and so on). From the statistical point of view, this corresponds to the identification of a value synthesizing the distribution.

The single value can be different according to the nature of data (Fig. 5.1).

If the number of indicators involved in the synthesis is more than one the index referred as centroid.

Actually, a distribution can be synthesized also by referring to its shape, in terms of *skewness* (asymmetry of the distribution) and/or *kurtosis* (tailedness of the distribution).

If the adopted index has particular limitations (as in the case of the mean), the synthesis may be obtained by identifying homogenous groups of micro-units and then aggregating individual values observed in them. This approach is based upon the *homogeneity* criterion[2]: within each level (area, group, and so on), individuals'

---

[2]Each sub-group represents a macro-unit defined in terms of *typology*. Identification of typologies requires particular analytical approaches, allowing homogeneous groups among individual cases to be identified (Aldenderfer and Blashfield 1984; Bailey 1994; Corter 1996; Hair et al. 1998):

– *segmentation analysis*, which can be conducted through different procedures (*Hierarchical Cluster Analysis, Q Analysis*);

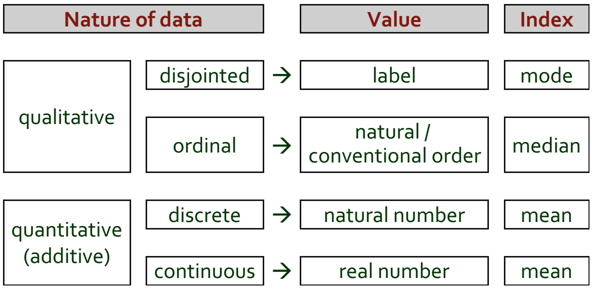


Fig. 5.1 Synthesis of units: different indexes according to nature of data

values are synthesized only if cases are homogeneous according to the involved indicator/s.

If more indicators are involved and not synthesized, each macro-unit can be represented by a profile of values, component values (generally proportions or incidences) describing the sub-groups. By using proportions, the sum of component values is constant.

Synthesizing Indicators Through Numerical Values: Dealing with Different Perspectives

The process of synthesizing indicators can be carried out by considering different perspectives which take into account

– the conceptual design that guided in defining the indicators (variables, dimensions, domains)
– the theoretical definition of the indicators (reflective or formative indicators)
– the technical issues of synthesis (weighting, aggregation techniques).

The Conceptual Perspective

By looking at the conceptual matrix illustrated in the previous chapter (Fig. 4.5), in which each row represents a dimension/sub-dimension and each column represents a domain, the synthesis of indicators (Fig. 5.2) can be achieved through the

– *partitioning analysis*, which can be conducted through different procedures, like K Means Methods, Iterative Reclassification Methods, "Sift and Shift" Methods, Convergent Methods.

Each analytical approach produces results that vary according to the decisions made in terms of (i) selected indicators; (ii) measures used in order to evaluate proximities between individual-points; (iii) method used in order to assign an individual-points to a group; (iv) criterion used in order to determine the number of groups; (v) criterion used in order to check the interpretability of the groups.

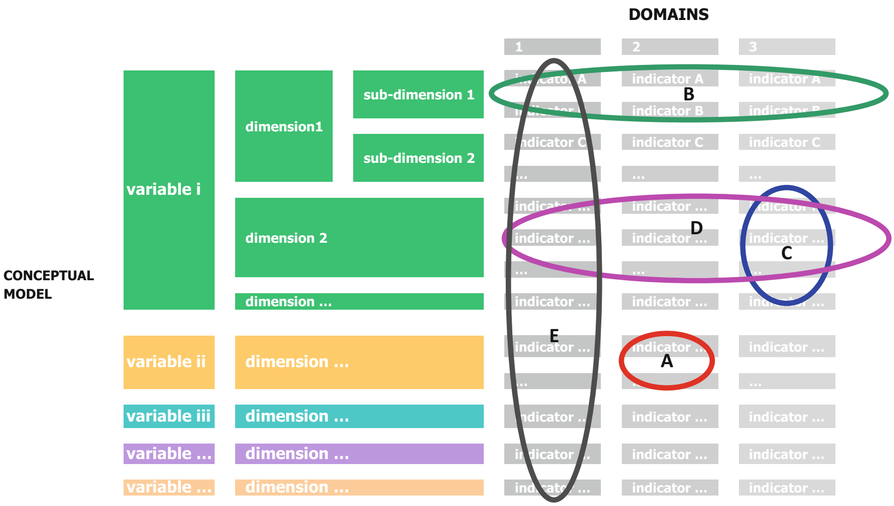


Fig. 5.2 The conceptual matrix: possible approaches to synthesis

following perspectives (illustrated through an example referring to subjective wellbeing):

- for each uni-dimensional variable the synthesis can involve indicators referring to
  - a single domain (e.g., satisfaction at work) → A
  - many/all domains (e.g., satisfaction in all domains) → B
- for each multidimensional variable the synthesis can involve indicators referring to
  - a single domain (e.g., subjective wellbeing at work) → C
  - many/all domains (e.g., subjective wellbeing in all domains) → D
- across variables
  - for a single domain (e.g., aggregating all indicators referring to the "work" domain) → E

More dimensions and domains (cells) are involved in the process, more difficult turns out to be the interpretation of the synthesis. This is particularly crucial in case the interpretation of the synthesis represents an important issue to consider.

The Model-of-Measurement Perspective

Aggregation of indicators should take into account the nature of the relationship between the defined concept and the selected indicators (Maggino 2009) which can be *reflective* or *formative*.

As we have seen, in the *reflective model*, indicators are seen as functions of the conceptual (latent) variable. Consequently, in synthesizing indicators in a reflective model it should be taken into account that they are highly correlated (internal consistency). As we know, reflective indicators are linearly related and interchangeable (the removal of an indicator does not change the essential nature of the underlying construct).

The statistical approach allowing this situation to be tested is factor analysis where an observed measure is assumed to be determined by latent factors. The fundamental equation of the factor model (for m indicators) is the following:

$$\sigma_{x_i}^2 = \sum_{j=1}^{m} \lambda_{x\,i}^2 \xi_j + \delta_{x_i}^2$$

where

$\sigma_{x_i}^2$ total variance of indicator x_i

$\lambda_{x_i \xi_j}$ factor loading of indicator x_i with reference to latent variable ξ_j

$\delta_{x_i}^2$ uniqueness (specific variance + error) of indicator x_i

The application of this analysis allows us to check the indicators consistent with each dimension of the defined concept. In case of multidimensional latent variables, the factor model allows the dimensionality to be tested; one score is consequently determined for each dimension. Traditionally, the reflective view is seen related to the development of scaling models applied especially in subjective measures (*scale construction*).

Any uncorrelated indicator cannot measure the defined conceptual construct and should be excluded. At the same time, it is possible to reduce their number in order to obtain simpler synthesis since the indicators in the reflective model are highly correlated and, consequently, interchangeable, (i.e., with less indicators).

In the *formative case*, indicators are seen as causes of the latent variable. This means that a concept is assumed to be defined by the selected indicators. The indicators are not assumed correlated to each other. In other words, two uncorrelated indicators can both contribute to the measurement of the same conceptual variable. Correlated indicators may turn out to be redundant in measuring the concept. Better, two correlated indicators included in a formative model actually let the conceptual contribution of the component measured by both weight more in the final synthesis.

The two different situations are in any case strictly related to the conceptual definition of the variable of interest. Let see an example in which the goal is to obtain a synthesis expressing "life satisfaction" (one of the dimension of the conceptual variable "subjective wellbeing) out of different indicators.

- *reflective synthesis*: life satisfaction is caused by individual personality traits (top down explanatory model); if true, satisfaction in domains should necessarily

correlate to each other. Factor analysis helps us in identifying indicators not correlated which should be excluded from the synthesis;
- *formative synthesis*: life satisfaction is formed by satisfaction in different domains (bottom up explanatory model); if true, satisfaction in domains should not necessarily correlate to each other. Principal Component Analysis can help us in order to select one indicator per each component to be synthesized.

The two different situations have important consequences in technical choices to be adopted in order to perform the synthesis.

In defining the procedure aimed at synthesizing formative indicators, four critical issues should be considered (Diamantopoulos and Winklhofer 2001):

- *Content specification*. It refers to the content domain that the synthetic indicator is intended to capture. In the ambit of formative model, content specification is inextricably linked with indicator specification.
- *Indicator specification*. Ideally, the indicators must cover the entire latent variable's content domain. The exclusion of an indicator is possible but causes the risk of changing latent variable specification. However, an excessive number of indicators is undesirable for difficulties in terms of data analysis. This issue is particular important especially in aggregative perspective.
- *Indicator collinearity*. Excessive collinearity among indicators makes it difficult to separate the distinct influence of the individual indicator on the latent variable. Multicollinear indicators turn out to be redundant and may cause the exclusion of one of them.
- *External validity*. In order to assess the measurement validity, the synthetic indicator should be related to other measures (external validity), since validity cannot be explored through the internal consistency perspective (which is typical of reflective approach). The basic idea is, in other words, to explore the quality of individual indicators by relating each of them with other indicators (external to the synthetic indicator): only the indicators significantly related to the variable of interest would be retained. Consequently, this process should be supported by a solid theoretical background. Another approach is that including some reflective indicators and estimates a multiple indicators and multiple causes (MIMIC) model (Diamantopoulos and Winklhofer 2001).

This approach finds its more known application in the composite indicator methodology.

The Technical Perspective

From the technical point of view, synthesis can be faced through two different approaches, aggregative-compensative and non-aggregative.

The Aggregative-Compensative Approach

By giving for grant that only one latent variable is measured, the aggregative approach requires applying a procedure requiring the following steps.

1. **Level of aggregation (*micro*/*macro level*)**

Identifying the level of synthesis is not a trivial issue. The decision depends on the meaning that should have the final aggregation. An example could help in illustrating the different meaning of syntheses accomplished through different sequences.

The affective component of subjective wellbeing is observed by collecting at individual level data concerning both positive and negative affects. Generally, the synthetic indicator, performed at individual level, is represented by the *affect balance*, defined as the difference between positive *affect*s and negative *affects*. Actually, even though the difference can be obtained also at macro level and since the obtained indicator should tell something related to the affective component of subjective wellbeing, the synthesis should be performed at micro level.

2. **Checking the dimensionality of the indicators to be aggregated (*dimensional analysis*)**

Selecting the indicators to be included in the aggregation represents a fundamental stage of the process of synthesis since it operationally defines the concept defining the phenomena that the aggregation is supposed to measure. The results allow the approach to synthesis to be evaluated. Actually, this step aims at

- testing (reflective model) or investigating (formative model) the level of complexity of the concept measured by the indicators in terms of dimensionality;
- refining the selection of the indicators showing the best statistical characteristics.

The two goals are generally pursued contextually through traditional analytical approaches. Beyond the criticisms concerning the metrics of data, the application of the traditional dimensional procedures puts other doubts, especially from the statistical perspective (as we see in another note in this chapter).

Dimensional analysis is based upon a single crucial information: correlation between indicators. The interpretation of the level of correlation is different according to the model of measurement identified.

According to the *reflective model*, the selected indicators are caused by the same latent variable. They can be considered *multiple measures* and assumed to contribute to the measurement of variable's major aspects allowing the variability of the defined latent variable to be covered. Consequently, highly correlated indicators prove that the model of measurement is *reflective*. The statistical evidence allowing the aggregation of reflective indicators is *internal consistency*. The latent construct can turn out to be multidimensional. In this case, the variance of each indicator can be explained by one dimension (but also by more dimensions). Aggregating indicators referring to each single dimension of the latent variable is conceptually admitted and the aggregated score can be easily interpreted.[3] Creating a single score out of all dimensions (i.e., aggregating all indicators referring to all dimensions of the latent variable) may be also consistent and meaningful.

In case of *formative models*, the correlation informs about the level of overlapping among indicators measuring the same unobservable variable.[4] Consequently, a high level of correlation between two indicators describes a redundancy and suggests the selection of only one of them. This decision should give preference

[3]The main and approach allowing to deal with reflective models is undoubtedly *Factor Analysis* (FA), which can be applied in order to test the hypothesized dimensional structure underlying the selected indicators. In particular, it allows indicators that fit better the latent dimensional structure to be synthesized. The approach is based upon the assumption that the total variance of each indicator represents a linear combination of three different components (additive assumption), *common variance* (due to the dimensional structure), *specific variance* (due to the specificity variance of each indicator), and *error*. Actually, this analysis allows, by estimating for each indicator the amount of common variance (*communality*), the reflective approach to be tested. Indicators turning out to be part of the same supposed underlying dimension can be meaningfully synthesized.

When the indicators refer to latent variables describing performances, a performance analysis, derived directly from the application of the *Item Response Theory*, can be applied in order to select indicators showing the best characteristics to discriminate among the units. In particular, the identified indicators can be distinguished from each other in terms of difficulty and discriminant capacity (Andersen 1972, 1973; Andrich 1988; Bock and Aitkin 1981; Hambleton et al. 1991; Lord 1974, 1984; Ludlow and Haley 1995; McDonald 1989; Rasch 1960; Sijtsma and Molenaar 2002; Swaminathan and Gifford 1982, 1985, 1986).

Very often, *Principal Component Analysis* (PCA) is applied to test dimensional structures by assimilating it to FA. This practice is strongly criticisable. In fact, the main goal of PCA is not to test a (dimensional) model but simply decompose the correlations among basic indicators in order to condense the variance among all the indicators as much as possible by calculating new linear variables, defined components. In other words, PCA describes the variation of a data set using a number of scores that is smaller than the number of the original indicators.

[4]From the technical point of view, several instruments are available:

- *Correlation Analysis*. It is useful in order to select indicators that are not redundant and to avoid multicollinearity (*double counting*) in composite indicator construction (Nardo et al. 2005a).
- *Principal Component Analysis*. The main goal of principal component analysis is to describe the variation of a data set using a number of scores that is smaller than the number of the original indicators. This approach is very often applied to test dimensional structures, even though this practice is strongly criticisable. This is done following the idea that this approach can be assimilated to Factor Analysis. The two approaches are actually, however, very different from each other. In particular, the main goal of Principal Component Analysis is not to test a (dimensional) model but simply to decompose the correlations among indicators in order to condense the variance among all the indicators as much as possible by calculating new linear variables, defined components.
- *MultiDimensional Scaling*. It allows the underlying dimensionality to be tested and a geometrical multidimensional representation (*map*) to be created for the whole group of indicators (Cox & Cox 1994; Kruskal & Wish 1978; Torgerson 1958).
- *Cluster Analysis*. In this context, it can be useful to identify meaningful groupings among indicators (Aldenderfer and Blashfield 1984; Bailey 1994; Corter 1996; Hair et al. 1998).

In some cases, the approaches can be combined (e.g., *tandem analysis* or *factorial k-means analysis*, Nardo et al. 2005a).

to the indicator allowing trend analysis and wide comparisons and proving to be available for a large number of cases are preferable.

From what has been said, the nature of a latent variable defined in the formative context is multidimensional. Consequently, aggregation of formative indicators raises some relevant issue. In general, when concepts are truly multidimensional, collapsing the indicators measuring them in just one synthetic indicator is very questionable. The nuances and ambiguities of the data would in fact be forced into a conceptual model where all the features affecting unidimensionality are considered as noise to be removed. Moreover, synthetic scores could be biased towards a small subset of basic indicators, failing to give a faithful representation of data.

3. **Defining the importance of each indicator in measuring the conceptual dimension** (***weighting Criteria***)

In case of formative model, stating the importance of indicators requires the adoption of a criterion able to reproduce as accurately as possible the contribution of each indicator to the construction of the synthesis. In this perspective, the definition of the weighting system can constitute an improvement and refinement of the adopted model of measurement. Cases' synthetic scores can sharply change from each other by simply changing the weights assigned to each indicator.

In this sense, apart from the applied approach, the defined weights represent judgment values. The researcher has to carefully evaluate and make formally explicit not only the methodology to be adopted but also the results that would have been obtained with other methodologies, also reasonably applicable.

From the technical point of view, the weighting procedure consists in defining and assigning a weight to each indicator. The weight will be used in the successive computation of the individual aggregate score when each weight is multiplied for the corresponding individual value of the indicator.

In order to define the weighting system, some decisions need to be adopted:

– proportional size of weights (equal or differential weighting)[5]
– approach to obtaining weights (objective or subjective)[6]
– level for obtaining and applying weights (individual or group)

---

[5]While *equal weighting* does not necessarily imply unitary weighting, the adoption of the *differential weighting* procedure do not necessarily corresponds to the identification of different weights but rather to the selection of the most appropriate approach in order to identify the weights among the following (Nardo et al. 2005a).

Assigning differential weights can be just as doubtful, especially when the decision is not supported by:

– theoretical reflections that endow a meaning on each indicator or consider its impact on the synthesis,
– methodological concerns that helps to identify the proper techniques, consistently with the theoretical structure.

[6]The approach is considered objective when the weights are determined through an analytic process, that is through

Independently from the approach adopted in order to define them, the weights can be kept constant or can be changed according to particular considerations concerning each application. In both cases, the researcher needs to rationalize the choice. The former approach can be adopted when the aim is to analyse the evolution of the examined phenomenon. The latter is useful when the aim – for example – concerns the definition of particular priorities.

In any case, we have to consider that a whole set of weights able to express in a perfect way the actual and relative contribution of each indicator to the measurement does not exist.

4. **Identifying the proper technique for aggregating indicators** (***aggregation technique***)

The selection of the proper aggregation technique should consider if it (i) admits or does not admit *compensability* among indicators; (ii) requires or does not require *comparability* (with reference to nature of data) among indicators; (iii) requires or does not require *homogeneity* in indicators' level of measurement

- An aggregating technique is *compensatory* when it allows low values in some indicators to be compensated by high values in other indicators.[7] A compensatory technique can be useful in some contexts especially when the purpose of applying indicators is to stimulate behaviours aimed at improving the overall performance by investing in those ambits showing lower values. The selection of the aggregating technique is an important choice also to avoid inconsistencies between the weights previously chosen – in terms of theoretical meaning and importance – and the way these weights are actually used. In other words, in

(i) statistical methods (*correlation, Principal Component Analysis*, *Data Envelopment Analysis*, *Unobserved Components Models*). The adoption of statistical methods in weighting socio-economic components has to be considered carefully since, by removing any control over the weighting procedure from the analysts, it may give a false appearance of mathematical objectivity that is actually difficult to achieve in social measurement (Sharpe and Salzman 2004).

(ii) multi-attribute models, like *Multi-Attribute Decision Making* (in particular, *Analytic Hierarchy Processes* – AHP) (Yoon and Ching-Lai 1995) or *Multi-Attribute Compositional Model* (in particular, *Conjoint Analysis*, CA).

The approach is considered subjective when shows the possibility to involve more individuals (experts or citizens) in the process of defining weighting systems for social indicators. These approaches are defined in the perspective of giving more legitimacy to social indicators by taking into account citizens' importance (values) and not – as usually done in the past – statistical importance.

[7]In the typical aggregating table (see below), we can observe all the possible synthetic scores obtained by aggregating two indicators (A and B) using simple addition or multiplication.

Some of the final scores, even though identical, are obtained through different combination of original scores. This means that obtained aggregated values do not allow us to return to original unit's profile. In other words, two units with different realities turn out to be identical and indistinct. The above figure suggests that also the multiplicative technique is compensatory, especially with reference to indicators showing low values.

order to continue interpreting the weights as “importance coefficients”, a non-compensatory aggregating procedure has to be preferred, such as a non-compensatory multi-criteria approach, like Multi-Criteria Analysis (MCA) (Nardo et al. 2005a).

- *Comparability* refers to the distributional characteristics of indicators, in particular to *directionality* and *functional form*.

Directionality refers to the direction by which each indicator measures the concept (i.e., positive or negative). The aggregative approach requires indicators uniformly oriented. In case an indicator needs to be re-oriented, it has to be submitted to the reflection procedure:

$$[(higher \cdot value \cdot observed) + 1] - (individual \cdot unit\text{'}s \cdot original \cdot value)$$

Functional forms represent how the changes are valued at different levels of an indicator’s distribution. If changes are valued in the same way, regardless of level, then the functional form is linear. If changes are valued differently, according to the level, the functional form is not linear. In other words, in some cases same absolute differences between observed values are valued differently and consequently can have different meaning (e.g., a change of €100 in terms of income can have a different meaning if it occurs at a high or at a low level of income).[8]

		B		
		1	2	3
A	4	5	6	7
	3	4	5	6
	2	3	4	5
	1	2	3	4

aggregation obtained through addition (additive approach)

		B		
		1	2	3
A	4	4	8	12
	3	3	6	9
	2	2	4	6
	1	1	2	3

aggregation obtained through multiplication (geometrical approach)

[8]If changes (Nardo et al. 2005a) are more significant at lower levels of the indicator, the functional form should be concave down (e.g. log or the nth root); on the opposite, if changes are more important at higher levels of the indicator, the functional form should be concave up (e.g. exponential or power). The standard choice is for log as the concave down function and power as the concave up function. Both the functional form are non-linear by definition. Applying the appropriate functional form helps to better interpreting the changes in the indicator. For example, a decrease of 5 years of life expectancy from a base level of 40 is more heavily penalized than the same decrease beginning at a level of 80 (Sharpe and Salzman 2004). Actually, many socio-economic indicators show non-linear functional forms, such as per capita GDP, measures of unemployment, poverty gaps and rates, measures of inequality (Sharpe and Salzman 2004). Anand and Sen (1997) state that, in measures of poverty deprivation “the relative impact of the deprivation ... would increase as the level of deprivation becomes sharper”.

- *Homogeneity* refers to the level of measurement adopted by the whole group of indicators. Almost all the aggregating techniques require homogeneous scales. Some techniques exist allowing the indicators' original scales to be transformed into an interpretable common scale. In order to select the proper approach, the data quality and properties and the objectives of the indicator should be taken into account.

The literature offers several **aggregation techniques** (Nardo et al. 2005a). The linear aggregation approach (additive techniques) is the most widely used. By contrast, multiplicative techniques (following the geometrical approach) and the technique based upon multi-criteria analysis (following the non-compensatory approach) allow the difficulties caused by compensation among the indicators to be overcome:

				Aggregating approaches			
				1. Linear aggregation		2. Geometrical aggregation	3. Non-compensatory aggregation
						Additive	Cumulative
Assumptions	Dimensionality	→	Relationships between indicators	Uni	Uni	Uni	Multi
	Model of measurement	→	Relationship between indicators and latent variable	Monotonic	Differential relationship	Monotonic	
	Compensation	→	Among indicators	Admitted	Not admitted (scalability of indicators)	Admitted	Not admitted
	Homogeneity	→	Of the level of measurement	Requested	Requested	Requested	Not requested

Criticisms of Aggregative-Compensative Approach

Unidimensionality from Multidimensionality

Despite its success, the aggregative approach has been deeply criticized as inappropriate and often inconsistent (Freudenberg 2003). Critics point out conceptual, methodological and technical issues especially concerning the difficulty in conveying into unidimensional measures all the relevant information pertaining to phenomena which are complex, dynamic, multidimensional, full of ambiguities and nuances, and which are represented by data being sensitive, qualitative (even when quantitatively measured) and containing errors and approximations.

In other words, a synthetic indicator obtained through the aggregative approach is hardly able to reflect the complexity of a socio-economic phenomenon and capture the complexity of the relationships between variables.

Process of Construction: Between Objectivity and Arbitrariness

The main argument supporting aggregative indicators is related to the need of communicating information in an easy way and adopting "objective" procedures. However, even though some decisions are strictly technical, it is quite difficult to make these decisions objective since they may involve different kind of concerns. In fact, as we have seen, each stage introduces some degree of arbitrariness in making decisions. In other words, although objectivity is always invoked as an essential requirement, in practice the procedures for computing aggregative indicators are far from being "aseptic".

The aggregative process is based mainly upon linear mathematical instruments, which require meeting assumptions concerning nature of data (i.e., quantitative), distributions (i.e., normal), and relationships (i.e., linear).

Nature of Data

The way ordinal indicators are managed can be mentioned among the examples of arbitrary methodological choices. As we know, data regarding socio-economic phenomena are mainly ordinal and discrete in their nature. Evaluating multidimensional poverty, measuring quality of products and services or assessing people satisfaction are just a few examples of issues faced by socio-economic scientists in their daily practice whose measurement produces data which are mainly ordinal in their nature. This is true also when data turn out to be apparently metric. In other words, ordinal data can represent the true expression of phenomena and not just a rough approximation of true and precise, yet non-observable, characteristics. Despite this, the aggregative approach is preferred in order to process and synthesize also ordinal indicators. On the contrary, it could be wise to accept the idea that the great part of socio-economic phenomena is actually observable through nuances and "ambiguities", which do not conciliate with quantitative metrics and which do not represent obstacle to be removed, but often are what really matters.

With the aim of pursuing metric analysis out of non-metric data, a lot of arbitrary choices are often taken in data analysis. In this perspective, strong assumptions are made about the metric nature of ordinal data, by stating the equivalence of intervals between the steps in ordinal scales. As an alternative, several methodological efforts are devoted to making ordinal measures quantitatively more precise. In particular, ordinal indicators are submitted to more or less sophisticated scaling tools with the ambition of transforming ordinal data into metric data. However, those tools tend to impose a quantitative latent model to data, which is not fully justifiable on any epistemological basis since they force and do not respect the qualitative nature of ordinal data. These procedures may sometimes lead to useful results, but they are often quite questionable, not being consistent with the intrinsic nature of data. Consequently, the efforts for getting more precise measures have the effect of frequently forcing the true nature of socio-economic phenomena.

Shape of Distributions and Nature of Relationships

The aggregative approach is based upon classical data analysis approaches which assume linear structures among indicators, analyzed usually through the analysis of covariance matrices. As a consequence, the results of these kinds of analysis are affected by some degree of arbitrariness and are not so easy to interpret, from a socio-economic point of view. Although these problems are well-known, and new methodologies are continuously under development in the socio-economic literature, they are still unsolved. Many approaches have been proposed, even involving highly sophisticated inferential tools, but none of them seems to be decisive. Basically, it can be asserted that the issue of ranking and evaluation in an ordinal setting is still an open problem, since the statistical methodologies, applied in the common practice or proposed at theoretical level, are unsatisfactory in many respects.

Definition of Weighting Systems

As we have seen, the necessity of selecting weights by grounding the choice on objective principles is frequently asserted (Nardo et al. 2005a; Sharpe and Salzman 2004), by giving preference to weighting systems produced through statistical tools. However, adopting purely statistical methods in weighting socio-economic components must be carefully considered. Removing any control over the weighting procedure from the analyst gives a possibly false appearance of objectivity that is actually difficult to achieve in socio-economic measurement (Sharpe and Salzman 2004). Moreover, since defining weights is often interpreted in the perspective of identifying individual and social values, the procedure should necessarily involve individuals' contributions in attributing importance to different domains. Sometimes, the choice and decision could be shared by a larger community (involving individuals in the process of social indicators construction). If indicators concern societal wellbeing, their construction turns out to be not just a technical problem, being part of a larger debate aimed at obtaining a larger *legitimacy*. In this perspective, the weighting issue can be even considered as a leverage of democratic participation to decisions. In their work (2007), Hagerty and Land stressed how building composite indicator should take into account and maximize the agreement among citizens concerning the importance of each elementary indicator. Choosing consistent weighting criteria is thus a subtitle issue, largely subjective and possibly data independent.

Aggregation Approach

The aggregating process, needed to get unidimensional scores out of multidimensional data, rises further methodological difficulties. The process is in fact quite controversial since:

– the indicators to be aggregated are rarely homogeneous in many respects (metrics, directionality, functional form, ...) and need not share common antecedents;
– the aggregating technique might introduce implicitly meaningless compensations and trade-offs among indicators;
– it is not clear how to combine ordinal variables and/or use numerical weights.

The Non-aggregative Approach

As we have seen, even though it represents the mainstream approach to synthesize indicators, many critical issues affect the aggregative-compensative methodology, which faces an impasse:

(i) implicitly or not, it is generally taken for granted that "evaluation implies aggregation"; thus
(ii) ordinal data must be scaled to numerical values, to be aggregated and processed in a (formally) effective way; unfortunately
(iii) this often proves inconsistent with the nature of phenomena and produces results that may be largely arbitrary, poorly meaningful and hardly interpretable.

Realizing the weakness of the syntheses built through the aggregative-compensative approach, statistical research has focused on developing alternative and more sophisticated analytic procedures, but almost always assuming the existence of a cardinal latent structure behind ordinal data. The resulting models are often very complicated and still affected by the epistemological and technical issues discussed above.

The way out to this impasse can be found in realizing that synthesis does not necessarily imply aggregation. In other words, non-aggregative approaches are needed, able to (i) respect the ordinal nature of the data and the process and trends of phenomena (not always linear but more frequently monotonic), (ii) avoid any aggregation among indicators, and (iii) producing a synthetic indicator.

New challenges and perspectives are emerging aimed at developing technical tools and strategies in order to reduce data structure, combine indicators and communicate the obtained "picture".

The Partially Ordered Sets Approach

In this perspective, one of the most useful references is the *Partial Order Theory*, a branch of discrete mathematics providing concepts and tools that fit very naturally the needs of ordinal data analysis. Non-aggregative approaches are focused not on dimensions but on *profiles*, which are combinations of ordinal scores, describing the «status» of an individual. Profiles can be mathematically described and analyzed through tools referring to that theory, in particular *Partially Ordered Set* (POSET), allow for the extraction of information directly out of the relational structure of the data and provide robust results, not requiring binding assumptions.

Through partial order tools it is possible to give an effective representation of data and their structure and to exploit the latter to extract, directly out of it, the information needed in the synthesizing process. In particular, the computations performed to assign numerical scores to the statistical units involve only the ordinal features of data, avoiding any scaling procedure or any other transformation of the kind. Consequently, the conclusions drawn through the application of POSET methodologies are much more meaningful, robust and consistent than those based upon traditional statistical tools.

The methodology is introduced and illustrated in other chapters of this volume.

Assessing the Robustness of the Synthesis

This stage aims at proving that the results obtained through the synthesis are not affected by the choices made along the process. The choices may concern not only the synthesizing process (definition of the weighting system, the selection of the aggregation technique and so on) but also the data management process (missing data imputation, normalization, etc.).

It could be interpreted as an attempt to objectify the choices, turned out to be inevitably subjective. Actually, this stage aims at defending the choices through evidences. However, this approach does not urge a methodological defence of the choices, in terms of scientific responsibility.

A. Uncertainty and Sensitivity Analysis

Independently from the adopted synthesizing approach, many choices have to be taken along the process; each choice may influence the robustness of the synthetic indicator.[9] Assessing the robustness allows us to evaluate roles and consequences of choices. The assessment procedure, which can be included in the wider field of *what-if analysis*, is conducted through two stages; each stage corresponds to a different analytical methodology (Nardo et al. 2005a):

1. *uncertainty analysis*: this method aims at analysing to what extent the synthetic indicator depends on the information composing it. In order to evaluate how the uncertainty sources influence the synthetic score, the procedure identifies different scenarios for each individual case; each scenario corresponds to a certain combination of choices that produces a certain synthetic value;

[9]The robustness is the capacity of the synthetic indicator to produce correct and stable measures (Edward and Newman 1982; Nardo et al. 2005a; Saisana et al. 2005; Saltelli et al. 2004; Sharpe and Salzman 2004; Tarantola et al. 2000).

2. *sensitivity analysis*: this method aims at evaluating the contribution of each identified source of uncertainty by decomposing the total variance of the synthetic score obtained; to this end, the procedure tests how much the synthetic score is sensitive to the different choices (small differences reveal low sensitivity).

The two methodologies, generally treated in separate contexts, are very popular in any scientific field that requires the development and assessment of models (financial applications, risk analyses, neural networks); in addiction, the *uncertainty analysis* is adopted and applied more frequently than the *sensitivity analysis* (Jamison and Sandbu 2001).

B. Assessing Discriminant Capacity

Assessing the discriminant capacity (Maggino 2007) aims at exploring ability of the synthetic score to:

– discriminate between cases and/or groups (supposed to be different with reference to the conceptual variable synthesized through indicators); this can be accomplished by applying the traditional approaches of statistical hypothesis testing;
– distribute all the cases without any concentration of individual scores in few segments of the continuum; to this end, some coefficients were defined (Guilford 1954; Maggino 2003, 2007);
– show values that are interpretable; this requires to identify particular values or reference scores, identified according to particular criteria; the reference scores, generally called **cut-point** (continuous data) or **cut-off** (discrete data), are particular useful when the synthetic indicator is applied for diagnostic and screening purposes.

Synthesizing Through Graphics

As constantly stated in this volume, representing a complex and multidimensional world through indicators produces a complex and multidimensional system. As we have seen, trying to synthesize indicators through numerical values has many advantages, among which we can mention

- answering the call by policy makers for condensed information
- improving the chance to get into the media with simple information
- allowing multi-dimensional phenomena to be represented in a uni-dimensional way
- allowing situations across time more to be easily compared
- comparing cases (e.g. countries) in a transitive way (ranking and benchmarking)

However, the many criticisms showed by the "numerical" approach invite to explore also another strategy of synthesis by using graphical instruments (visual display).

While the numerical approach concerns the "reduction" of many values in just one (or, at least, very few), the graphical perspective concerns the deduction of many values in a visual display which typically uses a bi-dimensional representation. The main issue is, how can we represent the world of measures on a mere flatland? (Tufte 1990).

Joint Representation of Indicators: Dashboards

Dashboards represent useful tools aimed at simultaneously representing, comparing and interpreting indicators' values

– through an analogical perspective
– by setting them on a standardized scale
– by representing them on a colour scale (e.g., a green-to-red colour scale).

Several software programmes (free or not) can be used in order to carry out the graphical representation through different images (car dashboards, analogical bars, digital bars, thermometers).

Whichever representation form is adopted, indicators' values are displayed through

• separated values (values are not aggregated), allowing weak and strong points to be identified,
• colours, allowing the analysis of relative performance (value to be displayed relatively to an expected value or a given level/targets)
• distributions, allowing assessment indicators' meaningfulness, outliers identification, etc.
• scatter plot graph, allowing simple correlation analysis between the indicators to be visualized. This function allows synergies (indicators whose "desirable" values are positively correlated) and potential conflicts (e.g. environment vs. many economic and social variables) to be identified.

Through the graphical display, dashboards allow comprehensive monitoring and evaluation of programmes, performances or policies, since

• highly complex systems of indicators can be represented by taking into account the hierarchical design,
• easy communications are possible through a catchy and simple graphical representation,
• indicators can be related to weights interpreted in terms of

  (a) *importance* (reflected by the size of the segments) and

(b) *performance result* (reflected by the colour, interpretable in terms of "good vs. bad")

- performances of different cases can be compared.

Of course, dashboard does not allow complex analysis to be accomplished concerning relationships between indicators and comparisons of performance over time (trends) or across units (inter-cases comparisons). Dashboards can be useful in the perspective of creating composite indicators.

A Step Forward: Visual Complexity

As we have seen, a system of indicators represents a complex structure essentially defined by three components, functions, elements, and interconnections (v. Chap. 4). Complexity affects not only the structure but also the whole data organization (many databases to be integrated) collection (many surveys might be necessary) and processing (several steps and analytical methodologies are required).

Operating a synthesis in a complex structure may represent a very difficult task, urging solutions which are able to apply different instruments in combination, including graphical devices.

Using graphical approach to synthesis could represent a step forward with respect to the analytical approaches, especially in its potential capacity to represent a system of indicators (or part of it) in particular ways able to support users in making statements and asking questions about the indicators and allowing a critical understanding the reality they try to represent. This represents a big challenge especially if we think that new sources of data (big data) would support the construction of indicators. (Lima 2011).

In particular, using visualization in order to support the need of having synthetic representations since it is able to manage large amount of data in a small space, to support the monitoring exercise and the recognition of patterns among indicators and exploring their dynamic interconnections.

Visual complexity may represent an alternative/complementary approach to synthesis. It is located exactly at the crossroad of different competences, i.e. image, word, number and art. This field of studies refers directly to the concept of complexity and recalls the main characteristics of a system.

Actually, visual complexity represents the intersection of two technical and cultural fields, networks and visualization (Lima 2011). While the first has some tradition in social sciences, the explosion of online social networks and the increasing availability of web information linked to each other have raised the need of having the possibility to visually represent them.

Instruments

The traditional statistical graphical instruments (charts, scatter plots, line charts, bar and pie charts, and so on) aim actually at synthesising data. However, as it happens with the traditional synthetic statistics, they are not able to represent data describing complex phenomena and relations. The instruments of visual complexity share some "linguistic" elements (lines, points, curves, and so on) but the context is different; in place of having a single synthetic score, the goal is to obtain a single leaf/plan of representation. The visual representation of complexity deals with (i) a high volume of data (number of indicators and the compound data collection processes behind them), (ii) a high number of data dimensions, (iii) many interrelations. In a sense, visual complexity can support the representation of systems of indicators and the difficult exercise of discovering patterns, connections, and structures in them. (Lima 2011).

Metaphors and Language Syntax

The "traditional" visual complexity tries to build narratives by adopting methodologies extracted from mainly two fascinating metaphors, usefully applicable to systems of indicators:

- the *three of life*, where the three metaphor is used as an instrument able to represent hierarchies and classify elements (particularly useful in hierarchical systems);
- the *networks*, where the net metaphor is used to represent diversity, decentralization, non-linearity.

In order to use visual methodologies to allow more synthetic views of indicators, it is important to consider their syntax, which concerns the visual language in terms of

- instruments and elements (colour, text, size, shape, contrast, transparency, position, orientation, layout and configuration)
- methodological approaches (arc diagram, area grouping, centralized burst, centralized ring, circle globe, circular ties, elliptical implosion, flow chart, organic rhizome, radial convergence, radial implosion, ramifications, scaling circles, segmented radial convergence, sphere,) for which visual experts are proposing different approaches, allowing both dimensions of synthesis (cases and indicators).

Perspectives and Applications

According to Lima's suggestions (2011), the general challenges of visual complexity can be identified in different applications, like representing complex and dynamic data, visualizing priorities, ambient visualization, and collective intelligence (cybernetics). Among them, the first two seem to be directly related to indicators, their construction, managements, synthesis, policy use and communication.

Seeing the World in Data In the past, the availability of data allowing the construction of indicators relied almost entirely on ad-hoc collections (surveys but not only). In the era of big data, the availability increased and opened new perspectives in data collection and use. In fact, digital and electronic instruments allow

- interactive data collection and communication (social data collection), launching new possibility to found a citizen science (participatory sensing) with serious consequences on social indicators definition and construction;
- management of continuous data flows, requiring proper analytical instruments and algorithms to distil information and transform it into meaningful data (extracting interesting points from the flow);
- visualization of complex data, making data more accessible and meaningful.

Visualizing Priorities As we have seen, the main goal of a system of indicators is monitoring. Actually a systematic and comprehensive monitoring of a system allows critical elements and situations to be identified. This function is particularly important if seen in the sustainability perspective, which requires not necessary ad-hoc indicators but the analysis of the interrelations of indicators composing a system. In this sense, visualization can play a crucial role.

Visual complexity is mainly based upon *visual analytics*, able to combine visual-interactive techniques (able to cope with large data volumes and based on discrete scales and indicators difficult to compare from the analytical point of view) and data analysis methods and analytical and computational processes (e.g. data mining) in a context requiring complex activities to be performed (e.g., sense making, reasoning, decision making).

. -. -. - .

In conclusion, visual complexity represents a big challenge in the field of indicators construction and synthesis since it has interesting potentialities in visually managing complex systems of indicators and translating large volumes of data, systematically organized, into "more digestible insights, creating an explicit bridge between data, [indicators] and knowledge." (Lima 2011).

In fact, the capacity to translate data into a form able to highlight important features, including commonalities and anomalies, makes visual representations a concrete challenge in synthesizing indicators by supporting not only the monitoring function but also the analytical reasoning process and decision making.

However, similarly to what we saw for the analytical approaches, the graphical approach to synthesising indicators is not extraneous to risks and criticisms since many decisions should be taken. In fact, in order to allow patterns, relations, dimensions to be communicated in a more efficient and clear way, many decisions should be taken in managing graphical instruments. For example, the relationship among indicators could be represented in terms of network visualization which may emphasize different aspects (density, organic growth, instability, dynamism) and/or different structure (symmetry, top-down, stable dimensions). One of the risks is that the synthetic representation of indicators could be created in order to obtain just more aesthetically interesting description.

By adopting Tufte's words, (2001) "what is to be sought in designs for the display of information is the clear portrayal of complexity. Not the complication of the simple; rather the task of the designer is to give visual access to the subtle and the difficult – that is, the revelation of the complex.".

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*, Sage University Paper series on quantitative applications in the social sciences, series no. 07–044. Beverly Hills:Sage.

Anand, S., & Sen, A. (1997). Concepts of human development and poverty: A multidimensional perspective. *Human Development Papers 1997*. New York: UNDP.

Andersen, E. B. (1972). The numerical solution of a set of conditional Estimation equations. *Journal of the Royal Statistical Society, Serie B, 34*, 42–54.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*(1), 123–140.

Andrich, D. (1988). *Rasch models for measurement*. Sage University Paper series on quantitative applications in the social sciences, series no. 07–068. Newbury Park: Sage.

Bailey, K.D. (1994). *Typologies and taxonomies. An introduction to classification techniques*. Sage University Paper series on quantitative applications in the social sciences, series no. 07–102. Thousand Oaks: Sage.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psycometrika, 46*(4), 443–459.

Corter, J. E. (1996). *Tree models of similarity and association*. Sage University Paper series on quantitative applications in the social sciences, series no. 07–112. Thousand Oaks: Sage.

Cox, T. F., & Cox, M. A. A. (1994). *Multidimensional scaling*. London: Chapman & Hall.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277.

Edward, W., & Newman, J. R. (1982). *Multiattribute evalutation*. Sage University Paper series on quantitative applications in the social sciences, series no. 07–026. Newbury Park: Sage.

Freudenberg, M. (2003). *Composite indicators of country performance: A critical assessment*. STI Working Paper, 2003/16, Industry Issues, Paris: OECD.

Guilford, J. P. (1954). *Psychometric methods*. New York/London: McGraw-Hill.

Hagerty, M. R., & Land, K. C. (2007). Constructing summary indices of quality of life: A model for the effect of heterogeneous importance weights. *Sociological Methods and Research, 35* (4), 455–496.

Hair, J.F. Jr., Anderson, R. E., Tatham, R. L., & Black, W.C. (1998, 5th ed.) *Multivariate data analysis*. Upper Saddle River: Prentice-Hall Inc.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991) *Fundamentals of item response theory*, measurement methods for the social sciences series, vol. 2. London: SAGE Publications.

Jamison, D., & Sandbu, M. (2001). WHO ranking of health system performance. *Science, 293*, 1595–1596.

Kalmijn, W. M., & Veenhoven, R. (2005). Measuring inequality of happiness in nations'. In search for proper statistics. *Journal of Happiness Studies, 6*(4), 357–396.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07–011. Newbury Park: Sage.
Lima, M. (2011). *Visual complexity. Mapping patterns of information*. New York: Princeton Architectural Press.
Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39*(2), 247–264.
Lord, F. M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21*(3), 239–243.
Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: Interpretation, use and transformation. *Educational and psychological measurement*. London: Sage Periodicals Press, *55*(6), 967–975.
Maggino, F. (2003). *Method effect in the measurement of subjective dimensions*. Firenze: Firenze University Press, Archivio E-Prints.
Maggino, F. (2007). *La rilevazione e l'analisi statistica del dato soggettivo*. Firenze: Firenze University Press.
Maggino, F. (2009). *The state of the art in indicators construction in the perspective of a comprehensive approach in measuring well-being of societies*. Firenze: Firenze University Press, Archivio E-Prints.
McDonald, R. P. (1989). Future directions for item response theory. *International Journal of Educational Research, 13*(2), 205–220.
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., (EC/JRC), Hoffman, A., & Giovannini, E (OECD) (2005a). *Handbook on constructing composite indicators: Methodology and userguide*, OECD, Statistics Working Paper.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A, 168*(2), 1–17.
Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice. A guide to assessing scientific models*. Chichester: Wiley.
Sharpe, A., & Salzman, J. (2004). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Ottawa: Center for the Study of Living Standards.
Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Measurement methods for the social sciences series, vol. 5. London: SAGE Publications.
Stoker, T. M. (1993). *Empirical approaches to the problem of aggregation over individuals* (pp. 1827–1874). XXXI: Journal of Economic Literature.
Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics, 7*(3), 175–191.
Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika, 50*(3), 349–364.
Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika, 51*(4), 589–601.
Tarantola, S., Jesinghaus, J., & Puolamaa, M. (2000). Global sensitivity analysis: A quality assurance tool in environmental policy modeling. In A. Saltelli, K. Chan, & M. Scott (Eds.), *Sensitivity analysis* (pp. 385–397). New York: Wiley.
Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York/London/Sydney: Wiley.
Tufte, E. R. (1990). *Envisioning information*. Cheshire: Graphic Press.
Tufte, E. R. (2001). *The visual display of quantitative information*. Cheshire: Graphic Press.
Veenhoven, R. (2005). Inequality of happiness in nations. *Journal of Happiness Studies, 6*(4), 351–355.
Yoon, K. P., & Ching-Lai, H. (1995). *Multiple attribute decision making*. Sage University Paper series on quantitative applications in the social sciences, series no. 07–104. Thousand Oaks: Sage.

Chapter 6
Scalability of Composite Indices: Complexity Complications and Findings from 15 Years of Monitoring Child and Youth Well-Being in the United States

Kenneth C. Land, Vicki L. Lamb, and Xiaolu Zang

The globalized world of the twenty-first century is characterized by complex human interactions and social conditions. At the same time, the social indicators movement of the 1960s and 1970s produced a tremendous increase in the richness of social data available for many countries, in which the key role of the *quality-of-life/well-being concept* in connecting social indicators to the study of subjective well-being has become became vividly evident (Land 2015b). One consequence is that the field of social indicators entered an era of construction and study of *composite or summary social indicators/indices* in the 1990s and 2000s. Often these indices attempt to summarize (objective and/or subjective) outcome or well-being indicators of a number of domains of life into a single index of the quality-of-life for the population or society as a whole or for some significant segment thereof (e.g., children and youth, the elderly, racial and minority groups, cities, states or regions within the nation, etc.). Composite indicators thus attempt to answer the questions that motivated the social indicators movement: How are we doing overall in terms of the quality-of-life? With respect to our past? With respect to our societal goals? With respect to other comparable units (e.g., cities, states, regions, nations)? And,

---

Paper presented at the Dealing with Complexity in Society: From Plurality of Data to Synthetic Indicators Conference, University of Padua, Italy, September 17–18, 2015.

K.C. Land (✉)
Social Science Research Institute and Department of Sociology, Duke University, Durham, NC 27708, USA
e-mail: kland@soc.duke.edu

V.L. Lamb
Department of Human Sciences, North Carolina Central University, Durham, NC 27707, USA
e-mail: vlamb@nccu.edu

X. Zang
Sanford School of Public Policy and Department of Sociology, Duke University, Durham, NC 27708, USA
e-mail: dewelve@gmail.com

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_6

within our societies, how are specific segments of the population (e.g., children, the elderly, racial and ethnic minorities, immigrants) doing?

This chapter addresses the question: Can properties of a society, such as composite indices, be scaled across time periods and levels of analysis – from the whole system to subunits thereof? Because societies are considered complex systems, indeed among the most complex, we address this question within the context of two general sets of equations of state for complex systems – a nonlinear deterministic dynamics model defined by difference or differential equations and a model that incorporates stochastic (uncertainty) elements into its specifications, thus leading to the various classes of statistical models that we use in our daily research. As a case study of the use of composite indices, the U.S. Child and Youth Well-Being Index (CWI) and its various empirical findings for the years 1975–2014 are reviewed. We then use the CWI as an example to address the question of whether the scalability of composite social indicators is more likely due to one of the complexity models or the other. We conclude that both complexity models appear to be applicable and useful for studying, interpreting, and scaling composite indicators.

The Theory of Models

As a framework within which to describe two classes of complexity models and associated scalability issues, we call upon the theory of models (Casti 1992a, b; Gorski 2004; Land 2001). This theory commences with the following general and encompassing definition: *models* are cognitive tools/ linguistic devices by which we order and organize experiences and observations. And the contemporary branch of the philosophy of science known as *scientific realism* (Suppe 1989) takes the view that the objective of science is to reveal the underlying structures of the world that generate different outcomes under different conditions, for which models may be useful tools.

Formal or mathematical models are specific types of cognitive tools for ordering our experiences, that is: *formal or mathematical models* encapsulate some slice of experiences/observations within the confines of the relationships constituting a formal system such as formal logic, mathematics, or statistics (Casti 1992a, p. 1). Consider a particular subset S of social life, and suppose that S can exist in a set of conceptually distinct abstract states $\Omega = \{\omega_1, \omega_2, \ldots\}$. The set Ω defines the *state space* of S. In quantitative (mathematical, computational, statistical) models, an observable of S is a rule f associating a real number with each ω in the state space Ω. More formally, an *observable* is a map $f: \Omega \rightarrow R$, where R denotes the set of real numbers.

In an absolute sense, for a full accounting of social life, we need an infinite number of observables $f_\alpha: \Omega \rightarrow R$, where the subscript α ranges over a possibly uncountable index set. That is, social life S is described by Ω and the possibly infinite set of observables $F = \{f_\alpha\}$. But it is impossible to deal with such a large set of observables, and it is not necessary to do so in order to build useful models and/or theories. In brief, in model construction, most of the possible observables in social life are thrown away and attention is focused on a proper subset A of F.

We now can characterize a *social system model* S^* as an abstract state-space Ω together with a finite set of observables $f_i\colon \Omega \rightarrow R,\ i = 1, 2, \ldots, n$ (Land 2001, p. 384). Symbolically,

$$S^* = \{\Omega, f_1, f_2, \ldots, f_n\}.$$

But there is more to the notion of a social system model than just the list of observables by which it is characterized. The essential "systemness" of S^* is contained in relationships that link the observables. These relationships are termed the *equations of state* for S^*. Formally, the equations of state can be written as

$$\Phi_i(f_1, f_2, \ldots, f_n) = 0, \qquad i = 1, 2, \ldots, m,$$

where the $\Phi_i(\cdot)$ are mathematical relationships expressing the dependency relations among the observables. This can be more compactly written as

$$\Phi(f) = 0. \tag{6.1}$$

To capture the scientific realist sense of a model, the observables may be regarded as in two classes. One class represents the *observable conditions* under which some process in state space occurs. In turn, these divide into the slow-changing conditions that can be represented as *constant parameters* in the time frame of the process and the more rapid changing conditions that can be called *inputs*. The other class represents the *observable outputs* of the process. Suppose that:

- the last m observables $f_{n-m+1}, \ldots, f_m$, called *endogenous* (or determined within the system under consideration),
- are functions of the remaining observables $f_1, f_2, \ldots, f_{n-m}$, where the latter are termed *exogenous* (or determined outside the system under consideration).

The endogenous elements of the model correspond to the *outputs* of the model and the exogenous terms correspond to the *inputs* with the given state description. In other words, suppose we can define m functional relations, with some finite number r of numerical parameters, $\beta_1, \beta_2, \ldots, \beta_r$, for determining values of the endogenous observables as a function of the exogenous observables. If we introduce the notation

$$\beta = (\beta_1, \beta_2, \ldots, \beta_r)$$

to denote the vector of parameters and the notation

$$x = (f_1, f_2, \ldots, f_{n-m})$$

and

$$y = \left(f_{n-m+1}, f_{n-m+2}, \cdots, f_n\right)$$

to denote vectors of the exogenous and endogenous observables, henceforth *variables*, respectively, the equations of state then become

$$y = \Phi_\beta(x). \tag{6.2}$$

Two Complexity Models

Within the context of the theory of models, we now turn to *the question of scalability*, that is: Can properties of a society, such as composite indices, be scaled across levels of analysis – from the whole system to subunits thereof? Because societies are considered complex systems, indeed among the most complex, we address the question of scalability within the context of two general sets of equations of state for complex systems.

Complexity Model 1: Nonlinear Dynamics If the equations of state (2) for the observables are nonlinear dynamic (difference or differential) equations, the model may be what has come to be known in recent decades as a *dynamic complexity model*. Studies of nonlinear dynamic models over the past 50 years have shown that even the simplest nonlinear models can exhibit:

- *sensitivity of solutions* (trajectories (over time) of the endogenous variables) to initial conditions,
- *bifurcations of flows over time* (points at which qualitative changes in the long-term behavior of the endogenous variables occur), and
- *chaos* (recurrent behavior that is not in an equilibrium) with *attractors* (sets of numerical values towards which the system tends to evolve) that are *strange or fractal* (highly irregular), and
- *catastrophe points* (points at which the trajectories of the endogenous variables shift discontinuously to another type of behavior).

An example is the Lorentz (1963) model in meteorology for convection due to atmospheric heating, which is defined by three state (differential) equations:

$$\frac{dx}{dt} = \sigma y - \sigma x, \quad \frac{dy}{dt} = \rho x - xz - y, \quad \frac{dz}{dt} = xy - \beta z, \tag{6.3}$$

where x (intensity of convective motion), y (temperature difference between ascending and descending currents), and z (distortion of the vertical gradient from linearity) are the state variables, t is time, and σ, ρ, and β are the system parameters. This model is famous in the nonlinear dynamic models literature because Lorenz (Lorentz 1963) was the first to show that its attractor has the foregoing properties of sensitivity of solutions to initial conditions and fractal attractors, with a corresponding explanation of the deterioration of accuracy of weather forecasts with time.[1] This demonstration stimulated much research on the properties of numerical solutions to nonlinear dynamic models – often popularized under the "catastrophe" and "chaos" labels (see, e.g., Gleick 1987).

Concepts and principles from dynamic complexity models also have been applied conceptually and qualitatively to the study of social dynamics. An example is a graphical representation of the rise and fall of the Roman Empire as a "cusp catastrophe" (Casti 1992a, p. 129). This depiction is based on Arnold Toynbee's (1961) classical *A Study of History* that emphasized the rise and fall of great civilizations as being a process of challenge and response. Based on Toynbee's historical analysis showing that an external threat can either increase or weaken a civilization's integrity, the cusp catastrophe representation shows small levels of challenge and high levels of political and economic integrity during the formative years of the Roman Empire. This is followed by periods of rising external challenge and declining levels of political and economic integrity as the Empire resists the challenges – leading to a catastrophic fall through the "cusp" from central Empire domination to complete destruction and external subjugation as the Empire is overthrown.

Complexity Model 2: Stochastic/Statistical Models A second complexity model incorporates *stochastic (uncertainty) elements* into the equations of state (2). For this, suppose we define an additive vector:

$$\varepsilon = (\varepsilon_{n-m+1}, \varepsilon_2, \ldots, \varepsilon_n)$$

of error terms for each equation (with the usual specifications on the error terms, namely, that the expected value of each ε_i, $E(\varepsilon_i) = 0$ with constant variance, $E(\varepsilon_i \varepsilon_i) = \sigma^2{}_i$, $i = 1, 2, \ldots, m$), one for each endogenous variable, to take explicitly into account the fact that there may be stochastic shocks to the equations of state due either to factors unaccounted for in our system model or to an intrinsic random element in the behavior of the endogenous variables. Then the equations of state (2) become:

$$y = \Phi\beta(x) + \varepsilon. \tag{6.4}$$

[1] As Casti (1992a, p. 306) notes, the classical Poincare-Bendixson Theorem states that no solution of a deterministic differential equation system of less than three dimensions (endogenous variables) can display the aperiodic trajectories characteristic of chaotic motion. Accordingly, the fractal attractors of chaotic systems do not occur in systems defined by one or two differential equations; Lorenz's discovery of a chaotic attractor was due to his study of the numerical solutions of the three equation system for the intensity of convective motion. Since the complexity of social systems usually requires models with multiple dimensions, chaotic trajectories for many outcomes can be expected.

Depending on the measurement properties of y and x and the specifications of the functional relationships in Φ and the stochastic component, Eq. (6.4) leads to various regression models, latent variable models, multilevel models, structural equation models–in other words, to all of the various classes of statistical models that social scientists use to model the complexity of the outcomes they study.

Scalability Within Nonlinear Dynamical Models

Nonlinear dynamical models provide a natural cognitive tool for thinking about social phenomena. The idea that the underlying dynamics of societies are governed by an infinite array of nonlinear differential equations that determine the temporal trajectories of a corresponding infinite array of endogenous variables–with all of the possibilities for chaotic, bifurcating, catastrophic behaviors that this conveys–is appealing.

On the surface, such a representation would appear to make impossible the scaling of composite measures across levels of analysis. That is, under these circumstances, the search for any correspondence between, say, trends in a composite indicator at the macro-societal level with trends at regional levels would appear doomed to failure. However, nonlinear dynamical models have a saving grace, that is, a property termed *self-similarity* – at all scales of measurement, the patterns that their dynamical trajectories (termed fractals because they can be highly irregular) trace out have the same or nearly the same degree of irregularity – which saves the day for scalability.

The strongest form of this property is *linear self-similarity*, in which case the subunits of a whole object (in the present case, a whole society) are exactly like the whole. Most social forms are not strictly linearly self-similar. They possess some degree of self-similarity across scales of analysis, but also some degree of uniqueness, as their fractals describe processes that may be chaotic or even totally random. However, the concept of self-similarity can be applied to random fractals, in which case it takes the form of *statistical self-similarity*–a pattern that repeats itself stochastically so that numerical or statistical measures are preserved across scales.

Scalability Within Stochastic/Statistical Models

Stochastic/statistical models also provide a natural cognitive tool for thinking about social phenomena. The idea that we have a limited slice of observations on the complex object we call a society, that is, a limited set of measurements, and that we have limited samples on which we must take averages at discrete times, and, therefore, that we must incorporate uncertainty/ randomness into our models is very appealing. Indeed, it is a cognitive tool idea that is such a part of social science culture that we use it almost unconsciously.

Within the stochastic/statistical models cognitive framework, scalability can be conceived as *an averaging process* in the sense that:

- a composite index at the societal level is an average at that level of analysis – conceived of as a statistic, that is the value of a random variable, and
- the values of the composite index at sub-societal levels of analysis also are averages/ random variables at those levels.

Scalability then can be analyzed as a relationship among averages–that is, among a family of random variables.

Case Study – The Child and Youth Well-Being (CWI) Index

In 1998, Dr. Ruby Takanishi, President of the Foundation for Child Development in New York City, posed to us the question: "We now have many indicators (literally dozens) of what is happening to kids in the United States, but we do not have a sense of whether things are getting better or worse overall. Can you do anything about this?" In response to this question, we began to develop the CWI. Specifically, we built the CWI on the foundations of an intersection of: (1) numerous databases of social indicator time series, many of which were initiated since the early-1970s, and (2) findings from subjective well-being studies.[2]

The CWI is a composite measure of levels and trends over time in the quality of life, or well-being, of America's children and young people. It consists of several interrelated summary or composite indices of annual time series of 28 social indicators of well-being. The principal objective of the CWI is to give a sense of the overall direction of change in the well-being of children and youth in the United States as compared to base years such as 1975 and 1995.

The CWI is designed to address the following types of questions: Overall, on average, how did child and youth well-being in the United States change in the last quarter of the twentieth century and into the present? Did it improve or deteriorate, and by how much? In which domains or areas of social life? For specific age groups? For particular race/ethnic groups? For each of the sexes? And did race/ethnic group and sex disparities increase or decrease?

The CWI is constructed as follows. First, annual time series data (from vital statistics and sample surveys) have been assembled on some 28 national level Key Indicators in seven Quality-Of-Life (QOL) Domains – a complete list of the Domains and Key Indicators is given in Table 6.1. These seven QOL Domains have been well-established in over three decades of empirical studies of subjective well-being, including studies of children and adolescents, by social psychologists and other social scientists. In this sense*, the CWI is an evidence-based measure of*

[2]Some prior publications on the CWI include Land et al. (2001, 2007), and the chapters in Land (2012). Annual reports of the CWI are posted on: http://www.soc.duke.edu/~cwi/

Table 6.1 Twenty-eight key national indicators of child and youth well-being

Family economic well-Being domain
Poverty rate (all families with children)
Secure parental employment rate
Median annual income (all families with children)
Rate of children with health insurance
Health domain
Infant mortality rate
Low birth weight rate
Mortality rate (ages 1–19)
Rate of children with very good or excellent health (as reported by parents)
Rate of children with activity limitations (as reported by parents)
Rate of overweight children and adolescents (ages 6–19)
Safety/behavioral domain
Teenage birth rate (ages 10–17)
Rate of violent crime victimization (ages 12–19)
Rate of violent crime offenders (ages 12–17)
Rate of cigarette smoking (Grade 12)
Rate of alcohol drinking (Grade 12)
Rate of illicit drug use (Grade 12)
Educational attainment domain
Reading test scores (ages 9, 13, and 17)
Mathematics test scores (ages 9, 13, and 17)
Community connectedness
Rate of persons who have received a high School diploma (ages 18–24)
Rate of youths not working and not in School (ages 16–19)
Rate of pre-kindergarten enrollment (ages 3–4)
Rate of persons who have received a Bachelor's degree (ages 25–29)
Rate of voting in presidential elections (ages 18–20)
Social relationships domain
Rate of children in families headed by a single parent
Rate of children who have moved within the last year (ages 1–18)
Emotional/spiritual well-Being domain
Suicide rate (ages 10–19)
Rate of weekly religious attendance (Grade 12)
Percent who report religion as Being very important (Grade 12)

Note: Unless otherwise noted, indicators refer to children ages 0–17 at last birthday

trends in averages of the social conditions encountered by children and youths in the United States.

Annual values of each of the 28 Key Indicators then are indexed by percentage change from their values in a base year such as 1975.[3] That is, subsequent annual

[3]Three indicators begin in the mid-1980s and use corresponding base years.

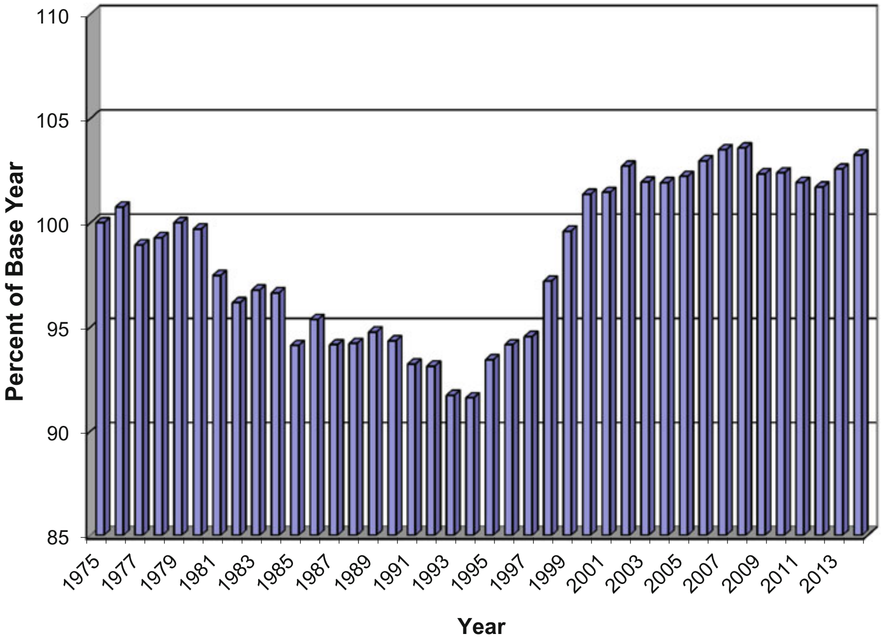


Fig. 6.1 Child well-being index, 1975–2014 (Source: Land 2015a)

observations are computed as percentages of the base year. The base year is assigned a value of 100. The directions of the indicator values are oriented such that a value greater (lesser) than 100 in subsequent years means the social condition measured has improved (deteriorated). The time series of the 28 Key Indicators are grouped together into the seven QOL Domains (see Table 6.1) and domain-specific summary well-being indices are constructed. Within these summary indices, each indicator is equally weighted (more about this later). The seven domain indices are then combined into the equally-weighted composite Child and Youth Well-being Index (CWI). The charts displayed in Figs. 6.1 and 6.2 respectively show changes over time from the base year 1975–2014 in the overall, composite Child and Youth Well-Being Index and its QOL Domain-Specific Indices.[4]

Figure 6.1 identifies a long "recession" in child and youth well-being in the years 1980–1994 followed by a rapid "recovery" in the years 1995–2002 and then a period of oscillations up and down in the years 2003–2013. In brief, just as the CWI allowed us be the first to signal that the steady increases in numerous Key Indicators in the period 1994–2002 were indicative not just of isolated trends (Land et al. 2001, 2007), but rather of an overall improvement in well-being, the CWI more recently signaled that this trend of overall improvement slowed and showed the imprint of macroeconomic expansions and contractions (Land 2012).

[4]The year 1975 is the earliest year for which statistical data on 25 of the 28 Key Indicators are available.

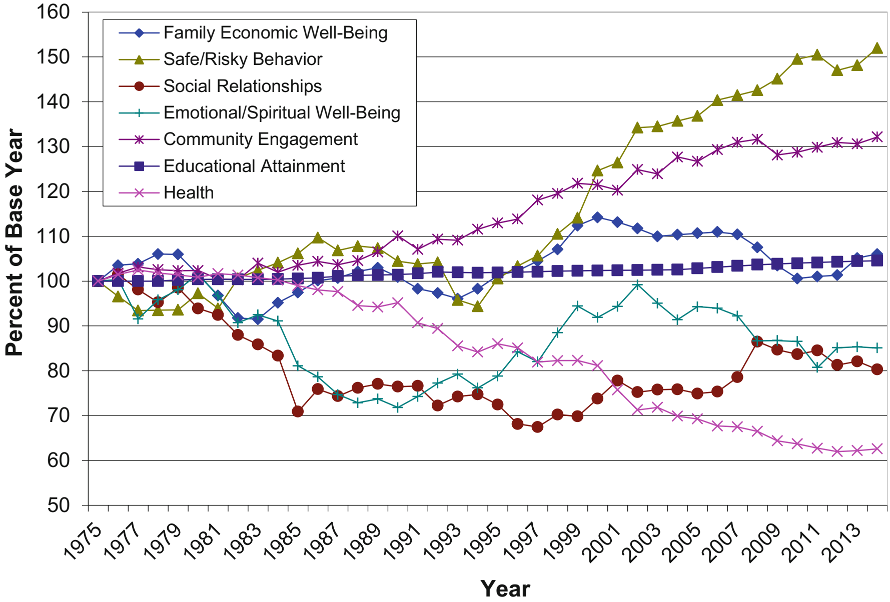


Fig. 6.2 Domain-specific summary indices, 1975–2014 (Source: Land 2015a)

Figure 6.2 helps us to identify the components of changes over time in child and youth well-being, specifically: two QOL Domains that show substantial improvements (safety/behavioral concerns and community engagement), two that declined and then stabilized or slightly improved (health and social relationships), two that oscillate (family economic well-being and, with longer cycles, emotional/spiritual well-being), and a domain shows a slight long-term improvement (educational attainment).

So, What Have We Learned? Five Lessons

Lesson 1 First, the subjective well-being research literature–which essentially lets individuals, including children and youths talk about their lives and what makes them feel happy, satisfied, and fulfilled with them–provides a good foundation for organizing indicators of child and youth well-being and indices of changes therein and can be used to evaluate the external validity of the CWI.

Evidence in support of this statement is shown in Fig. 6.3 which compares trends in the CWI with those of smoothed data on overall life satisfaction for High School Seniors from the Monitoring the Future Study (MTF) for the years 1976–2013. The MTF question, administered annually to 12th graders since 1975, is of the conventional global satisfaction with life form: "How satisfied are you with your life as a whole these days?" The answer range is a seven-point Likert rating scale:

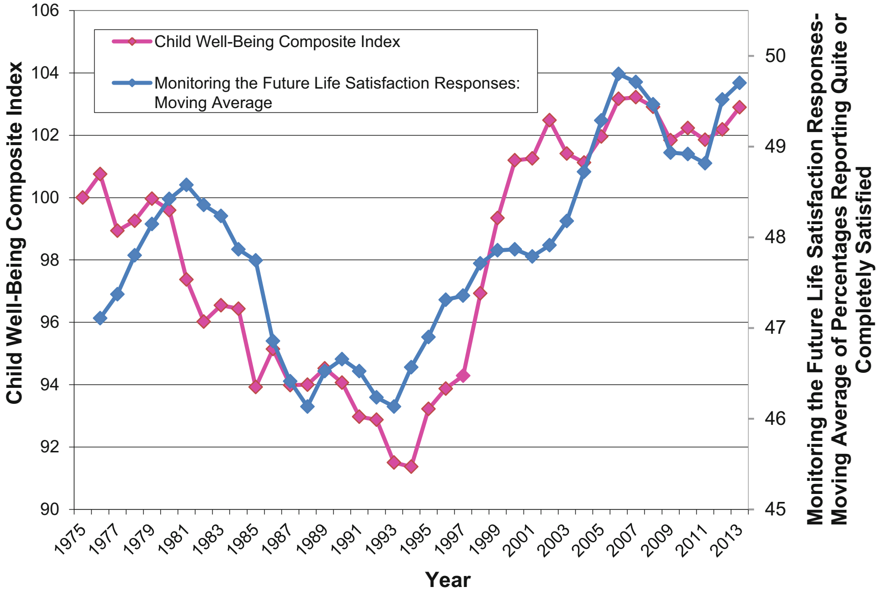


Fig. 6.3 CWI and smoothed MTF life satisfaction trend, 1976–2013 (Source: Land 2015a)

Completely Dissatisfied, Quite Dissatisfied, Somewhat Dissatisfied, Neither Satisfied or Dissatisfied, Somewhat Satisfied, Quite Satisfied, and Completely Satisfied. For comparisons with the CWI, we combined the last two response categories to calculate the percent of the 12th graders who respond that they either are Quite or Completely Satisfied in each year from 1975 to 2003. Because the annual MTF data are based on samples and the annual CWI is based on averages of numerous population and statistical averages, the latter varies more smoothly from year to year. Accordingly, in order to smooth out the MTF series to show its primary temporal trends, we applied three-point moving averages to the series three times.

Figure 6.3 shows a striking similarity of trends over time of the CWI – which is based on objective statistical time series of social indicators – and the only continuous empirical data on trends in the subjective well-being of children in American society across the past four decades (the correlation of the two series is 0.86). In other words, the CWI passes this external validity criterion as an indicator of trends in child and youth well-being in the U.S. In addition, Fig. 6.3 suggests that turning points and trends (ups and downs) in the CWI slightly lead the smoothed MTF life satisfaction data series.

Lesson 2 The equal-weighting strategy for calculating composite indices used by the CWI is surprisingly robust and has nice statistical properties. Specifically, in the context of a mathematical model of heterogeneous weighting schemes (corresponding to different values or preferences) among members of a population, Hagerty and Land (2007) showed that the equal-weighting strategy has a *minimax*

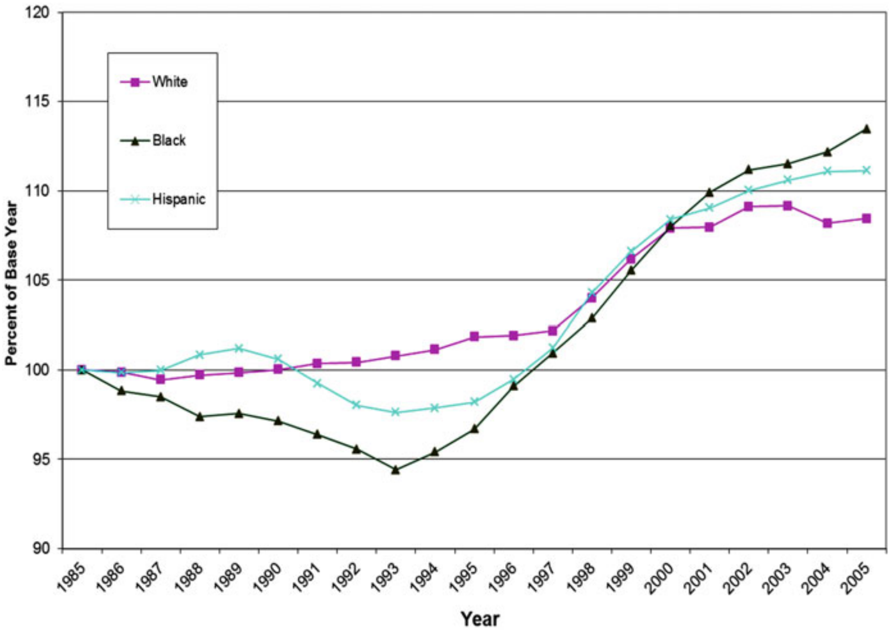


Fig. 6.4 Race/ethnic group-specific summary indices of child and youth well-being, 1985–2005 (Source: Land et al. 2012)

statistical property–equal weights minimize extreme disagreements among individuals (experts, children, parents) with respect to composite indices.

Lesson 3 Comparisons of levels and trends in child and youth well-being among major race/ethnic subpopulations can be made with the CWI. Figure 6.4 displays trends in CWI values for White, Hispanic, and Black subpopulations of children and youths. These trends show that the long-term recession in well-being cited above in Lesson 1 was particularly severe for Hispanic and Black children and youths and the subsequent long-term recovery was particularly strong.

Lesson 4 The CWI can be expanded to include additional Key Indicators in its seven QOL Domains. An example is the Expanded CWI which has 44 Indicators. The Expanded CWI uses new indicator data series that commenced in the U.S. in the mid-1990s and thus uses 1995 as a base year. The Expanded CWI also allows for the disaggregation of the CWI and comparison of trends therein by more specific age groups corresponding to infancy/early childhood (ages 0–5), middle childhood (ages 6–11), and adolescence (ages 12–17) as shown in Fig. 6.5. This figure shows that, for some time periods, improvements in the well-being of the two younger age groups tend to lead those in the adolescent age group, which is consistent with aging of the birth cohorts of children and indicative that improvements in well-being at younger ages tend to carry through to the later ages.

Lesson 5 The CWI also can be scaled to various sub-national units of analysis. For example, for the years 1995–2004, Fig. 6.6 displays trends in composite CWIs for

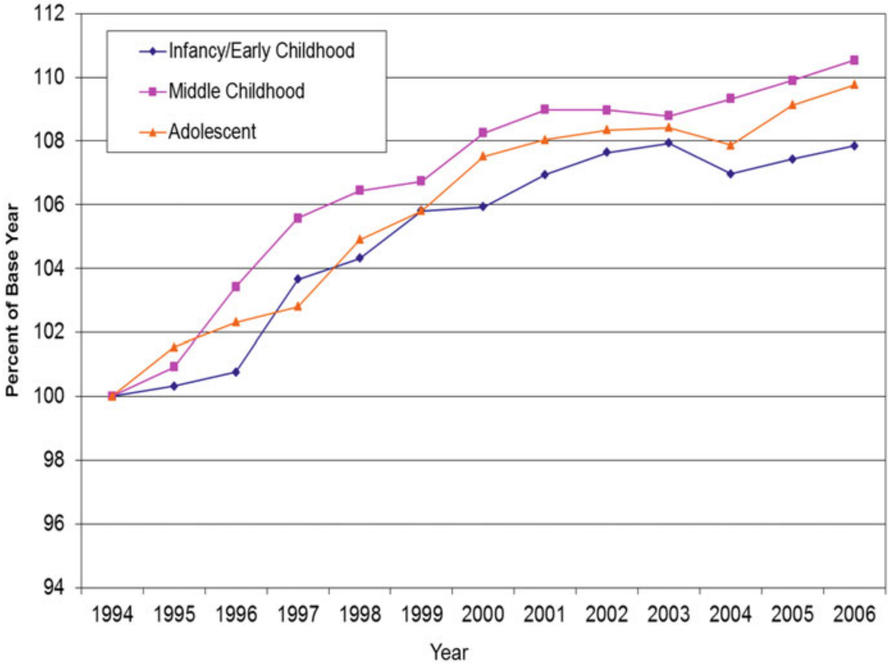


Fig. 6.5 Infancy/early childhood, middle childhood, and adolescent composite well-being Indices, 1994–2006 (Source: Land 2008)

the U.S. as a whole, the state of California as a whole, and the San Francisco Bay Area of California from Lee et al. (2009). The overall, composite U.S. index is the equally weighted average of 28 indicators shown above in Fig. 6.1. The six-component averages for the U.S., the state of California, and the Bay Area are based on six indicators that closely match to the national CWI (solid lines in Fig. 6.6): children poverty, infant mortality rate, infants born at low birth weight, child/youth death rate, teen birth rate, and youth suicide rate. The figure also includes other composite CWIs for California and the Bay Area based on 16 Key Indicators (dashed lines in Fig. 6.6). It can be seen that the trends over time are comparable for all of these composites, with the indices for both the state of California and the Bay Area showing greater increases/improvements in well-being than the U.S. as a whole for this time period.

Implications of CWI Findings for Scalability Models

Based on the foregoing findings regarding scalability of the CWI across population categories – race/ethnicity, child developmental stages, and geographical levels of analysis, can we adjudicate between the two complexity models – nonlinear deterministic dynamics and stochastic/statistical representations – exposited

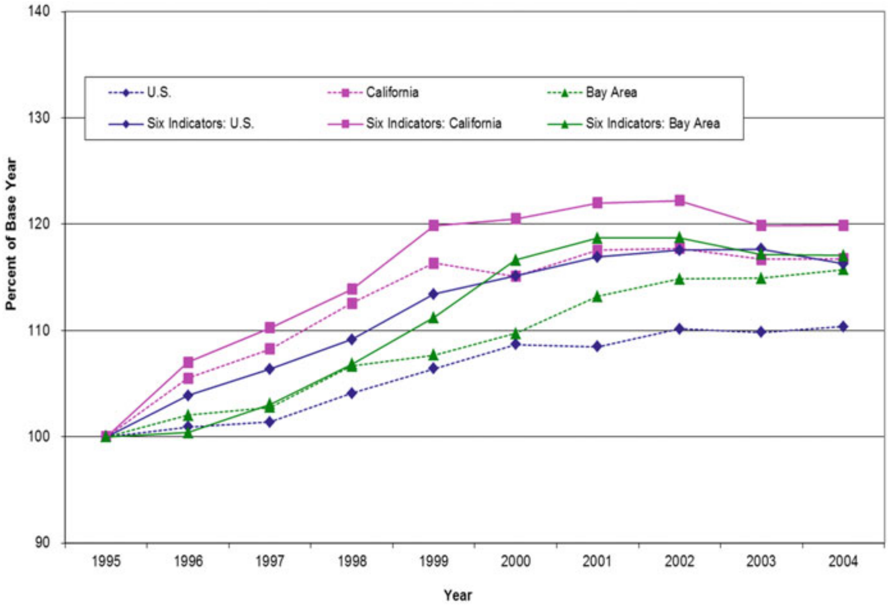


Fig. 6.6 Comparison of child and youth well-being index, U.S., California, and Bay Area, 1995–2004 (Source: Lee et al. 2009)

above? That is, using the CWI as an example, is the scalability of composite social indicators more likely due to one of these complexity models or the other?

Our response is that both models appear to be applicable and useful. To begin with, the CWI composite index and its well-being domain-specific indices are based averages of population and sample statistics, that is, random variables, in time series format measured at some level of population aggregation from local to national. As such, scalability questions pertain to the relationships among these averages, that is, to questions of statistical analysis of families of random variables. For example, in time-period/cross-sectional comparisons of Index values across population averages, the Index values for the U.S. as a whole are averages across the entire population. These may vary from the averages for specific sub-populations (e.g., race-ethnic groups, child developmental stages, or regional populations) as shown in the previous review of CWI findings, but the averages across all sub-populations will equal those for the whole population. Thus, the utility of the stochastic/statistical framework as a cognitive tool for studying questions of scalability of composite social indicators is evident.

In addition, however, the CWI time series has properties consistent with an underlying nonlinear deterministic dynamic generating model. For instance, one method for calculating the fractal dimension of a time series (Higuchi 1988) requires that the slope of the logarithm of a measure, $L(k)$, of the "length" of curves for subsets of a time series plotted against $Log\ k$ for a set of time intervals $k = 2,$

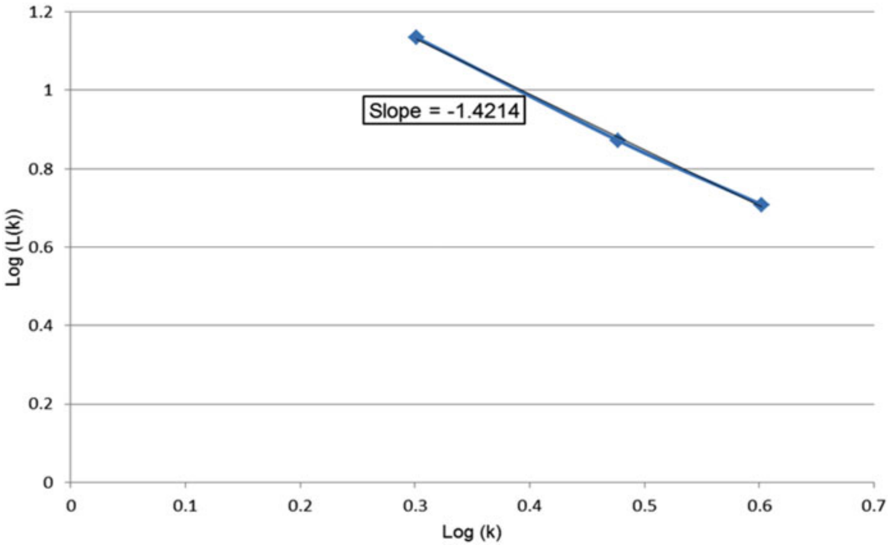


Fig. 6.7 Measure of fractal dimension of CWI time series, 1975–2014

3, … be negative and linear if the time series is fractal (see the Appendix for a definition of *L(k)*). In addition, if the slope is linear, its numerical value estimates the fractal (or Hausdorff) dimension *D* of the underlying time series, where "fractal dimension" measures the degree of roughness of a fractal shape and varies from one, indicating a smooth one-dimensional line, to two, indicating a highly irregular curve.

Applied to the CWI national time series shown above in Fig. 6.1, the graphical plot of *Log L(k)* and *Log (k)* is displayed in Fig. 6.7. This graphical plot is linear with a slope of −1.42. In other words, this application of the Higuchi (1988) index to the national-level CWI time series supports the inference that this series is fractal with dimension 1.42. This is indicative of a moderate level of irregularity in the series. It also is indicative of a moderate level of self-similarity across scales of analysis, which is consistent with the similar trends shown earlier for various population groups and geographical levels of analysis.

To interpret these findings, consider again the qualitative patterns of changes over time in the CWI as shown in Fig. 6.1. At an overall "macro" time scale, the temporal pattern of changes in the CWI over the 40 years of the graph in Fig. 6.1 shows a substantial long-term decline beginning in the early 1980s through 1994 and then a long-term recovery to levels at or above those of the earliest years of the chart. And embedded within this overall time trend are several "self-similar" periods of declines and increases in the CWI that are evident at smaller time scales and that have degrees of irregularity similar to that of the overall macro time scale – this is the fundamental property of fractals.

With respect to underlying generative dynamical models for the CWI time series, how should the estimated fractal dimension of 1.42 be interpreted? One possibility is that the underlying generative mechanism for the CWI may be a nonlinear dynamic model – like the nonlinear deterministic differential equation system of the meteorological model for atmospheric convection given earlier in Eq. (6.2). Another possibility: Recall that the Wiener process (Cox and Miller 1965, pp. 205–213) with drift μ is defined by:

$$X_t = \mu t + \sigma W_t$$

where X is the numerical values of a time series at time t, μ (the drift coefficient) and σ are real numbers, and W is a normal distribution of changes from time period to period around this trend in average value of the series. Substantively, the path of a time series governed by a Wiener process drifts up or down from time period to time period depending on the algebraic sign of μ (the expected value of the series) with variations from time period to time period around the drifting expected value that are distributed according to a normal frequency distribution. And the path of time series values generated by a one dimensional Wiener process is a fractal curve of dimension 1.5, which is close to the fractal dimension 1.42 estimated for the CWI time series. Applied to the CWI time series of Fig. 6.1, the drift coefficient μ would have a positive sign corresponding to the slight upward/improving trends in the CWI and indicative of an improvement in the overall quality-of-life/well-being of America's children and youth across the last four decades – around which the year-to-year changes are distributed as the normally distributed W part of the process. In brief, the CWI time series could be generated either by a nonlinear deterministic differential equation system, or by a stochastic/random time series process, or, more likely, by a combination of the two.

Conclusions

In sum, models as linguistic devices help us to organize our observations of societal phenomena and trends over time in the form of composite social indicators. Nonlinear dynamic complexity models may help us to understand the scalability of composite indicators due to underlying fractal properties. Stochastic/statistical complexity models also contribute to our understanding of composite indicators as members of families of random variables in the form of averages. Our study of these complexity models and their application to composite indices has focused on the Child and Youth Well-Being Index as a specific case-in-point. As with other composite well-being indices, the CWI and its well-being domain-specific indices are based on averages of population and sample statistics, that is, random variables, in a time series format measured at some level of population aggregation from local to national. As such, scalability questions pertain to the relationships among these averages, that is, to questions of statistical analysis of families of random variables. In addition, however, the CWI time series is shown to have "fractal" or "self-similar" properties consistent with an underlying nonlinear deterministic dynamic generating model.

Appendix

The Higuchi (1988) index for the degree of irregularity of a time series is defined as follows.

Denote a finite set of time series observations taken at a regular interval as:

$$X(1), X(2), X(3), \ldots, X(N).$$

From this referent time series, construct a new series, *X(m,k)* as follows:

$$X(m), X(m+k), X(m+2k), \ldots, X(m+[(N\text{-}m)/k]k) \ (m = 1, 2, \ldots k),$$

where [] denotes the last time series element not exceeding the index in brackets. Next, define the following measure for the length of the curve *X(m,k)*:

$$L_m(k) = \left\{ \frac{\left(\sum_{i=1}^{\left[\frac{N-m}{k}\right]} \left| X(m+ik) - X(m+(i-1)\cdot k) \right) \right| \frac{N-1}{\left[\frac{M-m}{k}\right]\cdot k} \right\}}{k} \right.$$

The *length of the curve for the time interval k, L(k)*, then is defined as the average over the *k* sets of $L_m(k)$. If *L(k)* is proportional to k^{-D}, then the curve is *fractal with dimension D*. To assess this, if *L(k)* is plotted against *k* on a doubly logarithmic scale, the points should fall on a straight line.

References

Casti, J. L. (1992a). *Reality rules I: Picturing the world in mathematics – The fundamentals*. New York: Wiley.

Casti, J. L. (1992b). *Reality rules II: Picturing the world in mathematics – The frontier*. New York: Wiley.

Cox, D. R., & Miller, H. D. (1965). *The theory of stochastic processes*. New York: Wiley.

Gleick, J. (1987). *Chaos: Making a new science*. New York: Viking Penguin.

Gorski, P. S. (2004). The poverty of deductivism: A constructive realist model of sociological explanations. *Sociological Methodology, 34*, 1–34.

Hagerty, M. R., & Land, K. C. (2007). Constructing summary indices of quality of life: A model for the effect of heterogeneous importance weights. *Sociological Methods and Research, 35* (May), 455–496.

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D, 31*, 277–283.

Land, K. C. (2001). Models and indicators. *Social Forces, 80*(December), 381–410.
Land, K. C. (2008). *Special focus report: Trends in infancy/early childhood and middle childhood well-being, 1994–2006*. New York: Foundation for Child Development.
Land, K. C. (Ed.). (2012). *The well-being of America's children: Developing and improving the child and youth well-being index*. New York: Springer.
Land, K. C. (2015a). *Duke child and youth well-being index (CWI) report 2015*. Durham: Duke University Center for Child and Family Policy.
Land, K. C. (2015b). Social indicators. In G. Ritzer (Ed.), *Wiley Blackwell encyclopedia of Sociology* (2nd ed.). Boston: Blackwell Publishing Company.
Land, K. C., Lamb, V. L., & Mustillo, S. K. (2001). Child and youth well-being in the United States, 1975–1998: Some findings from a new index. *Social Indicators Research, 56* (December), 241–320.
Land, K. C., Lamb, V. L., Meadows, S. O., & Taylor, A. (2007). Measuring trends in child well-being: An evidence-based approach. *Social Indicators Research, 80*, 105–132.
Land, K. C., Lamb, V. L., Meadows, S., Zheng, H., & Fu, Q. (2012). The CWI and its components: Empirical studies and findings. In K. C. Land (Ed.), *The well-being of America's children: Developing and improving the child and youth well-being index*. New York: Springer.
Lee, J., Lamb, V. L., & Land, K. C. (2009). Composite indices of changes in child and youth well-being in the San Francisco Bay Area and the state of California, 1995–2005. *Child Indicators Research, 2*(December), 353–374.
Lorentz, E. (1963). Deterministic non-periodic flows. *Journal of Atmospheric Science, 20*, 130–141.
Suppe, F. (1989). *The semantic conception of theories and scientific realism*. Urbana: University of Illinois Press.
Toynbee, A. J. (1961). *A study of history. 12 volumes*. New York: Oxford University Press.

Part III
Technical Issues

Chapter 7
Synthesis of Indicators: The Composite Indicators Approach

Matteo Mazziotta and Adriano Pareto

Why the Application of a Composite Indicators Is So Important?

In recent years, the debate on the measurement of multidimensional phenomena has caused, within the worldwide scientific Community of developed countries, a renewed interest. It is common awareness that a number of socio-economic phenomena cannot be measured by a single descriptive indicator and that, instead, they should be represented with a multiplicity of aspects or dimensions. Phenomena such as development, progress, poverty, social inequality, well-being, quality of life, etc., require, to be measured, the 'combination' of different dimensions, to be considered together as components of the phenomenon (Mazziotta and Pareto 2013). In fact, the complex and multidimensional nature of these phenomena requires the definition of intermediate objectives whose achievement can be observed and measured by individual indicators. The mathematical combination (or aggregation as it is termed) of a set of indicators that represent the different dimensions of a phenomenon to be measured can be obtained by applying methodologies known as *composite indicators* or *composite index* (Saisana and Tarantola 2002; Salzman 2003; OECD 2008).

As is known, building a composite index is a delicate task and full of pitfalls: from the obstacles regarding the availability of data and the choice of individual indicators, to their treatment in order to compare (normalization) and aggregate them (weighting and aggregation). Despite the problems mentioned, the composite indices are widely used by several international organizations for measuring

M. Mazziotta • A. Pareto (✉)
Italian National Institute of Statistics, Viale Oceano Pacifico 171, Rome, Italy
e-mail: mazziott@istat.it; pareto@istat.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_7

economic, environmental and social phenomena and, therefore, they provide an extremely relevant tool and in the course of evolution (OECD 2008).

Many scientists dispute the use of composite indices that lead to the determination of a single value for each geographic area, preferring the so-called dashboard (as in the case of monitoring the state of health of a vehicle: oil level, gasoline, water temperature, etc.). In the case of dashboard, it is possible to identify various dimensions of the phenomenon, all relevant, without which, they are further aggregated. From the statistical point of view, it is an incontrovertible choice but from the standpoint of political and media is a heavy limitation. The easy-disclosure in the media and the immediate understanding by the user are certainly the strengths of a unique index. Obviously, both approaches have strengths and weaknesses. The dashboard manages complexity not using synthetic measures so that certainly it defects from the communication point of view. In this case the question without answer is: "Is well-being increased or decreased?". The composite index manages also the complexity but it reduces the dimensions in space with an evident loss of information; however, the composite indicator allows a single measure that is more communicative. A composite indicator, before the theoretical and methodological aspects, has a problem: is it possible to measure well-being with a formula? The answer is probably yes if a paradigm of work is strictly respected (see Sects. 7.2 and 7.3). In literature, for example, many attempts to measure well-being do not respect a paradigm of work and arrive to unreliable and questionable conclusions. This aspect causes the failure of many alternative measures to GDP.

Mission "Replace GDP"

The debate presented above has convinced scientists that the economic measure for excellence (GDP) is not able to represent the well-being or the progress of a society, much less to express the quality of life of a geographical area or a community. This debate has produced worldwide a considerable literature with more than a 100 alternative indices, published by government organizations (and others), academia and business press, but despite this, it seems that the popularity of GDP has not been minimally scratched.

In fact, the GDP is based on very solid theoretical bases, while many alternative indices are guilty of clarity from the stage of definition of the phenomenon; in many circumstances, not having a shared socio-economic theory behind, taking into account dozens of indicators so that all possible aspects are considered (and then no one).

The publication, in September 2009 of the report by the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz Commission), set up by the former French President Nicolas Sarkozy, was crucial for developing several studies about "Beyond GDP" scenarios. The Commission's aim has been to identify the limits of GDP as an indicator of economic performance and societal progress, to consider what additional information might be required for the production of more relevant indicators of social progress, to assess the feasibility of alternative measurement tools, and to discuss how to present the statistical information in an appropriate way (Giovannini and Rondinella 2012).

In truth, even before the Stiglitz Commission, several attempts to measure phenomena ‘close to’ well-being (progress, quality of life, happiness, etc.) have been made and published from scientists and prestigious institutions. These attempts can be divided in four groups and some of these studies are presented below. The first proposes to adjust the GDP: MEW – Measure of Economic Well-being (Nordaus and Tobin 1972), ISEW – Index of Sustainable Economic Welfare (Daly and Cobb 1989), GPI – Genuine Progress Indicator. Another group of studies takes into account aspects such as social and environmental activities or directly the level of (perceived) satisfaction of individuals: Happy Life Expectancy (Veenhoven 1996) and Happy Planet Index. A third group includes measures that represent a composite indicator including GDP: the best known among these are HDI – Human Development Index (UNDP 1990, 2001, 2010), BLI – Better Life Index (OECD 2011) and GNH – Gross National Happiness. Finally, the fourth approach argues that it is preferable to measure different dimensions with a set of indicators (dashboard) rather than to get a single synthetic measure (Rinaldi and Zelli 2014), for example the Millennium Development Goals (MDGs) by UNDP. For a detailed review of composite indicators see Bandura (2008).

In the Italian panorama, the first report on “Equitable and Sustainable Well-being” (BES) by the Committee composed by Istat (Italian National Institute of Statistics) and CNEL (Italian Council for Economics and Labour) was published in March 2013. It consists in a dashboard of 134 individual indicators divided in 12 domains. The third BES report, published in December 2015, presents a composite indicator for each domain of well-being (Istat 2015). Also in Italy, since 2003, the “Campaign Sbilanciamoci!” has published the Index of the Regional Quality of Development (QUARS) with the aim of providing a multidimensional measure of the development of Italian regions, based on 41 individual indicators divided in seven domains and synthesized by a simple arithmetic mean (Gnesi et al. 2010). One of the indices with greater media coverage in Italy is the measure of the Quality of Life (QoL) which, every year, the economic newspaper “Il Sole 24ore” publishes at the provincial level. It is based on 36 individual indicators divided in six domains and synthesized by a simple arithmetic mean.

The Use (Good and Bad) of the Composite Indicators

The construction of a composite indicator is a good solution but a paradigm of work must be strictly followed. It is a complex task whose phases involve several alternatives and possibilities that affect the quality and reliability of the results. The main problems, in this approach, concern the choice of theoretical framework, the availability of the data (in space and over time), the selection of the more representative indicators and their treatment in order to compare and aggregate them (Fanchette 1974).

The paradigm of work is based on the following steps (OECD 2008; Mazziotta and Pareto 2012):

1. Defining the phenomenon to be measured. The definition of the concept should give a clear sense of what is being measured by the composite index. It should refer to a theoretical framework, linking various sub-groups and underlying indicators.
2. Selecting a group of individual indicators. Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility, etc. (Maggino 2014). It is necessary to consider that socio-economic phenomena, as well-being, follow a *formative approach* according to which the latent factor (well-being) depends on the indicators that 'explain' it and not vice versa (Diamantopoulos et al. 2008).
3. Normalizing the individual indicators. This step aims to make the indicators comparable and to define the polarity. Normalization is required prior to any data aggregation as the indicators in a data set often have different measurement units. We want to normalize the indicators so that an increase in the normalized indicators corresponds to an increase in composite index.
4. Aggregating the normalized indicators. It is the combination of all components to form one or more composite indices (mathematical functions). Different aggregation methods are possible and the choice must be conditioned by the nature of the indicators into the *formative approach*.
5. Validating the composite index. This step aims to assess the robustness of the composite index, in terms of capacity to produce correct and stable measures, and its discriminant capacity.

It is important to emphasize that the theoretical part is not separate from the statistical-methodological one: then, the choice of the individual indicators is not independent from the choice of the aggregation method. Unfortunately, many methods in the literature do not comply with this restriction and, for example, they use the factor analysis as a method of synthesis into a *formative model* (see Sect. 7.2).

No universal method exists for composite indices construction. In each case their construction is much determined by the particular application, including formal elements and incorporates some expert knowledge on the phenomenon. Nevertheless, the advantages of composite indices are clear, and they can be summarized in unidimensional measurement of the phenomenon, easy interpretation with respect to a battery of many individual indicators and simplification of the data analysis.

A basic rule to keep in mind is 'garbage in garbage out' that is, if the original matrix contains garbage then the composite index produces garbage. If a phenomenon is poorly defined, then he will certainly be poorly measured. Despite this, the reverse is not true. If the phenomenon is well defined and the matrix is composed of elementary indicators of good quality, then it is not always true that the composite index is valid. It depends on the statistical methodology used which must be 'well-matched' with the theoretical framework on which is based the phenomenon to be measured.

The 'Perfect' Composite Indicator Does Not Exist (Properties, Criteria of Comparison, etc.)

As mentioned in Sect. 7.1.2, no universal method exists for composite indices construction. The best composite indicator is the one that respects the objectives required by the researcher or the commitment. The paradigm of work requires that some questions are asked before starting work. The responses influence the path to be followed in order to obtain the best possible solution of composite index. All the answers can influence both the choice of the individual indicators and the methodology to normalize and synthesize them.

- Do you need territorial comparisons? If yes, the individual indicators chosen must be available for the required territorial disaggregation and this may affect the use or not of some measures.
- Do you need comparisons over time? If yes, the individual indicators chosen must be available for the required time series and above all only some normalization methods allow performing effectively and correctly by statistical points of view comparisons over time between composite indices.
- Are the individual indicators non-substitutable? Or, is the compensation between the indicators admitted? Usually, in the measurement of socio-economic phenomena, the formative approach is required and then the compensation is not admitted. Therefore, if the individual indicators are non-substitutable the choice of the aggregation method must be based on this factor. In this case, the arithmetic mean and the linear models are not eligible. For example, the HDI and the Human Poverty Index - HPI are characterized by indicators non-substitutable and the aggregation methods (power mean, respectively, of order 0 and 3) do not allow compensation between them.
- What is the audience to which the analysis is targeted? The client and recipient of the composite indicator should influence the choice of the statistical synthesis method of the individual indicators. The simplicity of calculation, the immediate use and easy interpretation of output results are conditions essential when the study is addressed to a broad audience not accustomed to technicalities: the reader should immediately understand both the methodology used and the meaning of the obtained results. If the study is addressed towards an academic audience then the methodology can certainly be more complex and the results have 'shades of reading'. In all cases, the transparency of method and calculation must be respected because otherwise the composite indicator is a fraud.
- Is the method robust? The first rule for constructing a good composite index is the compliance with the aims of the study. However other rules must be respected: the index should be robust i.e. it must incorporate the changes but

not be too influenced by outliers. The method must be stable (but not too much) to the variations of the input matrix. It is very important to choose the most robust method through sensitivity analysis (influence analysis or others similar techniques).

The answers to these questions should guide the research toward the most effective method for reducing the multidimensionality of the phenomenon. It is not possible to ignore either one of these questions because the risk of altering the reality is very high.

In particular, the attention has to be focused on the search of the most suitable method depending on the following factors: type of indicators (substitutable/non-substitutable), type of aggregation (simple/complex), type of comparisons to be made (relative/absolute), type of weights of the indicators (subjective /objective) as described in Sect. 7.3.

The Steps Characterizing the Composite Indicators Construction

We have seen that the main steps for constructing a composite index are the following (Salzman 2003; OECD 2008; Mazziotta and Pareto 2012)[1]: (1) Defining the phenomenon to be measured, (2) Selecting a group of individual indicators, (3) Normalizing the individual indicators, (4) Aggregating the normalized indicators, and (5) Validating the composite index.

The Definition of the Phenomenon

The definition of the phenomenon should give a clear sense of what is being measured by the composite index. It should refer to a theoretical framework, linking various sub-groups and underlying indicators. A fundamental issue, often overlooked in composite index construction, is the identification of the model measurement, in order to specify the relationship between the phenomenon to be measured (latent variable) and its measures (individual indicators). In this respect, if causality is from the phenomenon to the indicators we have a *reflective* measurement model; if causality is from the indicators to the concept we have a *formative* model (Diamantopoulos et al. 2008).

The reflective measurement model is most widely used in psychological and management sciences. Typical examples of reflective scenarios include measures of

[1]Some authors describe a greater number of steps (e.g., imputation of missing data). We report only the fundamental steps.

intelligence, attitudes and personality that are assessed by eliciting responses to indicators. A fundamental characteristic of reflective models is that a change in the latent variable causes variation in all individual indicators simultaneously.

The formative model is common in economics and sociology. A typical example of formative model is the socioeconomic status (SES), which is defined as a combination of education, income, occupation, and residence. If any one of these indicators increases, SES would increase (even if the other indicators did not change); conversely, if a person's SES increases, this would not necessarily be accompanied by an increase in all four indicators (Diamantopoulos and Winklhofer 2001).

Defining the model of measurement is very important, because it is closely related with the selection and aggregation steps.

The Selection of the Indicators

In this step, the number and nature of the components that will make up part of the composite index need to be determined. Then, the specific indicators employed in estimating each of the component index must be selected. Such selection is generally based on theory, empirical analysis, pragmatism or intuitive appeal (Booysen 2002).

The strengths and weaknesses of a composite index largely derive from the quality of the underlying indicators. Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility, etc. (OECD 2008).

The selection step is the result of a trade-off between possible redundancies caused by overlapping information and the risk of losing information. A statistical approach to the choice of indicators involves calculating the correlation between potential indicators, and including the ones that are less correlated in order to minimize redundancy (Salzman 2003). However, the selection process depends on the measurement model used: in a *reflective* model, all the individual indicators must be intercorrelated; whereas in a *formative* model they can show negative or zero correlations (Diamantopoulos et al. 2008).

The Normalization

Normalization step aims to make the indicators comparable. Normalization is required before any data aggregation as the indicators in a data set often have different measurement units and ranges. In such cases, without normalization, composite indices will be biased towards variables with high ranges (implicit weighting scheme) and meaningful changes in a value may significantly affect the composite index. Therefore, it is necessary to bring the indicators to the same standard, by transforming them into pure, dimensionless, numbers. Another motivation for the normalization is the fact that some indicators may be positively correlated with the phenomenon to be measured (positive polarity), whereas others may be negatively correlated with it (negative polarity). We want to normalize the indicators so that an increase in the normalized indicators corresponds to an increase in the composite index (Salzman 2003).

Formally, we have to move from the data matrix $\mathbf{X} = \{x_{ij}\}$, with n rows (statistical units) and m columns (individual indicators), to the normalized matrix $\mathbf{Y} = \{y_{ij}\}$:

$$\mathbf{X}_{n,m} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nj} & \dots & x_{nm} \end{pmatrix} \Rightarrow \mathbf{Y}_{n,m} = \begin{pmatrix} y_{11} & \dots & y_{1j} & \dots & y_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ y_{i1} & \dots & y_{ij} & \dots & y_{im} \\ \dots & \dots & \dots & \dots & \dots \\ y_{n1} & \dots & y_{nj} & \dots & y_{nm} \end{pmatrix}$$

where x_{ij} is the original value of indicator j for unit i and y_{ij} is the normalized value of indicator j for unit i.

There are various normalization methods, some of which transform the range or variance of the indicators to a common basis and others which emphasizes percentage change. The following classification is here used: *no normalization*, *ranking*, *standardization* (or *Z-scores*), *re-scaling* (or *Min-Max*), *distance from a reference* (or *Indicization*).

The researcher must identify the most suitable normalization method to apply to the problem at hand, taking into account its properties and robustness against possible outliers in the data. Different normalization methods will produce different results for the composite index. Therefore, a robustness analysis should be carried out to assess their impact on the results (Freudenberg 2003).

The Polarity Issue

The polarity of an individual indicator is the sign of the relation between the indicator and the phenomenon to be measured. For example, in the case of development, the 'Life expectancy' has positive polarity, whereas the 'Infant mortality rate' has negative polarity. When a composite index must be constructed, all the individual indicators must have positive polarity, so it is necessary to 'invert' the sign of the indicators with negative polarity. Inversion of polarity may be performed before normalizing or jointly. However, in most of cases the results are identical. There are two basic methods for inverting polarity: (a) linear transformation, and (b) non-linear transformation.

(a) Linear transformation takes the complement with respect to maximum value, as follow:

$$x'_{ij} = \max_{i}\left(x_{ij}\right) - x_{ij} \tag{7.1}$$

where $\max_{i}\left(x_{ij}\right)$ is the maximum of indicator j. This is the simplest technique and it allows to save the same 'distance' between units, with a different origin. It is particularly used with ranking, standardization and re-scaling.

(b) Non-linear transformation takes the reciprocal of the value:

$$x'_{ij} = \frac{1}{x_{ij}} \tag{7.2}$$

This technique is typically used with indicization, but it modifies the 'distances' between units and thus it can be criticized. Furthermore, it requires all values are greater than 0.

Sometimes, polarity of an indicator may be positive below a certain threshold and negative above it or vice versa. For example, in the case of gender parity, the 'Percentage of women elected in Parliament on the total of the elects' has positive polarity below 50% and negative polarity above 50%. We call this the 'Double-polarity question'.

The simplest method for moving from a double-polarity to a standard case (positive or negative polarity) is the triangular transformation.

Triangular transformation has the form:

$$x'_{ij} = \mathrm{abs}\left(\lambda_j\text{-}x_{ij}\right) \tag{7.3}$$

where λ_j is the threshold for indicator j. If the obtained polarity is negative, an additional linear or non-linear transformation is required.

In Fig. 7.1 three examples of linear transformation (a), non-linear transformation (b), and triangular transformation (c) are shown.

No Normalization

The first method, no normalization, involves an aggregation of original data. This may be a good technique if all the indicators have the same unit of measurement and similar ranges or they are expressed as percentages or ratios. Otherwise, aggregating individual indicators without normalization will cause the index to be dominated by implicit weights coming from the units and range used to measure indicators.

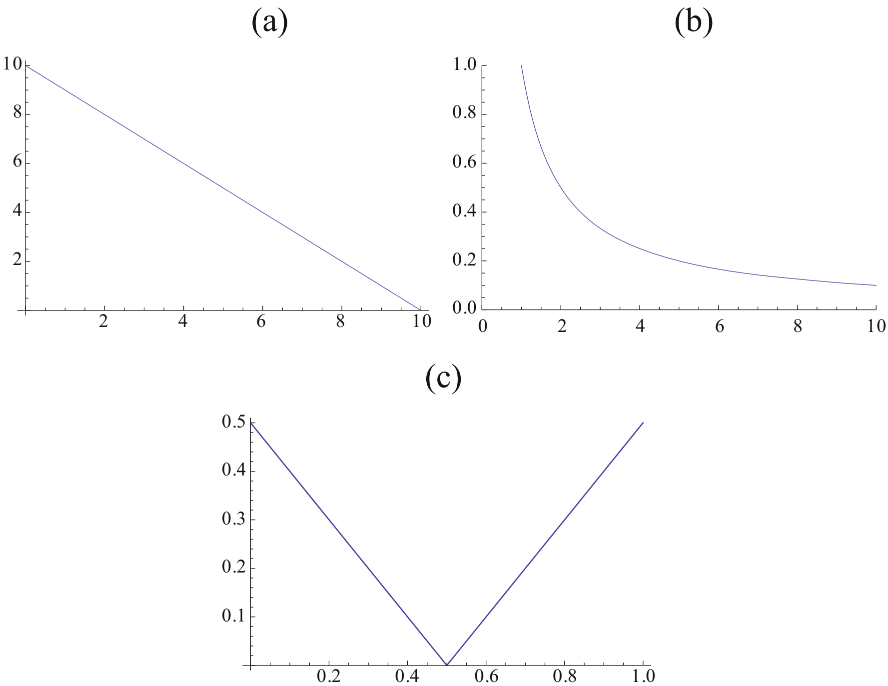


Fig. 7.1 Linear, non-linear and triangular transformation

Ranking

This method simply ranks units for each indicator as follows:

$$y_{ij} = \mathrm{rank}\left(x_{ij}\right) \tag{7.4}$$

where rank(x_{ij}) is the rank of unit i with respect to indicator j. Units with the same value receive a rank equal to the mean of the ranks they span, so that the sum of the ranks is $n(n+1)/2$. If indicator j has negative polarity, the rank order must be reversed. This is equivalent to apply (7.1) or (7.2) and then (7.4). Ranking is based on ordinal levels and it is not affected by outliers. However, differences between the units cannot be evaluated as absolute level information is lost. So, the method allows the performance of units to be followed over time only in terms of relative positions (rankings).

Standardization (or *Z-scores*)

Standardization converts indicators to a common scale with a mean of zero and standard deviation of one. The formula is:

$$y_{ij} = \frac{x_{ij} - \mathrm{M}_{x_j}}{\mathrm{S}_{x_j}} \tag{7.5}$$

where M_{x_j} and S_{x_j} are, respectively, the mean and standard deviation of indicator j. If indicator j has negative polarity, formula (7.5) is multiplied by -1. This is equivalent to apply (7.1) and then (7.5). Standard scores may be further adjusted if calculations yield awkward values. For example, we can multiply each score by ten and add 100 to obtain more visually manageable scores (Booysen 2002). Standardization does not transform indicators to a common range. So, it is allows extreme values to influence the results because the range between the minimum and maximum standard scores will vary for each indicator.

Re-scaling (or *Min-Max*)

Re-scaling normalizes indicators to have an identical range [0, 1] as follows:

$$y_{ij} = \frac{x_{ij} - \min_i (x_{ij})}{\max_i (x_{ij}) - \min_i (x_{ij})} \tag{7.6}$$

where $\min_i (x_{ij})$ and $\max_i (x_{ij})$ are, respectively, a minimum and a maximum value that represent the possible range of indicator j (*goalposts*). If indicator j has negative polarity, the complement of (7.6) with respect to 1 is calculated.[2] This is equivalent to apply (7.1) and then (7.6). The *goalposts* can be selected relative to the observed minimum and maximum values of the indicator, be it for a specific year or over an extended period of time. Alternatively, they can be fixed by experts. Re-scaling is based on the range and it is sensitive to outliers. On the other hand, the range for indicators with very little variation will increase and these will contribute more to the composite indicator than they would using another method.

Distance from a Reference (or *Indicization*)

This method takes the percentage ratio between original values and a reference for each indicator. The indicized value is given by:

$$y_{ij} = \frac{x_{ij}}{x_{oj}} 100 \tag{7.7}$$

[2] The 'complement with respect to 1' is the number to add to make 1.

where x_{oj} is the reference value for indicator j – generally, the maximum or an external benchmark. In this method, the reference is given a value of 100 and units receive a score depending on their distance from it. Values greater (less) than 100 indicate above (below) reference performance. If indicator j has negative polarity, formula (7.2) can be preliminarily applied; however indicization is recommended only for indicators with positive polarity. Moreover, it is less robust to the influence of outliers than other methods.

In Table 7.1 is reported an example of normalization with some hypothetical data for five statistical units. The table provides the normalized indicator by the different methods, for positive and negative polarity, and the basic statistics of the normalized values (with the characteristics in bold).

Note that normalized indicators by ranking have a mean of $(n+1)/2=(5+1)/2=3$. Z-scores have a mean of 0 and standard deviation of 1, so that the variability 'effect' is nullified. Re-scaled indicators range between 0 and 1 and variability does not change by inverting polarity (standard deviation = 0.34); however the mean of the normalized values for negative polarity is the complement with respect to 1 of the mean for positive polarity, so the coefficient of variation (CV) is different (91.9 versus 55.1). Finally, indicized indicators have a maximum of 100 and save the original CV, but only for positive polarity (CV = 68.9).

The main pros and cons of different normalization methods are summarized in Table 7.2.

Potential problems include the loss of interval level information (e.g., ranking), sensitivity to outliers (e.g., standardization, re-scaling and indicization), and implicit weighting (e.g., indicization). The different transformations will therefore have significant effects on the construction of the composite index, and important incentive effects on the behaviour of units being assessed (Jacobs et al. 2004).

The Aggregation

Aggregation is the combination of all the components to form one or more composite indices. This step requires the definition of the importance of each individual indicator (weighting system) and the identification of the technique (compensatory, partially compensatory or non-compensatory) for summarizing the individual indicator values into a single number.

Formally, we have to move from the normalized matrix $\mathbf{Y} = \{y_{ij}\}$, with n rows (statistical units) and m columns (normalized indicators), to the vector $\mathbf{C} = \{c_i\}$, with n rows:

Table 7.1 Comparing normalization methods

		Normalized indicator (y)							
		Positive polarity (+)				Negative polarity (−)			
Unit	Original indicator (x)	Ranking	Z-scores	Re-scaling	Indicization	Ranking	Z-scores	Re-scaling	Indicization
1	450.0	1.0	1.81	1.00	100.0	5.0	−1.81	0.00	11.1
2	200.0	2.5	0.00	0.38	44.4	3.5	0.00	0.63	25.0
3	200.0	2.5	0.00	0.38	44.4	3.5	0.00	0.63	25.0
4	100.0	4.0	−0.73	0.13	22.2	2.0	0.73	0.88	50.0
5	50.0	5.0	−1.09	0.00	11.1	1.0	1.09	1.00	100.0
Min	50.0	1.0	−1.09	**0.00**	11.1	1.0	−1.81	**0.00**	11.1
Max	450.0	5.0	1.81	**1.00**	**100.0**	5.0	1.09	**1.00**	**100.0**
Mean	200.0	**3.0**	**0.00**	0.38	44.4	**3.0**	**0.00**	0.63	42.2
Std	137.8	1.4	**1.00**	0.34	30.6	1.4	**1.00**	0.34	31.5
CV (%)	68.9	45.9	–	91.9	**68.9**	45.9	–	55.1	74.6

Table 7.2 Pros and Cons of normalization methods

Normalization method	Pros	Cons
Ranking	Applicable to indicators with positive, negative and zero values	Loss of information (from interval/ratio scale to ordinal scale)
	Suitable both for bounded and unbounded indicators[a]	Assumes equal intervals between consecutive values
	No/low implicit weighting (normalized indicators have equal or similar variances)	Aggregation by a mathematical function is questionable for ordinal data
	Insensitive to outliers	
Standardization (or Z-scores)	Applicable to indicators with positive, negative and zero values	Not very suitable for bounded indicators
	No implicit weighting (normalized indicators have equal variances)	Produces negative values
		Sensitive to outliers
Re-scaling (or Min-Max)	Applicable to indicators with positive, negative and zero values	Not very suitable for unbounded indicators
	Low implicit weighting (normalized indicators have similar variances)	The mean reference can be lost
		Sensitive to outliers (the range depends on extreme values)
Distance from a reference (or Indicization)	Suitable both for bounded and unbounded indicators	Not applicable to indicators with negative values (zero values are accepted only for indicators with positive polarity)
	Saves the coefficient of variation (only for indicators with positive polarity)	High implicit weighting (normalized indicators have different variances)
		Very sensitive to outliers

[a]Indicators can be divided in 'bounded' and 'unbounded'. We say that an indicator is 'bounded' when it ranges between fixed values. An example of bounded indicator is the 'Employment rate' that always ranges between 0 and 100. We say that an indicator is 'unbounded' when there are no predetermined upper or lower limits. An example of unbounded indicator is the 'Household disposable income', because there is theoretically no limit to how high the income could be

$$\mathbf{Y}_{n,m} = \begin{pmatrix} y_{11} & \cdots & y_{1j} & \cdots & y_{1m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{i1} & \cdots & y_{ij} & \cdots & y_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{n1} & \cdots & y_{nj} & \cdots & y_{nm} \end{pmatrix} \Rightarrow \mathbf{C}_n = \begin{pmatrix} c_1 \\ \cdots \\ c_i \\ \cdots \\ c_n \end{pmatrix}$$

where c_i is the value of the composite index for unit i.

The literature offers a wide variety of aggregation methods, each with its pros and cons. They range from the simple arithmetic o geometric mean to multivariate statistical methods. In this chapter, some traditional and more recent methods are reported: the *Power mean of order r*, *Wroclaw Taxonomic Method*, *Mean-Min Function*, *Mazziotta-Pareto Index*, and *Principal Component Analysis*.

Aggregation is the most important and delicate step of the procedure. In this stage, the choices of the researcher assume a fundamental role, from a methodological point of view, as even minimal changes in the method applied can have major impact on the result. Therefore, data aggregation has always been an interesting but controversial topic in composite index construction (Saltelli 2007).

The Weighting System

In addition to the implicit weights introduced during normalization, explicit weights may be defined during aggregation. The aim with explicit weighting is that weights should reflect the relative importance (significance, reliability or other characteristics) of the individual indicators. The weights given to different indicators heavily influence the outcomes of the composite index. So, weights ideally should be selected according to an underlying theoretical framework for the composite index.

The most widely used techniques for weighting individual indicators are the following: (a) no explicit weighting (equal weighting approach), (b) expert weighting, and (c) Principal Component Analysis (PCA).

(a) If no explicit weighting is defined other than that implicitly introduced during the normalization, equal weights are applied to all individual indicators. This implies that all indicators in the composite index have equal importance, which may not be the case. However, if there are no statistical or empirical grounds for choosing different weights, this may be a valid approach in some contexts.[3]
(b) Expert weighting is typically set by a group of specialists who define weights for each indicator. The values determined by specialists are then averaged. Weights are sometimes defined by policy makers or social surveys about how meaningful or important individual indicators are to people.
(c) PCA can be used to set weights by using the coefficients of the first principal component. This is an empirical and relatively more objective option for weight selection and it has the advantage of determining that set of weights which explains the largest variation in the original indicators.[4]

Since different weighting systems imply different results and, given the subjectivity inherent in many of these criteria, no explicit weighting should be the norm and the burden of proof should fall on differential weighting (Booysen 2002).

[3]Note that the equal weighting approach may give extra weight to certain performance aspects if several individual indicators are in effect measuring the same attribute. As a remedy, indicators could be tested for statistical correlations, and lower weights could be given to variables strongly correlated with each other. On the other hand, correlations may merely show that unit performance on these indicators is similar (Freudenberg 2003).

[4]Although PCA has a number of excellent mathematical properties, its use in weighting components of social indices is dubious. For example, it may lead to indicators which have little variation being assigned small weights, irrespective of their possible contextual importance (Salzman 2003).

The Compensability Issue

A fundamental issue concerning composite index construction is the degree of compensability or substitutability of the individual indicators.

The components of a composite index are called ‘substitutable’ if a deficit in one component may be compensated by a surplus in another (e.g., a low value of “Proportion of people who have participated in religious or spiritual activities” can be offset by a high value of “Proportion of people who have participated in meetings of cultural or recreational associations” and vice versa). Similarly, the components of a composite index are called ‘non-substitutable’ if a compensation among them is not allowed (e.g., a low value of “Life expectancy at birth” cannot be offset by a high value of “GDP per capita” and vice versa).[5] Thus we can define an aggregation approach as ‘compensatory’ or ‘non-compensatory’ depending on whether it permits compensability or not (Casadio Tarabusi and Guarini 2013).

Compensability is closely related with the concept of unbalance, i.e., a disequilibrium among the indicators that are used to build the composite index. In any composite index each dimension is introduced to represent a relevant aspect of the phenomenon considered, therefore a measure of unbalance among dimensions may help the overall understanding of the phenomenon. In a non-compensatory or partially compensatory[6] approach, all the dimensions of the phenomenon must be balanced and an aggregation function that takes unbalance into account, in terms of penalization, is often used.

A compensatory approach involves the use of additive methods, such as the arithmetic mean. A non-compensatory or partially compensatory approach generally requires the use of non-linear functions, such as the geometric mean (OECD 2008) or Multi-Criteria Analysis (MCA) (Munda and Nardo 2009).

Power Mean of Order r

The power mean of order r aggregates normalized indicator as follows:

$$\mathrm{M}_i^r = \left(\sum_{j=1}^{m} y_{ij}^r w_j \right)^{\frac{1}{r}}$$

where w_j is the weight of indicator j $(0 < w_j < 1)$ and $\sum_{j=1}^{m} w_j = 1$.

[5]Note that compensability/non-compensability does not imply dependence/independence and vice-versa. For example, “Hospital beds (per 1000 people)” and “Hospital doctors (per 1000 people)” are two dependent (positively correlated) indicators but they are non-substitutable, because a deficit in beds cannot be compensated by a surplus in doctors and vice-versa (Mazziotta and Pareto 2015a).

[6]Note that a ‘partially compensatory’ approach can be considered ‘non-compensatory’, since it is not full compensatory.

Table 7.3 Special cases of the power mean of order r

Order	Formula	Aggregation function	Approach	Penalization	
				Intensity	Direction
$r \to -\infty$	$M_i^{-\infty} = \min_j (y_{ij})$	Minimum	Non-compensatory	Maximum	Downward
$r = -1$	$M_i^{-1} = \left(\sum_{j=1}^{m} \frac{w_j}{y_{ij}}\right)^{-1}$	Harmonic mean	Partially compensatory	High	Downward
$r \to 0$	$M_i^0 = \prod_{j=1}^{m} y_{ij}^{w_j}$	Geometric mean	Partially compensatory	low	Downward
$r = 1$	$M_i^1 = \sum_{j=1}^{m} y_{ij} w_j$	Arithmetic mean	Compensatory	None	–
$r = 2$	$M_i^2 = \left(\sum_{j=1}^{m} y_{ij}^2 w_j\right)^{\frac{1}{2}}$	Quadratic mean	Partially compensatory	low	Upward
$r = 3$	$M_i^3 = \left(\sum_{j=1}^{m} y_{ij}^3 w_j\right)^{\frac{1}{3}}$	Cubic mean	Partially compensatory	High	Upward
$r \to +\infty$	$M_i^{+\infty} = \max_j (y_{ij})$	Maximum	Non-compensatory	Maximum	Upward

For $r = 1$, we have an additive averaging. In particular, if $w_j = \frac{1}{m}$, then M_i^1 is the *simple arithmetic mean*. This technique is advantageous because of its methodological transparency, but it implies full compensability, such that poor performance in some indicators can be compensated for by sufficiently high values in other indicators.

In Table 7.3 are reported some special cases of power mean of order r. The table also provides the type of approach and the features (intensity and direction) of the penalization for unbalanced values. If the composite index to be constructed is 'positive', i.e., increasing values of the index correspond to an improvement of the phenomenon (e.g., socio-economic development), a downward penalization must be used. On the contrary, if the composite index is 'negative', i.e., increasing values of the index correspond to a worsening of the phenomenon (e.g., poverty), an upward penalization must be used. In any cases, an unbalance among indicators values will have a negative effect on the value of the index.[7]

Due to the penalization (upward or downward), we have:

$$M_i^{-\infty} \leq \ldots \leq M_i^{-1} \leq M_i^0 \leq M_i^1 \leq M_i^2 \leq M_i^3 \leq \ldots \leq M_i^{+\infty}$$

and the means are equal if and only if $y_{ij} = y_{ik}$ $(j \neq k)$.

[7]Note that a simple non-compensatory approach uses the minimum (maximum) value of the normalized indicators so that the other values cannot increase (decrease) the value of the index. This function realizes the maximum penalization for unbalanced values of the indicators (Casadio Tarabusi and Guarini 2013).

Note that not all aggregation functions are compatible with all normalization methods. For example, if the individual indicators are transformed in z-scores (standardization), they cannot be aggregated by a geometric mean because it is defined only for sets of positive values.

One approach commonly used in economics is to calculate the Jevons Index (geometric mean of indicized indicators). This method allows to build, for each unit, two closely interrelated composite indices: a 'static' index for space comparisons, and a 'dynamic' index for time comparisons (Mazziotta and Pareto 2016).

Given a set of individual indicators with positive polarity, let x_{ij}^t denote the value of the indicator j for unit i, at time t, where $x_{ij}^t > 0$ ($j = 1, \ldots, m$; $i = 1, \ldots, n$; $t = t_0$, t_1). The 'static' composite index may be defined as follows:

$$\mathrm{SJ}_i^t = \prod_{j=1}^{m} \left(\frac{x_{ij}^t}{x_{oj}^t} 100 \right)^{\frac{1}{m}}$$

where x_{oj}^t is the reference value for indicator j at time t (e.g., the average).

In order to compare the data from time t_0 to t_1, for each unit, we can construct a 'dynamic' composite index given by:

$$\mathrm{DJ}_i^{t_1/t_0} = \prod_{j=1}^{m} \left(\frac{x_{ij}^{t_1}}{x_{ij}^{t_0}} 100 \right)^{\frac{1}{m}}.$$

For the 'circularity' or 'transitivity' property of the *index number* theory, SJ and DJ are linked by the relation:

$$\mathrm{DJ}_i^{t_1/t_0} = \left(\mathrm{SJ}_i^{t_1} / \mathrm{SJ}_i^{t_0} \right) \mathrm{DJ}_o^{t_1/t_0}.$$

SJ and DJ are meaningful only for indicators with positive values. They give more weight to the low values and penalize downwards the unbalance among components.

Examples of well-known composite indices based on the power mean of order r are the United Nations' HDI (geometric mean of re-scaled values) and HPI (cubic mean of re-scaled values).

Wroclaw Taxonomic Method

This method was developed by a group of Polish mathematicians and applied to the aggregation of indicators of economic development (Harbison et al. 1970). It rests on the concept of 'ideal unit': a hypothetical unit that has, for each indicator, the most desirable value within the data set (optimal score).

The Euclidean distance from each unit to the 'ideal unit' is then calculated as follows:

$$\mathrm{D}_i = \sqrt{\sum_{j=1}^{m} \left(y_{ij} - y_{oj} \right)^2}$$

where y_{ij} is the standardized value by (7.5) and y_{oj} is equal to $\min_i \left(y_{ij} \right)$ or $\max_i \left(y_{ij} \right)$ according to whether indicator j has negative or positive polarity. The composite index for unit i is given by:

$$\mathrm{d}_i = \frac{\mathrm{D}_i}{\mathrm{M_D} + 2\mathrm{S_D}}$$

where $\mathrm{M_D}$ and $\mathrm{S_D}$ are, respectively, the mean and standard deviation of the distances D_i.

The index is equal to zero when the distance between a given unit and the 'ideal unit' is null (all the values coincide). The higher is the index, the greater is the difference between the two units. The main weakness of this method is the criterion for defining the 'ideal unit' (Silvio-Pomenta 1973).

Mean-Min Function

The Mean-Min Function (MMF) is a two-parameter function that incorporates two extreme cases of penalization of unbalance: the zero penalization represented by the arithmetic mean (compensatory approach) and the maximum penalization represented by the minimum function (non-compensatory approach). The function penalizes downwards and all other possible cases are intermediate.

The composite index is defined as:

$$\mathrm{MMF}_i = \mathrm{M}_{y_i} - \alpha \left(\sqrt{\left(\mathrm{M}_{y_i} - \min_j \left\{ y_{ij} \right\} \right)^2 + \beta^2} - \beta \right) \quad (0 \le \alpha \le 1; \;\; \beta \ge 0)$$

where M_{y_i} is the mean of the normalized values for unit i, and the parameters α and β are respectively related to the intensity of penalization of unbalance and degree of complementarity between indicators (Casadio Tarabusi and Guarini 2013).

The function reduces to the arithmetic mean for $\alpha = 0$ (in this case β is irrelevant) and to the minimum function for $\alpha = 1$ and $\beta = 0$. So, the interval of definition of the values of the composite index is: $\min_j \left\{ y_{ij} \right\} \le \mathrm{MMF}_i \le \mathrm{M}_{y_i}$.

The MMF is independent from the choice of the normalization method. By choosing the values of parameters appropriately one should obtain the aggregation function that best suits the specific theoretical approach. However, there is not a general rule for tuning these values (Mazziotta and Pareto 2015b).

Mazziotta-Pareto Index

The Mazziotta-Pareto Index (MPI) is a composite index for summarizing a set of indicators that are assumed to be not fully substitutable. It is based on a non-linear function which, starting from the arithmetic mean of the normalized indicators, introduces a penalty for the units with unbalanced values of the indicators (De Muro et al. 2011). Two version of the index have been proposed: (a) MPI, and (b) adjusted MPI (AMPI). The first version is the best solution for a 'static' analysis (e.g., a single-year analysis), whereas the second one is the best solution for a 'dynamic' analysis (e.g., a multi-year analysis).

(a) MPI

The MPI is based on the following normalization:

$$z_{ij} = 100 + 10y_{ij}$$

where y_{ij} is given by (7.5).[8]

Denoting with M_{z_i}, S_{z_i}, cv_{z_i}, respectively, the mean, standard deviation, and coefficient of variation of the normalized values for unit i, the composite index is given by:

$$\mathrm{MPI}_i^{+/-} = \mathrm{M}_{z_i} \pm \mathrm{S}_{z_i}\mathrm{cv}_{z_i}$$

where the sign $\pm$ depends on the kind of phenomenon to be measured. If a downward penalization is required, then the MPI^- is used, else the MPI^+ is used.

Therefore, the MPI decomposes the score of each unit in two parts: mean level (M_{z_i}) and penalty ($\mathrm{S}_{z_i}\mathrm{cv}_{z_i}$). The penalty is a function of the indicators' variability in relation to the mean value ('horizontal variability') and it is used to penalize the units. The aim is to reward the units that, mean being equal, have a greater balance among the indicators values.

(b) AMPI

The AMPI normalizes indicators as follows:

$$r_{ij} = y_{ij}60 + 70$$

where y_{ij} is given by (7.6). To facilitate the interpretation of results, the 'goalposts' can be chosen so that 100 represents a reference value (e.g., the average in a given year). Let Inf_{x_j} and Sup_{x_j} be the minimum and maximum of indicator j across all time periods considered, and Ref_{x_j} be the reference value for indicator j. Then the 'goalposts' are defined as: $\mathrm{Ref}_{x_j} \pm \Delta$, where and $\Delta = \left(\mathrm{Sup}_{x_j} - \mathrm{Inf}_{x_j}\right)/2$.[9]

[8] Normalized indicators have a mean of 100 and standard deviation of 10.

[9] Normalized indicators range approximately between 70 and 130.

Denoting with M_{r_i}, S_{r_i}, cv_{r_i}, respectively, the mean, standard deviation, and coefficient of variation of the normalized values for unit *i*, the composite index is given by:

$$\mathrm{AMPI}_i^{+/-} = \mathrm{M}_{r_i} \pm \mathrm{S}_{r_i}\mathrm{cv}_{r_i}$$

where the sign $\pm$ depends on the kind of phenomenon to be measured. If a downward penalization is required, then the AMPI^- is used, else the AMPI^+ is used.

The main difference between MPI and AMPI is the normalization method. The MPI is based on a standardization of the individual indicators that is repeated independently for each time period, so it is not possible to appreciate any absolute change in unit performance. The AMPI is based on a re-scaling and measures absolute variations with respect to prefixed goalposts. Moreover, the AMPI allows to compute the score of each unit independently of the others, in contrast to the MPI where the mean and standard deviation of the individual indicators are required. For a comparison between the two versions, see Mazziotta and Pareto (2015a).

Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical method that, starting from a large number of individual indicators, allows to identify a small number of composite indices (principal components or factors) that explain most of the variance observed (Dunteman 1989). The first principal component is often used as the 'best' composite index. It is defined as:

$$\mathrm{C}_{i1} = \sum_{j=1}^{m} a_{j1}\, x_{ij}$$

where a_{j1} is the weight of indicator *j* for factor 1.

This composite index has many optimal mathematical properties. The most important is that it explains the largest portion of variance of the individual indicators. This is obtained by maximizing the sum of the squares of the coefficients of correlation between the composite index and the individual indicators. However, the first principal component accounts for a limited part of the variance in the data, so we can lose a consistent amount of information. Moreover, the PCA based index is often 'elitist' (Mishra 2007), with a strong tendency to represent highly intercorrelated indicators and to neglect the others, irrespective of their possible contextual importance. So many highly important but poorly intercorrelated indicators may be unrepresented by the composite index.

An alternative method is the weighted mean of the factors (Giudici and Avrini 2002). This approach consists in aggregating individual indicators by a weighted mean of factor scores, with weights proportional to the variance explained by each of the components. The composite index for unit i is:

$$\mathrm{S}_i = \frac{\sum_{h=1}^{p} \mathrm{C}_{ih}\lambda_h}{\sum_{h=1}^{p} \lambda_h}$$

where C_{ih} is the value of factor h for unit i, λ_h is the percentage of variance explained by factor h, and p is the number of considered factors ($p \leq m$). If $p = m$, no information is loss. The method assigns decreasing order of importance to the factors, according to their amount of variance explained.

The Validation

Validation step aims to assess the robustness of the composite index, in terms of capacity to produce correct and stable measures, and its discriminant capacity. As seen above, the outcomes and rankings of individual units on the composite index may largely depend on the decisions taken at each of the preceding steps (selection of individual indicators, normalization and aggregation). For this reason, statistical analyses should be conducted to explore the robustness of rankings to the inclusion and exclusion of individual indicators and setting different decision rules to construct the composite index (Freudenberg 2003).

Robustness of a composite index is assessed by two different methodologies: Uncertainty analysis (UA) and Sensitivity analysis (SA). UA focuses on how uncertainty in the input factors propagates through the structure of the composite index and affects the results. SA studies how much each individual source of uncertainty contributes to the output variance (Saisana et al. 2005). UA and SA can be used synergistically and iteratively during composite index construction to help in indicator selection, add transparency to the index construction process, and explore the robustness of alternative composite index designs and rankings (USAID 2014).

Discriminant capacity of a composite index is assessed by exploring its capacity in: (a) discriminating between units and/or groups; (b) distributing all the units without any concentration of individual scores in a few segments of the continuum; (c) showing values that are interpretable in terms of selectivity through the identification of particular reference values or *cut-points* (Maggino and Zumbo 2012).[10]

[10]Point (a) can be verified by applying the traditional approaches of statistical hypothesis testing, whereas specific coefficients were proposed for evaluating (b) (Guilford 1954). Receiver operating characteristic (ROC) analysis allows to identify discriminant *cut-points* in (c).

Uncertainty Analysis (UA)

UA is essentially based on simulations that are carried out on the various equations that constitute the underlying model. A valid approach for evaluating output uncertainty is the Monte Carlo method, which is based on multiple evaluations of the model with a set of randomly selected input factors (OECD 2008).

The steps of the procedure are summarized below:

1. Identify k input factors F_i $(i = 1, \ldots, k)$ that can introduce uncertainty in the results (e.g., errors in individual indicators, exclusion of an individual indicator, etc.).
2. Assign a probability density function to each input factor (e.g., normal distribution for the errors in individual indicators; discrete uniform distribution to select the individual indicator to be excluded, etc.).
3. Generate randomly L combinations or samples of independent input factors F_1^l, $F_2^l, \ldots, F_k^l$ $(l = 1, 2, \ldots, L)$ and calculate the corresponding value of the composite index for each unit $c_i^l = f\left(x_{i1}, x_{i2}, \ldots, x_{im}; F_1^l, F_2^l, \ldots, F_k^l\right)(i = 1, 2, \ldots, n; l = 1, 2, \ldots, L)$.
4. Calculate the average shift in countries' ranks $\bar{\mathrm{R}}^l = \frac{1}{n}\sum_{i=1}^{n}\left|r_i^l - r_i\right|$ $(l = 1, 2, \ldots, L)$, where $r_i^l = \mathrm{rank}\left(c_i^l\right)$ is the rank assigned by the composite index to unit i for sample l and r_i is the original rank of unit i.
5. Analyse the distribution of $\bar{\mathrm{R}}^l$ and/or r_i^l $(l = 1, 2, \ldots, L)$. The main characteristics of this distributions, such as the mean and variance, are estimated with an level of precision related to the size of the simulation L. In general, the lower the variance, the greater the robustness.

A particular case of UA is the Influence analysis (IA) that aims to empirically quantify the 'weight' of each individual indicator in the calculation of the composite index. Given m individual indicators, the IA performs steps 3 and 4, with $L = m$, by excluding each time indicator l. The value of $\bar{\mathrm{R}}^l$ represents the 'weight' of indicator l (Mazziotta et al. 2010).

Sensitivity Analysis (SA)

SA examines the degree of influence of each input factor on the composite index, thereby helping to reveal how much each individual source of uncertainty contributes to the output variance (OECD 2008).

The importance of a given input factor F_i can be measured via the so-called *sensitivity index*, which is defined as the fractional contribution to the model output variance due to the uncertainty in F_i. For k independent input factors, the sensitivity indices can be computed by using the following decomposition formula for the total variance of the output ($\bar{\mathrm{R}}^l$ or r_i^l):

$$\mathrm{V} = \sum_i \mathrm{V}_i + \sum_i \sum_{j>i} \mathrm{V}_{ij} + \ldots + \mathrm{V}_{12\ldots k}$$

where V_i is the output variance due to the uncertainty in F_i, V_{ij} is the output variance due to uncertainty of the interaction between F_i and F_j, and so on.

A first measure of the fraction of the output variance V that is accounted for by the uncertainty in F_i is the *first-order sensitivity index* for the factor F_i defined as:

$$\mathrm{S}_i = \frac{\mathrm{V}_i}{\mathrm{V}}$$

A measure that concentrates in one single term all the interactions involving a given factor F_i is the *total effect sensitivity index*, given by:

$$\mathrm{S}_{Ti} = \frac{\mathrm{V}_i}{\mathrm{V}} + \sum_{j \neq i} \frac{\mathrm{V}_{ij}}{\mathrm{V}} + \ldots + \frac{\mathrm{V}_{12\ldots k}}{\mathrm{V}} = \mathrm{S}_i + \sum_{j \neq i} \mathrm{S}_{ij} + \ldots + \mathrm{S}_{12\ldots k}$$

where S_{ij} is the *second-order sensitivity index* for the factors F_i and F_j, and so on.

If the model has no interactions among its input factors (*additive* model), we have $\mathrm{S}_{Ti} = \mathrm{S}_i$ $(i = 1, 2, \ldots, k)$ and $\sum_i \mathrm{S}_{Ti} = 1$. In general, $\sum_i \mathrm{S}_{Ti} \geq 1$, and a significant difference between S_{Ti} and S_i signals an important interaction role for the factor F_i in the output.

Estimators for both (S_i, S_{Ti}) are provided by a variety of methods, such as the method of Sobol (Saisana et al. 2005).

Best Practices

As we have seen above, there does not exist a composite index universally valid for all areas of application, since its validity depends on the strategic objectives of the research. In this Section we propose a scheme with some general guidelines to follow for constructing a composite index.

The main factors to take into account in the choice of the method to be adopted for summarizing a set of individual indicators are as follows (Mazziotta and Pareto 2013):

– type of indicators (substitutable/non-substitutable);
– type of aggregation (simple/complex);
– type of comparisons (absolute/relative);
– type of weights (objective/subjective).

There is not always a 'well-established' solution, and sometimes it may be necessary to renounce to some requirements, to satisfy others.

Type of Indicators

It is one of the main factors that affect the choice of the aggregation method. If the individual indicators are substitutable, then a compensatory approach is indicated, else a non-compensatory or partially compensatory approach is required.

Type of Aggregation

The choice of the aggregation method also depends on the aim of the work and on the type of ‘users’ (researchers or people). Generally, an aggregation method can be considered ‘simple’ or ‘complex’. We say that an aggregation method is ‘simple’ when a easily understandable mathematical function is used (e.g., the HDI). On the contrary, an aggregation method is said to be ‘complex’ if a sophisticated model or multivariate method is used (e.g., PCA).

Type of Comparisons

Data normalization firstly depends on the type of comparisons required. All the normalization methods allow for space comparisons, whereas time comparisons of the units may be difficult to make or to interpret. Comparisons over time may be ‘absolute’ or ‘relative’. We say that a time comparison is ‘relative’ when the composite index values, at time t, depend on one or more endogenous parameters (e.g., mean and variance of the individual indicators at time t). Similarly, we say that a time comparison is ‘absolute’ when the composite index values, at time t, depend on one or more exogenous parameters (e.g., minimum and maximum of the individual indicators fixed by the researcher). Ranking and standardization allow only for relative comparisons since they are based exclusively on values of the individual indicators at time t. Other methods, such as re-scaling or indicization, require that the minimum and maximum (e.g., the ‘goalposts’ of the HDI) or the base of index numbers are independent from the time t, in order to perform comparisons in absolute terms (Tarantola 2008).

Type of Weights

The question of the choice of a weighting system in order to weigh the individual indicators, according to their different importance in expressing the considered phenomenon, necessarily involves the introduction of an arbitrary component.

A subjective weighting can be adopted, implicitly, by assigning the same weight to all the components (equal weighting) or, explicitly, by a group of experts. Alternatively, an objective weighting can be applied, implicitly, by choosing a normalization method that assigns a weight proportional to the variability of the indicator or, explicitly, by multivariate statistical methods, such as PCA.

Figure 7.2 shows the flow chart for the choice of the 'best' method in constructing a composite index, with the main possible solutions (normalization, weighting and aggregation) for each 'path' followed (assumptions and requirements).

If the phenomenon to be measured is decomposable into more dimensions, each of them is represented by a subset of individual indicators, it may be more convenient to build a composite index for each dimension (or 'pillar') and then obtain the overall index by means of the aggregation of the partial composite indices. In this case, it is possible to follow a compensatory approach within each dimension and a non-compensatory or partially compensatory approach among the various dimensions.

The most used aggregation methods for substitutable indicators are the additive ones, such as the arithmetic mean (simple) or PCA (complex). For non-substitutable indicators, non-linear methods are instead used, such as multiplicative functions (simple) or MCA (complex).

Focusing on methods based on the use of mathematical functions, the type of normalization depends on the nature of the space-time comparisons to do and on the weight to be assigned to the individual indicators.

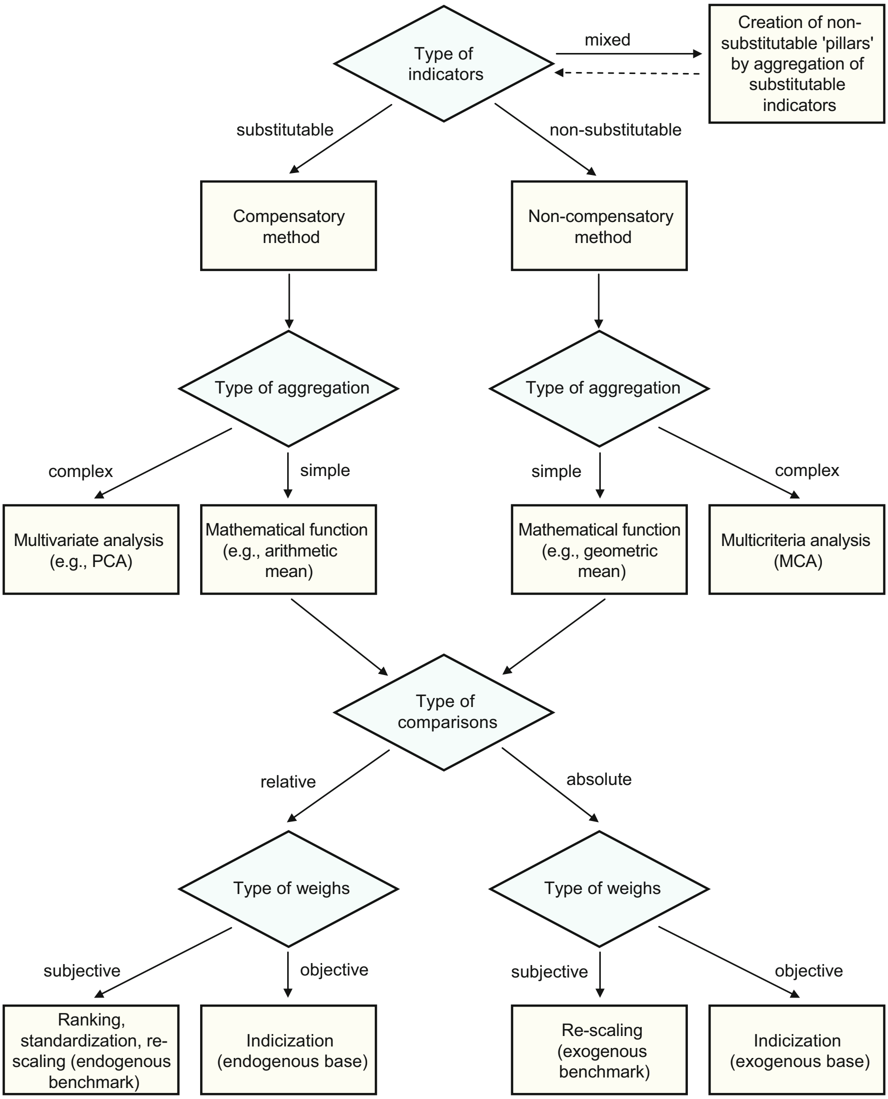
For relative comparisons with subjective weighting (equal or different weights), we recommend ranking, standardization or re-scaling with endogenous goalposts. For assigning objective weights proportional to the variability of the indicators is more suitable an indicization where it is assumed as a base the mean, the maximum value or another reference value of the distribution (endogenous base).

For absolute comparisons, it is not possible use ranking or standardization. In the case of subjective weighting, it is necessary to resort to a re-scaling with minimum and maximum values independent of the distribution (exogenous benchmark), whereas in the case of objective weighting, a indicization with externally fixed base may be a good solution (exogenous base).

An Application To Socio-economic Well-Being Data

In this Section, an application for constructing a composite index of well-being for OECD countries, in 2014, is presented (OECD 2013).

Four basic dimensions of well-being are considered: Health, Education, Jobs, Income. A formative measurement model is adopted and individual indicators are the following: "Life expectancy (years)" (X_1), "Educational attainment (% of people, aged 15–64)" (X_2), "Employment rate (% of people, aged 15–64)" (X_3), "Household disposable income (USD – PPPs adjusted)" (X_4).



Source: Mazziotta and Pareto, 2013

Fig. 7.2 Flow chart for the choice of the ‘best’ method

In Table 7.4 is reported the list of indicators and their description. Note that all individual indicators have positive polarity.

In order to assess the performance of different methods, we compared the following indices:

Table 7.4 Individual indicators of well-being and definitions

Individual indicator	Name	Definition
X_1	Life expectancy	It is the standard measure of the length of people's life. Life-expectancy measures how long on average people could expect to live based on the age specific mortality rates currently prevailing. Life-expectancy can be computed at birth and at various ages
X_2	Educational attainment	It profiles the education of the adult population as captured through formal educational qualifications. Educational attainment is measured as the percentage of the adult population (15–64 years of age) holding at least an upper secondary degree, as defined by the OECD-ISCED classification
X_3	Employment rate	It is the share of the working age population (people aged from 15 to 64 in most OECD countries) who are currently employed in a paid job. Employed persons are those aged 15 and over who declare having worked in gainful employment for at least 1 h in the previous week, following the standard ILO definition
X_4	Household disposable income	It includes income from work, property, imputed rents attributed to home owners and social benefits in cash, net of direct taxes and social security contributions paid by households; it also includes the social transfers in kind, such as education and health care, that households receive from governments. Income is measured net of the depreciation of capital goods that households use in production

Source: www.oecdbetterlifeindex.org

1. arithmetic mean of re-scaled values in [0, 1] (M_{0-1});
2. arithmetic mean of standardized values (Mz);
3. geometric mean of indicized values (SJ);
4. Mazziotta-Pareto Index with negative penalty (MPI^{-}).

Methods 1 and 2 are compensatory, whereas 3 and 4 are partially compensatory. Table 7.5 shows the data matrix **X** and the composite indices (values and ranks).

Overall, the results are concordant, except for SJ where an implicit weighting due to indicization is applied (Table 7.6). The Spearman rank correlation coefficient between M_{0-1} and Mz- is $\rho = 0.998$ (mean absolute difference of rank = 0.367, i.e., the rank of each unit changes, on average, of 0.367 positions between the two methods), whereas the more similar index to SJ is the MPI^{-}, with $\rho = 0.998$ (mean absolute difference of rank = 1.600).

Finally, an influence analysis was performed to assess the robustness of the composite indices. As we can see in Table 7.7, SJ is the more sensitive index (standard deviation = 1.13; range = 2.80), because indicization gives weights proportional to the variability, and then some indicators are very influential and others no. On the contrary, in a partially compensatory approach, the MPI^{-} tends to assign equal weight or importance to each indicator and it is less sensitive to the

Table 7.5 Individual indicators and composite indices of well-being

	Individual indicator[a]				Composite index							
					M_{0-1}		Mz		SJ		MPI^{-}	
Country	X_1	X_2	X_3	X_4	Value	Rank	Value	Rank	Value	Rank	Value	Rank
Australia	82.0	74.0	72.0	31,197	0.754	7.5	0.658	8.5	108.962	9.0	106.349	8.0
Austria	81.1	82.0	73.0	29,256	0.752	9.0	0.658	8.5	110.090	7.0	106.523	7.0
Belgium	80.5	71.0	62.0	27,811	0.575	16.0	−0.054	16.0	100.478	15.0	99.220	16.0
Canada	81.0	89.0	72.0	30,212	0.779	5.0	0.768	5.0	112.850	4.0	107.635	5.0
Chile	78.3	72.0	62.0	13,762	0.382	27.0	−0.822	26.0	83.984	28.0	91.559	26.0
Czech Rep.	78.0	92.0	67.0	17,262	0.537	19.0	−0.193	19.0	96.253	18.0	97.346	19.0
Denmark	79.9	77.0	73.0	25,172	0.657	14.0	0.283	13.0	103.986	12.0	102.667	13.0
Estonia	76.3	89.0	67.0	14,382	0.447	23.0	−0.542	22.0	90.700	24.0	93.406	22.0
Finland	80.6	84.0	70.0	26,904	0.697	11.0	0.437	11.0	107.161	10.0	104.359	10.0
France	82.2	72.0	64.0	29,322	0.662	13.0	0.281	14.0	103.526	14.0	102.448	14.0
Germany	80.8	86.0	73.0	30,721	0.774	6.0	0.751	6.0	112.673	5.0	107.427	6.0
Greece	80.8	67.0	51.0	19,095	0.387	26.0	−0.828	27.0	85.931	27.0	90.802	27.0
Hungary	75.0	82.0	57.0	15,240	0.296	29.0	−1.136	29.0	86.216	26.0	87.631	29.0
Ireland	80.6	73.0	59.0	23,721	0.522	20.0	−0.276	20.0	96.065	19.0	96.989	20.0
Italy	82.7	56.0	58.0	24,724	0.511	21.0	−0.336	21.0	91.036	23.0	95.511	21.0
Japan	82.7	93.0	71.0	25,066	0.790	3.5	0.787	4.0	109.078	8.0	107.731	4.0
Korea	81.1	81.0	64.0	18,035	0.562	17.0	−0.122	18.0	94.103	20.0	98.491	18.0
Mexico	74.4	36.0	61.0	12,850	0.094	30.0	−1.939	30.0	68.262	30.0	79.867	30.0
Netherlands	81.3	72.0	75.0	25,697	0.700	10.0	0.443	10.0	103.931	13.0	104.134	11.0
New Zealand	81.2	74.0	72.0	21,773	0.642	15.0	0.205	15.0	99.349	16.0	101.843	15.0
Norway	81.4	82.0	76.0	32,093	0.814	2.0	0.911	2.0	113.912	3.0	108.925	2.0
Poland	76.9	89.0	60.0	16,234	0.419	24.0	−0.660	24.0	91.123	22.0	92.549	24.0
Portugal	80.8	35.0	62.0	18,806	0.344	28.0	−0.988	28.0	76.418	29.0	88.641	28.0

(continued)

Table 7.5 (continued)

Country	Individual indicator[a]				Composite index							
					M_{0-1}		Mz		SJ		MPI^{-}	
	X_1	X_2	X_3	X_4	Value	Rank	Value	Rank	Value	Rank	Value	Rank
Slovak Rep.	76.1	91.0	60.0	17,228	0.413	25.0	−0.675	25.0	92.759	21.0	92.149	25.0
Slovenia	80.1	84.0	64.0	19,692	0.561	18.0	−0.116	17.0	96.772	17.0	98.658	17.0
Spain	82.4	54.0	56.0	22,799	0.458	22.0	−0.551	23.0	87.551	25.0	93.295	23.0
Sweden	81.9	87.0	74.0	27,546	0.790	3.5	0.802	3.0	110.708	6.0	107.986	3.0
Switzerland	82.8	86.0	79.0	30,745	0.888	1.0	1.190	1.0	115.648	2.0	111.735	1.0
United Kingdom	81.1	77.0	71.0	25,828	0.681	12.0	0.367	12.0	104.320	11.0	103.615	12.0
United States	78.7	89.0	67.0	39,531	0.754	7.5	0.695	7.0	117.692	1.0	105.746	9.0
Average	**80.1**	**76.5**	**66.4**	**23,757**	**0.588**		**0.000**		**100.000**		**100.000**	

[a]Source: www.oecdbetterlifeindex.org

Table 7.6 Comparison of rankings by different composite indices

Composite index	M_{0-1}	Mz	SJ	MPI^{-}
Rank correlation				
M_{0-1}	1.000	0.998	0.966	0.996
Mz	0.998	1.000	0.970	0.998
SJ	0.966	0.970	1.000	0.966
MPI^{-}	0.996	0.998	0.966	1.000
Mean absolute difference of rank				
M_{0-1}	0.000	0.367	1.833	0.500
M_z	0.367	0.000	1.633	0.200
SJ	1.833	1.633	0.000	1.600
MPI^{-}	0.500	0.200	1.600	0.000

Table 7.7 Influence analysis. Average shift in countries' ranks when excluding an indicator

Individual indicator	M_{0-1}	Mz	SJ	MPI^{-}
X_1	1.53	1.53	3.20	1.40
X_2	1.47	1.70	0.93	1.80
X_3	2.07	2.20	2.47	2.07
X_4	2.30	1.90	0.40	1.93
Mean	**1.84**	**1.83**	**1.75**	**1.80**
Standard deviation	**0.35**	**0.25**	**1.13**	**0.25**
Range	**0.83**	**0.67**	**2.80**	**0.67**

inclusion or exclusion of individual indicators (standard deviation = 0.25; range = 0.67).

Acknowledgements The chapter is the result of the common work of the authors: in particular, Matteo Mazziotta has written Sects. 7.1 and 7.4; Adriano Pareto has written Sects. 7.2 and 7.3.

References

Bandura, R. (2008). A *survey of composite indices measuring country performance: 2008 update*. A UNDP/ODS Working Paper, New York.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research, 59*, 115–151.

Casadio Tarabusi, E., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research, 112*, 19–45.

Daly, H. E., & Cobb, J. J. (1989). *For the common good: Redirecting the economy toward community, the environmental and a sustainable future*. Boston: Beacon Press.

De Muro, P., Mazziotta, M., & Pareto, A. (2011). Composite indices of development and poverty: An application to MDGs. *Social Indicators Research, 104*, 1–18.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*, 269–277.
Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203–1218.
Dunteman, G. H. (1989). *Principal components analysis*. Newbury Park: Sage Publications.
Fanchette, S. (1974). Social indicators: Problems of methodology and selection. In *Social indicators: Problems of definition and of selection*, *Reports and papers in the social sciences* (p. 30). Paris: UNESCO.
Freudenberg, M. (2003). *Composite indicators of country performance: A critical assessment*. OECD, DSTI/IND, 5, Paris.
Giovannini, E., & Rondinella, T. (2012). Measuring equitable and sustainable well-being in Italy. In F. Maggino & G. Nuvolati (Eds.), *Quality of life in Italy: Research and reflections* (pp. 27–40). New York: Springer.
Giudici, P., & Avrini, F. (2002). Modelli statistici per la costruzione di indicatori della qualità della vita: aspetti metodologici. *Rivista di statistica ufficiale, 1*, 61–80.
Gnesi, C., Segre, E., & Villa, A. (2010). *Come si vive in Italia? Rapporto QUARS 2010 – Indice di qualità regionale dello sviluppo*. www.sbilanciamoci.org
Guilford, J. P. (1954). *Psychometric methods*. New York/London: McGraw-Hill.
Harbison, F. H., Maruhnic, J., & Resnick, J. R. (1970). *Quantitative analyses of modernization and development*. Princeton: Princeton University Press.
Istat. (2015). *BES 2015, Il benessere equo e sostenibile in italia*. Available at: http://www.istat.it/it/files/2015/12/Rapporto_BES_2015.pdf
Jacobs, R., Smith, P., & Goddard, M. (2004). Measuring performance: An examination of composite performance indicators. The University of York, CHE Technical Paper Series, 29, York.
Maggino, F. (2014). Indicator development and construction. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3190–3197). Dordrecht: Springer.
Maggino, F., & Zumbo, B. D. (2012). Measuring the quality of life and the construction of social indicators. In K. C. Land et al. (Eds.), *Handbook of social indicators and quality of life research* (pp. 201–238). New York: Springer.
Mazziotta, M., & Pareto, A. (2012). A non-compensatory approach for the measurement of the quality of life. In F. Maggino & G. Nuvolati (Eds.), *Quality of life in italy: Research and reflections* (pp. 27–40). New York: Springer.
Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one? *Rivista Italiana di Economia Demografia e Statistica, LXVII*(2), 67–80.
Mazziotta, M., & Pareto, A. (2015a). On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social Indicators Research*. doi:10.1007/s11205-015-0998-2.
Mazziotta, M., & Pareto, A. (2015b). Comparing two non-compensatory composite indices to measure changes over time: A case study. *Statistika, 95*, 44–53.
Mazziotta, M., & Pareto, A. (2016). A well-being index based on the weighted product method. In G. Alleva & A. Giommi (Eds.), *Topics in theoretical and applied statistics, studies in theoretical and applied statistics*. Cham: Springer. doi:10.1007/978-3-319-27274-0_22.
Mazziotta, C., Mazziotta, M., Pareto, A., & Vidoli, F. (2010). La sintesi di indicatori territoriali di dotazione infrastrutturale: metodi di costruzione e procedure di ponderazione a confronto. *Review of Economics and Statistics for Regional Studies, 1*, 7–33.
Mishra, S. K. (2007). *A comparative study of various inclusive indices and the index constructed by the principal components analysis*. MPRA Paper, 3377. Available at: http://mpra.ub.uni-muenchen.de/3377
Munda, G., & Nardo, M. (2009). Non-compensatory/non-linear composite indicators for ranking countries: A defensible setting. *Applied Economics, 41*, 1513–1523.

Nordaus, W. D., & Tobin, J. (1972). Is growth obsolete? In: National bureau of economic research, *Economic research: Retrospect and prospect* (pp. 1–80). New York: Columbia University Press.

OECD. (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: OECD Publications.

OECD. (2011). *Compendium of OECD well-being indicators*. Paris: OECD.

OECD. (2013). *How's life? 2013: Measuring well-being*. Paris: OECD Publishing. doi:10.1787/9789264201392-en.

Rinaldi, A., & Zelli, R. (2014). *Misurare il Benessere*. Roma: Donzelli editore.

Saisana, M., & Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. European Commission-JRC, EUR 20408 EN, Ispra.

Saisana, M., Tarantola, S., & Saltelli, A. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A, 168*, 307–323.

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research, 81*, 65–77.

Salzman, J. (2003). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Technical Report, Center for the Study of Living Standards, Ottawa.

Silvio-Pomenta, J. F. (1973). *Typological study using the Wroclaw Taxonomic Method (A study of regional disprities in Venezuela)*. SHC/WS/316, UNESCO, Paris.

Tarantola, S. (2008). *European innovation scoreboard: Strategies to measure country progress over time*. JRC Scientific and Technical Reports, EUR 23526 EN, Luxembourg.

UNDP. (1990). *Human development report 1990: Concept and measurement of human development*. Oxford: Oxford University Press.

UNDP. (2001). *Human development report 2001*. New York: Oxford University Press.

UNDP. (2010). *Human development report 2010*. New York: Palgrave MacMillan.

USAID. (2014). *Design and use of composite indices in assessments of climate change vulnerability and resilience*. Washington, DC: U.S. Agency for International Development.

Veenhoven, R. (1996). Happy life-expectancy. A comprehensive measure of quality of life in nations. *Social Indicators Research, 39*, 1–58.

Chapter 8
Synthesis of Indicators: The Non-aggregative Approach

Marco Fattore

Introduction

The need for new tools in synthetic indicators construction in the social sciences is deeply related to the problem of describing and understanding increasingly complex societal facts. On the one hand, official surveys, administrative data, web data, open data, to say a few, are now easily available to social scientists in the form of wide and complex multidimensional indicator systems. On the other hand, as data complexity grows, the need to get effective synthetic views, capable to enhance decision-making, increases as well. New procedures for data treatment are necessary, to overcome the limitations of older approaches that are designed for simpler data systems, are based on the "synthesis-as-aggregation" paradigm and employ composite indicators as their main statistical tool. To make a concrete example, consider the "beyond GDP" perspective to well-being and to societal evaluation. Going "beyond GDP" invariably requires dealing with multidimensional systems of ordinal data (e.g. pertaining to ownership of goods, access to services, self-perception of health and economic status...), ruling out the possibility to directly apply the composite indicator approach to measurement (Fattore 2015). In this and similar contexts, two main issues arise.

1. Ordinal attributes cannot be aggregated through linear combinations, averages or other functionals, designed for numerical variables. In fact, ordinal scores cannot be summed, multiplied by scalars or composed in other ways. For this reason, they are often transformed into numerical scores, through more or less sophisticated scaling tools, before aggregation. Unfortunately, there are evidences that such procedures may lead to controversial results (Madden 2010). Moreover, one could legitimately ask why concepts naturally conceived in

M. Fattore (✉)
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
e-mail: marco.fattore@unimib.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_8

ordinal terms should be forced into numerical settings. Is the idea of ordinal scores as rough manifestations of underlying continuous traits always well founded? Or is it actually motivated by the lack of consistent and effective procedures for the treatment of ordinal data? Such problems go beyond the setting of well-being measurement and arise in many other fields as well. For example, in marketing and customer segmentation, in ecological and environmental studies, in risk management and, more generally, in ordinal multi-criteria decision-making (Bruggemann et al. 1999; Annoni and Bruggemann 2009; Bruggemann and Patil 2010, 2011; Bruggemann and Voigt 2012; Bruggemann and Carlsen 2014; Carlsen and Bruggemann 2014). It is in fact a feature of modern information society that most of data we deal with are of a discrete and qualitative kind. The absence of statistical tools and procedures to manage such data types consistently may well turn into severe limitations in our capability to exploit the great amount of information they convey.

2. Independently of their nature, it is a matter of fact that many data systems available to social scientists often comprise weakly interdependent attributes. The absence of strong interconnections in a multi-indicator system prevents from achieving effective dimension reductions through aggregation procedures. Consequently, and independently of the models or of the procedures they are computed from (e.g. latent variables or structural equation models, PLS path modeling or other formative aggregation tools), composite indicators are inherently inappropriate in these situations, being aggregative and compensative. This leads to a fundamental question: is attribute aggregation the only road to synthesis? The answer to this question motivates most of the present chapter.

All in all, for several and different reasons, composite indicators may often be inadequate for the needs of social scientists. It is the aim of this chapter to illustrate and discuss possible alternatives to the composite indicator paradigm, particularly when dealing with ordinal attributes. Specifically, we will focus on the role of Partial Order Theory in the development of non-aggregative synthetic indicators, i.e. indicators which are not built as aggregation of attribute scores. The goal is to show that the theory of partially ordered sets does represent the natural setting for addressing evaluation problems on multidimensional systems of ordinal data and, more generally, on multi-indicator systems, even with mixed data types. We will keep the text at a very simple level, avoiding deep mathematical discussions. The aim of the chapter is in fact to introduce readers "lightly" to a different approach to indicator construction, making them aware of alternative methodologies and providing a "tour" of possible statistical tools. Consider, however, that concepts and results illustrated in the following are mathematically well founded and technical details can be found in cited references. As a matter of fact, applications of the theory of partial orders to statistics and the social sciences are at a very early stage. Much is still to be investigated, discovered and formalized and perhaps only the simplest results have been worked out insofar. Nevertheless, they are enough to appreciate the potentialities of the approach and to extend the conceptual and practical toolbox of social scientists (Fattore and Maggino 2014). The chapter is organized as follows. In Paragraph 2, a brief account of Partial Order Theory is given, to introduce some basic concepts and set the notation. In Paragraph 3, it is discussed how to build simple synthetic indicators, to rank statistical units based on their attribute scores, without aggregating them. It is also briefly introduced a recently developed methodology to perform multidimensional assessments on multi-indicator systems. Such a methodology has been developed in the context of multidimensional deprivation evaluation and represents an alternative to (and, in a sense, an extension of) the much revered Counting Approach of Alkire and Foster (2011), which is sketched, and compared to the partial order approach, in Paragraph 4. Some computational issues, that must be considered when applying partial order tools to data analysis, are discussed in Paragraph 5. Finally, Paragraph 6 concludes.

What Is a Partially Ordered Set?

Partially ordered sets are among the most common and ubiquitous mathematical objects used (albeit implicitly) in daily life. To make a simple example, suppose you are to decide about your holidays. It is likely that you do prefer some locations to others, but that in some cases you may not express any preference, leaving the alternatives unordered. In other words, you are not able to rank all of the locations under consideration, and some of them remain "incomparable". Holiday destinations, partly ranked according to your choice criteria, are an example of a partially ordered set. A typical example in the social sciences is when a set of statistical units are scored against a set of ordinal attributes and resulting achievement profiles have conflicting scores. An individual A owing a house, a car and a telephone may be classified as "better" (whatever this means in specific contexts) than an individual B possessing a house, a telephone, but no car. However, what if individual A has no telephone? In this case, A and B have conflicting achievements, they are no more comparable and cannot be trivially ranked in terms of (say) well-being. This example is the prototype of what we are going to handle in the rest of the chapter. More generally, however, partially ordered sets arise from any system of attributes, comprising ordinal, numerical or mixed data types, when conflicting scores exist in the profiles of statistical units. This motivates the interest towards partial orders and their potential relevance in data analysis.

Formal Definitions

Let X be a set and let "$\leq$" be a binary relation on it. Binary relation "$\leq$" is called a *partial order* if it satisfies the properties of *reflexivity* ($x \leq x$, for any x in X), antisymmetry (if $x \leq y$ and $y \leq x$ then $x = y$) and *transitivity* (if $x \leq y$ and $y \leq z$ then $x \leq z$). The set X endowed with relation $\leq$ is called a *partially ordered set*, or a *poset* for short (Neggers and Kim 1998, Davey and Priestely 2002). In this chapter,

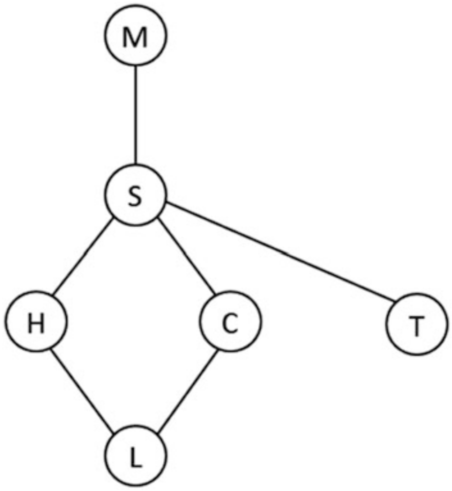


Fig. 8.1 Holiday locations: example of a Hasse diagram on six elements

we will consider only partial orders on finite sets. In this case, partially ordered sets can be conveniently depicted by means of Hasse diagrams, which are a kind of directed graph. Consider again the holiday example, and suppose there are six possible locations: mountain (M), seaside (S), hills (H), lake (L), countryside (C) and town (T). A possible poset on {M, S, H, L, C, T} is depicted in Fig. 8.1, in terms of its Hasse diagram. The diagram must be read from top to bottom; if a downward path links element x to element y, then x and y are *comparable* (and, in this example, x is preferred to y); otherwise, x and y are *incomparable*. M is the preferred destination (it is preferred to any other location); S is preferred to H, L, C and T; H and C are preferred to L and incomparable with T which, in turn, is incomparable with L. A finite poset may have a *maximum* element (an element which is above any others, like M in the example) and a minimum element (an element which is below any others). More generally, finite posets always have *maximal* elements (elements which are below no other elements) and *minimal* elements (elements which are above no other elements, like L and T in the example of Fig. 8.1). A subset of poset elements which are mutually comparable is called a *chain* (e.g. M, S, H and L in the holiday example). Conversely, a subset of poset elements which are mutually incomparable is called an *antichain* (e.g. H, C and T in the example). A poset which is a chain is also called a *linear* (or *complete*, or *total*) *order*.

Most of the posets in the social sciences arise from so-called multi-indicator systems. A *multi-indicator system* (MIS) is simply a set of indicators or attributes (no matter whether ordinal, numerical or even of mixed types) defined on a set of statistical units. Each unit is thus assigned a profile over the MIS, i.e. a sequence of scores on the set of indicators. In general, the resulting set of profiles can be ordered only partially, since conflicting scores in two profiles may arise, leading to incomparability. To be concrete, consider the following construction. Let A_1, A_2 and A_3 be three ordinal attributes, each scored on a four-degree ordinal scale: “1 - low”, “2

- medium", "3 - high", "4 – very high". To give a social meaning to these attributes, let A_1 refer to *schooling level*, A_2 to *social participation* and A_3 to *work stability*, so that the MIS globally refers to *social integration*. Combining the scores, $4^3 = 64$ different achievement profiles are obtained. They can be partially ordered according to the following natural criterion:

$$\mathbf{p} \leq \mathbf{q} \text{ if and only if } p_i \leq q_i \text{ for } i = 1, 2, 3 \tag{8.1}$$

where $\mathbf{p}$ and $\mathbf{q}$ are two profiles over A_1, A_2 and A_3 and p_i, q_i are their scores on A_i. If $\mathbf{p} \leq \mathbf{q}$ and $p_i < q_i$ on at least one attribute A_i, then we write $\mathbf{p} < \mathbf{q}$. The set of profiles Π, partially ordered by (), can be called the "achievement poset" (Fattore 2015). Its Hasse diagram is depicted in Fig. 8.2. The diagram has both a maximum (or *top*) profile, namely 444, and a minimum (or *bottom*) profile, i.e. 111. For future reference, in Fig. 8.3 we also de pict the Hasse diagram of the set of 8 profiles defined over three binary attributes.

In MISes, more statistical units may usually share the same profile. Thus, to each element of the partial order one can associate the corresponding number of

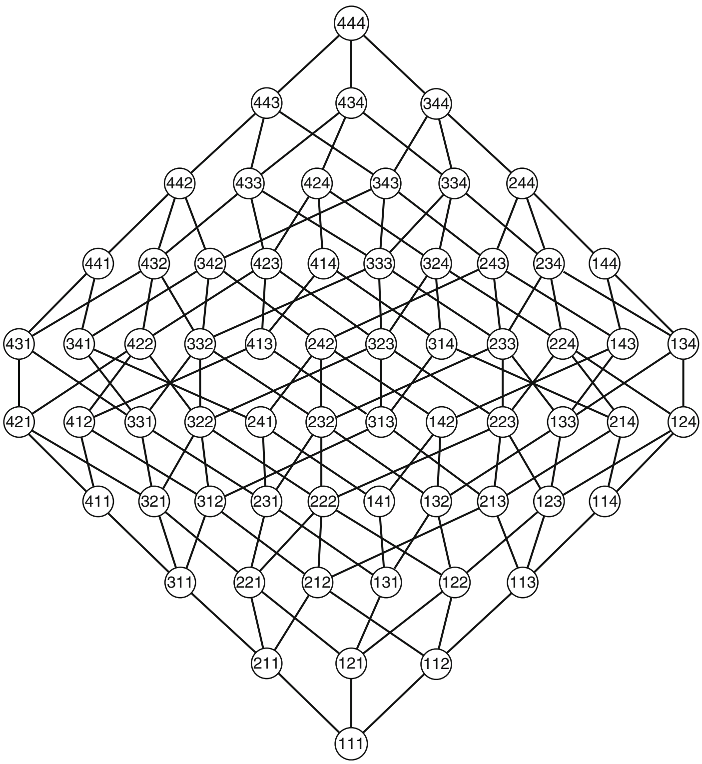


Fig. 8.2 Hasse diagram of the achievement poset Π on three attributes measured on four-degrees scales

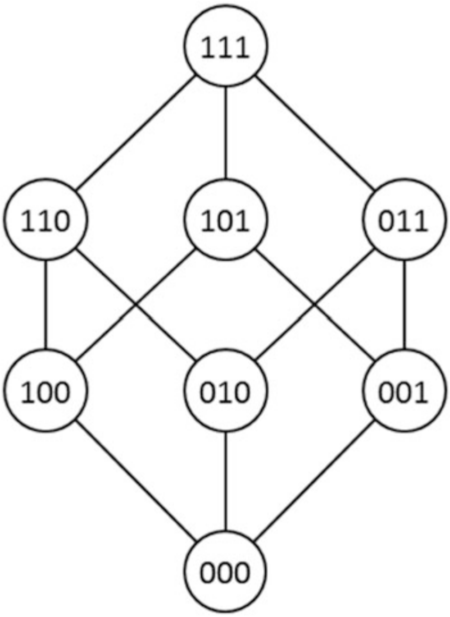


Fig. 8.3 Hasse diagram of the achievement poset on three binary attributes

statistical units, defining a frequency distribution over the poset. The underlying partially ordered set is thus the *ground set* of the frequency distribution; the fact that this set is ordered only partially must be taken into account properly, when working out synthetic indicators.

Before showing, in the next paragraph, how posets may be actually employed for computing indicators, we provide some general comments. Informally speaking, a poset comprises two kinds of information about its elements. First, there is a "vertical" information related to the existence of comparabilities. For example, one can assert that profile 243 in Fig. 8.2 is "better" (in a social integration perspective) than profile 133. Comparabilities are thus linked to the notion of "intensity" of the trait of interest. On the other hand, incomparabilities reveal the existence of "ambiguities" within the set of poset elements. This "horizontal" information is directly linked to the vagueness (or even fuzziness) of social comparisons and evaluation. Given a profile, say 142, there are many alternative ways for a score configuration to be "different" from it. In a very informal sense, some configurations may be "better" or "worse" then 142, others can be "partly better" or "partly worse" than 142 and this may occur at different degrees. There are nuances and ambiguities that, in posets, reflect in the different "relational position" of elements, within the network of comparabilities and incomparabilities that defines the partial order (and that is made visible in the Hasse diagram). Such considerations should be enough to realize the conceptual (and operative) relevance of Partial Order Theory in the social sciences. Moreover, this point of view is also consistent with the following assertions made by Sen (1992, pages 48–49), who directly refers to partial orders in his discussion about well-being and inequality:

"Indeed, the nature of interpersonal comparisons of well-being as well as the task of inequality evaluation as a discipline may admit incompleteness as a regular part of the respective exercises. An approach that can rank the well-being of every person against that of every other in a straightforward way, or one that can compare inequalities without any room for ambiguity or incompleteness, may well be at odds with the nature of these ideas. Both well-being and inequality are broad and partly opaque concepts. Trying to reflect them in the form of totally complete and clear-cut orderings can do less than justice to the nature of these concepts. [. . .] The use of partial ordering has two different types of justification in interpersonal comparison or in inequality evaluation. First, as has been just discussed, the ideas of well-being and inequality may have enough ambiguity and fuzziness to make it a mistake to look for a complete ordering of either. This may be called the "fundamental reason for incompleteness". Second, even if it is not a mistake to look for one complete ordering, we may not be able in practice to identify it."

In the next paragraphs, we provide examples of how synthetic indicators may be built starting from partially ordered structures. We focus on two main alternatives, the first based on the simple concept of *average rank* and the second based on the somehow more sophisticated notion of *mutual rank probability*. Both of them draw upon the existence of some linear orders naturally associated to posets, the so-called *linear extensions*, that we must introduce first.

Linear Extensions of a Finite Partially Ordered Set

Consider the poset Q depicted in Fig. 8.4a. It is composed of five elements and comprises nine chains and three antichains. It has also two maximal elements and a minimum element. Suppose to order elements of Q into a linear order L, in such a way that if $\mathbf{p} \leq \mathbf{q}$ in Q, then it is also $\mathbf{p} \leq \mathbf{q}$ in L. In other words, suppose to completely order elements {a, b, c, d, e}, respecting the constraints imposed by the partial order relation of Q. The resulting complete order L is called a *linear extension* of Q. L is "linear" since it is a chain; it is also an "extension" of Q, since it turns incomparabilities of Q into comparabilities, i.e., in mathematical terms, it adds comparabilities to the partial order relation of Q. A direct check shows that there are only 5 different linear extensions of Q; they are depicted in Fig. 8.4b. The most interesting property of linear extensions is that their intersection (i.e. the set of comparabilities they have in common) equals the original poset Q. In other words, Q comprises all and only those comparabilities its linear extensions "agree upon". Moreover, no other poset has the same set of linear extensions of Q. So, in practice, its linear extensions identify uniquely Q. This is a special case of a general theorem (Schroeder 2002), stating that two finite posets have different sets of linear extensions and that any finite poset is the intersection of its linear extensions. Informally stated, "one partially ordered set" is equivalent to "many non-partially (i.e. linearly) ordered sets", namely linear extensions. The consequence of this result is that one can study properties of a poset by looking at the

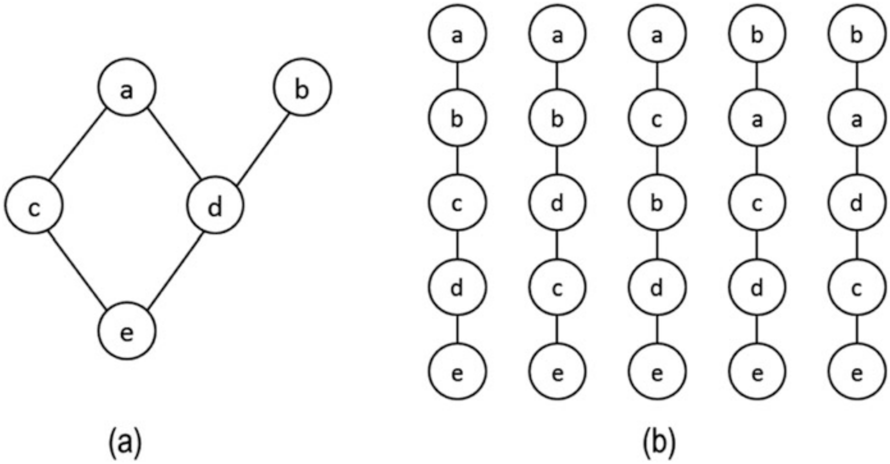


Fig. 8.4 A poset on 5 elements (**a**) and the set of its linear extensions (**b**)

set of such complete orders. Being linear, these are easy to treat and analyze. Linear extensions are thus the building blocks of finite posets and, in a sense, they play the role of univariate variables in the classical composite indicator approach.

Building Synthetic Indicators from Partially Ordered Sets

In applications of composite indicators, the usual aim is to assess the level of a latent trait by aggregating attribute scores. This is often pursued following formative or reflective schemes, in which attributes are composed through linear combinations, i.e. weighted means. As mentioned before, this way of doing may be problematic under different points of view. In particular, the resulting aggregate scores are not always easily interpretable. This issue arises under many forms in different statistical models, involving latent variables for evaluation purposes (think for example to the well-known indeterminacy problems in structural equation models or to the conceptual difficulties in estimating IRT and Rasch models). Basically, the problem is that there are no natural scales against which to measure latent traits, so that "comparisons" (to benchmarks or inter-units) are often more meaningful than "absolute" measurements. In the ordinal case, this evidence is even stronger, due to the non-numerical nature of the data. This does not imply that numerical scores cannot be associated to ordinal MISes and that evaluations cannot be pursued on them; on the contrary, this means that such scores must reflect the results of multidimensional comparisons between ordinal profiles, rather than measurements against an "absolute" scale. Using an expression that will be made clear in the following, synthetic indicators must be built "out of the partial order structure" and not from profiles seen as "stand-alone entities". This is the analogue of involving the covariance structure of a numerical MIS to achieve dimension reduction, the only difference being the "law" defining the "data space", i.e. a partial order relation instead of a scalar product.

Average Rank

In very general terms, evaluation over posets can be often seen as a way to derive a complete order out of a partial order. In fact, once evaluation scores are assigned to poset elements, these can be ordered, producing a linear order (if ties are not present). In many cases, moreover, it is the final ranking, rather than the precise scores, what really matters (e.g. think of a policy-maker that must allocate funds for a social program, to more deprived people).

A very simple way to accomplish this task and to associate to each element of a finite poset P a score representing its "position" on a "low-high" axis is the following:

1. List all the linear extensions of P.
2. For each linear extension, compute the rank of the elements of P (the rank of an element x is defined as 1 plus the number of elements below **p** in the complete order).
3. For each element of P, compute the average rank (*avg*) over the linear extensions of P.

Table 8.1 reports the above computations for the poset depicted in Fig. 8.4a. Clearly, the average rank of a poset element **p**, *avg*.(**p**), is comprised between 1 and the number M of poset elements. The function *avg*.(·) has also the fundamental property of being strictly order preserving, i.e. whenever profile **p** is below profile **q** in the input poset, then *avg*.(**p**) is less than *avg*.(**q**). The proof of this result is trivial, noticing that in any linear extension of P, **p** is ranked below **q** (e.g. elements "a" and "c" in the example). When the poset P is obtained from ordinal variables (as in Fig. 8.2 or Fig. 8.3), it is likely that several statistical units share the same profiles. Each statistical unit is then assigned the average rank of the corresponding profile, so that the final result is not a complete order of units, but an "order with ties" (i.e. a so-called "quasi-order"). Notice also that the information about the average rank, i.e. about the location of profiles, is really extracted out of the partial order relation itself. In fact, the difference in the average ranks of different profiles depends upon their different positions within the Hasse diagram of the poset (see caption of Table 8.1).

The procedure leading to the average rank can be generalized to other parameters, like the median rank. Once the rank distribution over linear extensions is computed for each poset profile, one can in fact easily compute different statistics.

As stated previously, the general problem of associating a rank to elements of a poset P may be basically seen as the problem of picking a particular linear extensions out of the set of linear extensions of P. Computing the average rank is

Table 8.1 Computation of the average rank for the elements of the poset depicted in Fig. 8.4a

Element	Lin. ext. 1	Lin. ext. 2	Lin. ext. 3	Lin. ext. 4	Lin. ext. 5	Average rank
a	5	5	5	4	4	4.6
b	4	4	3	5	5	4.2
c	3	2	4	3	2	2.8
d	2	3	2	2	3	2.4
e	1	1	1	1	1	1

For each element, its rank in each linear extension is reported, together with the corresponding average. Notice that in each linear extension the rank is 1 for the minimum and 5 for the maximum, i.e. ranks are counted from the bottom, according to the definition given in the text. Interestingly, elements c and d have different average ranks, although they share the same level in the Hasse diagram. This is due to the existence of the comparability $d < b$, which makes the positions of elements c and d not "equivalent". This shows that information is extracted from the whole relational pattern of the poset

just a way to achieve this. There are nevertheless many other ways to accomplish the task. For a different, and somehow more complex, procedure with application to decision-making, see Patille and Taille (2004), who propose to derive the final linear order, comparing profile ranking distributions and not just average ranks.

We end this paragraph just citing a different application of linear extensions to the analysis of frequency distribution over posets. In some cases, in fact, it may be of interest to identify "location parameters" of frequency distributions defined over partially ordered sets. For example, in a study on multidimensional deprivation, one may be interested in determining one or more benchmark profiles, (e.g. profiles representing the "median" profile or a specific "quantile profile"), in view of the identification of deprivation thresholds, deprived profiles and, hence, deprived individuals. Due to the existence of incomparabilities, it is in general not possible to define quantiles as we usually do in the unidimensional case. However, one can associate to each profile the "probability" of "being the specified quantile" over the set of linear extensions of the poset. To this goal, it is sufficient to:

1. List the linear extensions of the poset.
2. Identify in each of them the profile representing the quantile of interest, given the frequency distribution over the profiles.
3. Associate to each profile the fraction of linear extensions in which it corresponds to the quantile of interest.

This procedures leads quite naturally to a fuzzy extension of the notion of quantile, as discussed in Fattore (2008).

Evaluation Over Ordinal Multi-indicator Systems

The computation of the average rank relies on counting positions of profiles in linear extensions. We now introduce an evaluation procedure based on the computation of mutual rank probabilities and suitable for evaluation processes based on multidimensional comparisons. Consider the problem of assessing multidimensional deprivation, over a set of k ordinal deprivation attributes. The general procedure followed in poverty studies basically involves (i) the identification of a deprivation threshold and (ii) the subsequent comparison of each statistical unit to it, so as to identify deprived or non-deprived individuals. A measure of deprivation severity may be also computed, in terms of distance of a statistical unit from the threshold. Aggregate indicators, such as Head Count Ratio and Poverty Gap, may then be easily worked out. This logic thread is borrowed from monetary poverty analysis, where the notion of "comparison to a threshold" is, at least conceptually, easy to implement, due to unidimensionality. In a multidimensional setting, there are basically two different ways to pursue this logic. The first is to compute some composite indicator, reducing each k-dimensional profile to a single score and then to apply the above unidimensional procedure. The problems with this approach have been discussed in the Introduction: in the case of ordinal data one cannot sum scores and, in general, weak interdependencies among attributes make compensative procedures scarcely effective and possibly misleading. A second approach is to avoid any preliminary unidimensional reduction and to address deprivation assessment in terms of multidimensional comparisons between units' profiles and some reference profiles, assumed as deprivation benchmarks. The idea of poverty evaluation as "comparison to a benchmark" is kept, but it is turned into practice with no attribute aggregation. This second approach is based on the theory of partially ordered sets and will be sketched in the following. Before introducing it, however, we must cite a third way to perform multidimensional deprivation measurements, namely the Counting Approach of Alkire and Foster, which has gained great popularity in the last years. This procedure, that will be briefly described later, is somehow halfway the composite and the poset approach. In fact, it avoids composing attributes in an aggregate deprivation variable prior to threshold comparison, but achieve this at the cost of reducing the deprivation assessment to a simple counting of binary deprivations over the attributes. As discussed later in the chapter, this leads to lose a great deal of information about deprivation.

The Poset Approach to the Evaluation of Multidimensional Deprivation

We describe the poset approach to deprivation measurement through a simple example. Consider 3 ordinal variables, pertaining to well-being, e.g. *labour status*, *economic status* and *health status*, and suppose to assess each of them on a 3-degree scales (in real applications, scales have usually 4 (or more) degrees, but here we want to keep examples simple, to give visual accounts of the evaluation procedure).

The starting point of the assessment process is to build the "achievement profiles" (i.e. the configurations of individual scores over A_1,A_2 and A_3) and to arrange them into the "achievement poset" Π (Fattore 2015). The Hasse diagram of Π is depicted in Fig. 8.5. The achievement poset simply represents a partial order

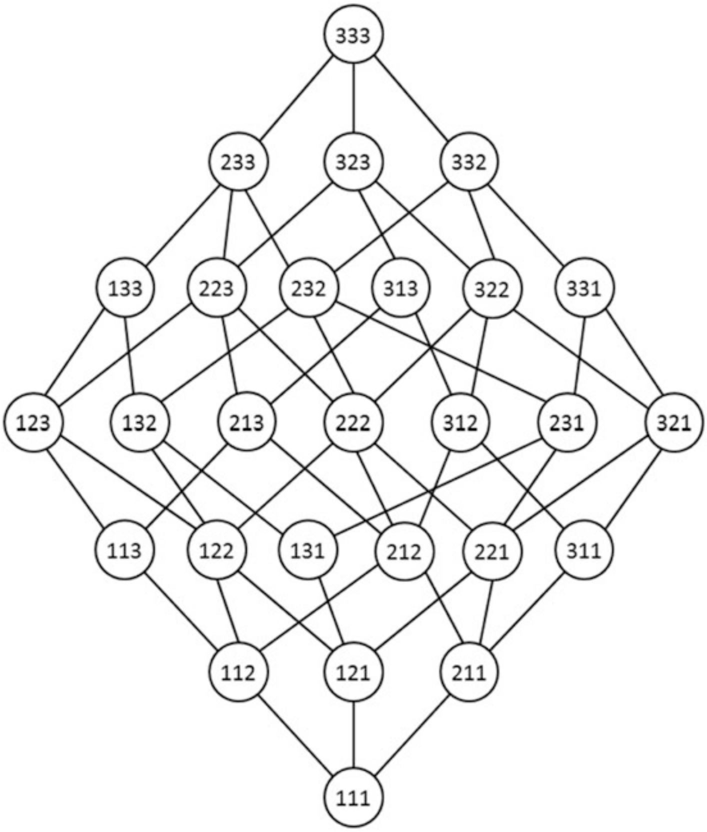


Fig. 8.5 Hasse diagram of the achievement poset over three ordinal attributes on three-degrees scales

relation in which two profiles **p** and **q** are comparable (say $\mathbf{p} < \mathbf{q}$), if the achievements of **p** are lower than or equal to those of **q** and at least one achievement is strictly lower (as in Eq. 8.1). As already noticed, the achievement poset comprises both chains and antichains, reflecting the existence of both a "vertical" dimension and a "horizontal" dimension in profile comparisons. As it will be clear in the following, this leads naturally to a fuzzy perspective on multidimensional deprivation (Qizilbash 2006; Lemmi and Betti 2006).

The achievement poset is a pure mathematical object and, in itself, it conveys no information about the relative deprivation of single profiles. To give it a socio-economic meaning, a threshold must be set, to identify profiles that are "undoubtedly" deprived, leaving to the evaluation procedure to "decide" about the deprivation of all of the other profiles. The identification of such benchmark profiles is left to policy-makers and it is not of concern here. We simply note that more than one profile may be set as deprivation benchmark, since individuals may be deprived in different ways. In practice, deprivation benchmarks constitute an antichain of the

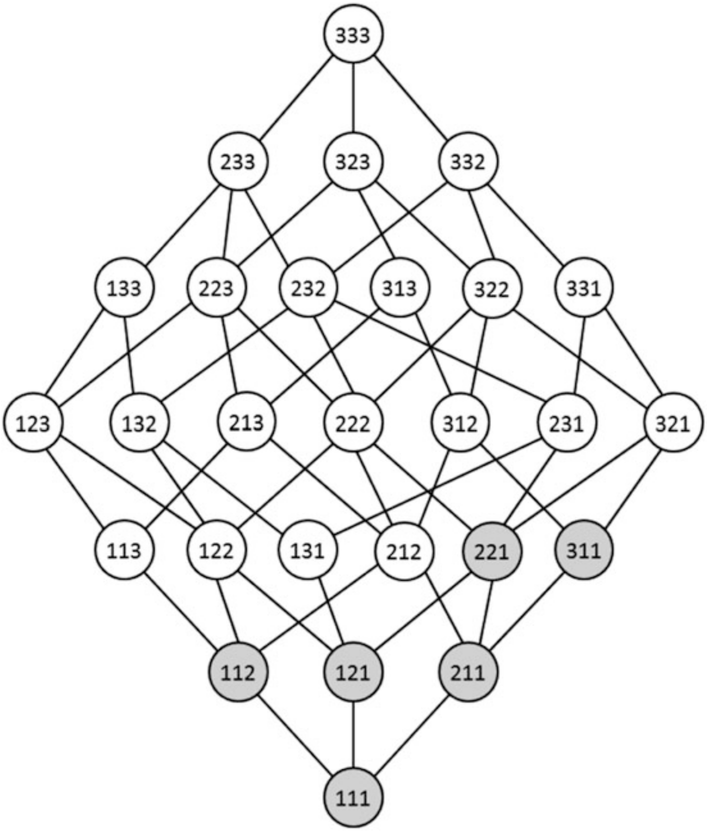


Fig. 8.6 Hasse diagram of the achievement poset of Fig. 8.5 with deprivation threshold {112, 221, 311}. In grey, completely deprived profiles, i.e. profile on or below the threshold antichain

achievement poset; such an antichain is the analogous of the deprivation threshold in classical approaches (Fattore 2015; Fattore et al. 2011, 2012). Figure 8.6 depicts the achievement poset with a possible threshold, namely the antichain {112, 221, 311}.

Setting the threshold is the way a "minimum" amount of information is injected into the evaluation procedure. Deprivation benchmarks represent deprived profiles; consequently, profiles below any threshold element (i.e. the so-called *down-set* of the threshold, see again Fig. 8.6) are deprived as well (since their achievements are worse than those of one or more deprived profiles). The deprivation status of all of the other profiles of the achievement poset must instead be assessed in an indirect way, considering their "relational position" in the Hasse diagram, with respect to the threshold antichain.

Formally, the aim is to define an *identification function* idn(·) over Π with values in [0,1], measuring to what degree an achievement profile is deprived, *given* the threshold. Notice that the identification function does not measure deprivation intensity, but the membership of a profile to the set of deprived profiles. In this respect, the evaluation procedure is fuzzy in nature, as anticipated. The most naturally way to define the identification function exploits again the set of linear extensions of the achievement poset. In brief, the idea runs as follows:

1. List all the linear extensions of the achievement poset.
2. Being linear orders, in each linear extension the classification of a profile as deprived or not is unambiguous. In fact, given a linear extension h, either a profile $\mathbf{p}$ is above all the elements of the threshold in h, or it coincides with or is below (at least) one element of the threshold. In the first case, the profile is classified as not deprived in h; in the second case it is classified as deprived.
3. Counting the fraction of linear extensions where a profile $\mathbf{p}$ is classified as deprived, one gets the desired identification score of the profile.
4. The identification score of a profile is then inherited by all statistical units sharing that profile, providing a statistical distribution of deprivation degrees over the population.
5. Once identification scores are obtained for each profile and statistical unit, usual aggregate indicators over the population can be computed.

The resulting identification function for the Hasse diagram of Fig. 8.6 is depicted in Fig. 8.7. As it can be noticed by direct inspection, it is not linear (it somehow resembles the shape of a sigmoid function, which is typical of measurement devices) and it is not constant over the levels of the achievement poset. In fact, the selection of the threshold breaks the symmetry of the poset, making profiles on the same level "non-equivalent" (or "non-equideprived" so as to say). The identification function equals 1 (full deprivation) on the profiles on or below the threshold and equals 0 (no deprivation) on profiles which are above all the elements of the threshold. All of the other profiles get scores between 0 and 1, according to the fuzzy spirit of the evaluation exercise.

A few other comments are in order. First, the identification function is (anti-) order preserving, i.e. if profile $\mathbf{q}$ is above profile $\mathbf{p}$ in the achievement poset, then $\mathrm{idn}(\mathbf{q}) \leq \mathrm{idn}(\mathbf{p})$. In other words, if achievements improve, the deprivation score cannot increase. Second, the identification function is defined without aggregating attribute scores, but counting over linear extensions. Linear extensions act as classifiers; the fuzziness and nuances of deprivation turn into the existence of classifiers classifying the same profile differently. Using the same principle as above, one can also compute a measure of deprivation severity averaging, over linear extensions, the distance of a profile from non-deprivation in each linear extension. Details can be found in (Fattore 2015), where it is also shown how exogenous information on the importance of deprivation attributes can be introduced into the evaluation procedure, without employing numerical weights. As a result, one gets a complete procedure for the measurement of deprivation (or of other latent traits) on ordinal MISes. The evaluation procedure is "totally ordinal", involving no scaling procedures and no numerical weighting; still, it sticks to the logic thread of any evaluation process based on the "comparison to a threshold" principle.

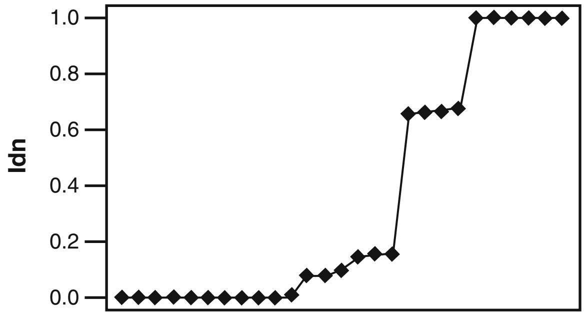


Fig. 8.7 Identification function (deprivation degrees) for the poset depicted in Fig. 8.6. Each dot corresponds to a profile (profiles are listed on the abscissa in increasing identification score; score configurations of poset elements on the abscissa are not reported)

Partial Order Theory and the Counting Approach of Alkire and Foster

Among the attempts to overcome the composite indicator paradigm for evaluation purposes on multi-indicator systems, the most relevant, at least in terms of its spreading across the social science community, is the so-called Counting Approach of Alkire and Foster (2011). In a few years, this procedure has become the reference one for the assessment of multidimensional poverty, mainly in poor and developing countries. Given its relevance, it is instructive to sketch and compare it to the partial order approach (a full discussion of the Counting Approach can be found in the cited reference).

With reference to deprivation, given a multi-indicator system comprising k attributes $A_1,..,A_K$, one selects a set of k cut-offs $c_1,...,c_K$ defining k deprivation thresholds, one for each deprivation attribute. Statistical units whose achievement on attribute A_i is less or equal to c_i are declared deprived on that attribute. This way, each unit is associated a binary profile on k binary attributes $B_1,..,B_K$ which are the binary counterparts of $A_1,..,A_K$ (i.e. a binary profile scores 1 on B_i if the original profile on deprivation attributes scores not higher than c_i on A_i). Next, an overall cut-off c is set between 1 and k: statistical units whose number of deprivations is at least equal to c are finally declared as deprived. The Alkire-Foster identification process is thus composed of two subsequent steps: first, attributes (and profiles, as well) are "binarized"; second, deprivations are counted. A number of synthetic indicators may then be computed, to get an overall picture of deprivation in the population of interest. In particular, and in addition to the Head Count Ratio (fraction of poor individuals), the Counting Approach allows for a measure of deprivation intensity (obviously linked to the number of deprivations suffered by statistical units). It is also possible to account for attribute relevance, introducing a set of k weights, to be attached to attributes when deprivation intensity is measured (technical details may be found in the cited reference).

What is more interesting here, is to compare the Alkire-Foster procedure to the partial order approach. Although both start by considering the achievement profiles of statistical units, the two methodologies differ in the way the evaluation exercise is addressed, ordinal data are treated and deprivation information is extracted. As a matter of fact, the Counting Approach sticks to a more classical view on deprivation assessment, in fact:

1. It conceives global deprivation as the co-occurrence of deprivations over single attributes, i.e. it is "attribute-driven", rather than "profile-driven". Information is not extracted from the "relational pattern" (i.e. from the achievement poset), but from a global feature (number of deprivations) of profiles, viewed as "stand-alone entities".
2. It is aggregative-compensative in nature: deprivations are simply "summed up" in a "compositional" way.
3. It circumvents the problem of properly treating ordinal attributes by "binarizing" them. However, by collapsing ordinal features into binary variables, it loses a great deal of information about deprivation. This in turn leads to a crisp picture of it, which does not reflect its possible vague or fuzzy nature properly.
4. Attribute importance is accounted for by using weights. Not only weighting schemes may be too rigid to reflect preferences or policy priorities (for example, using weights it is impossible to set an attribute as indifferent to two other attributes, the first of which is more relevant than the second); it is also inconsistent when applied to ordinal, and even binary, attributes.

To show explicitly which kind of "reductions" and information losses are implied by the binarization/counting procedure of Alkire and Foster, it is useful to cast it into partial order terms. Let us consider again the deprivation example of Fig. 8.5 and let us put, for matter of convenience, $c_1 = c_2 = c_3 = 2$ and $c = 2$ (i.e. a statistical unit is deprived on an attribute, if and only if its achievement on that attribute is equal or lower than 2; a statistical unit is identified as deprived if it has at least two deprivations). Turning attributes A_1, A_2, A_3 into their binary counterparts B_1, B_2, B_3 according to the attribute cut-offs, the set of possible achievement profiles reduces to just $2^3 = 8$ elements (compared to 27 original profiles of the "full" achievement poset). The Hasse diagram of such a "reduced" achievement poset is that depicted in Fig. 8.3, it is also replicated in Fig. 8.8, where the entire binarization process implied by the Alkire-Foster procedure is represented. Comparing the diagrams of the "full" and "reduced" achievement posets clearly reveals the approximations involved in the binarization step and in the final deprivation counting. In particular, it is noticeable that achievement profiles are either classified as deprived or not; nuances are not taken into account and the resulting identification function is binary (deprivation scores are either 0 or 1), ruling out any fuzzy perspective in the assessment exercise. In fact, all profile incomparabilities are resolved by means of the counting procedure: profiles with the same number of deprivations (i.e. of 1 s in the diagram of Fig. 8.3, reported also in Fig. 8.8) are collapsed into the same "equivalence" class and classes are then accordingly arranged in a complete order (Fig. 8.8).

To conclude, the Counting Approach suffers of many drawbacks typical of the composite indicator methodology, like other counting approaches (Cerioli and Zani 1990). But it has its merits too. It is very simple to apply, posing no computational problems even with a high number of attributes. In a sense, it contributes to

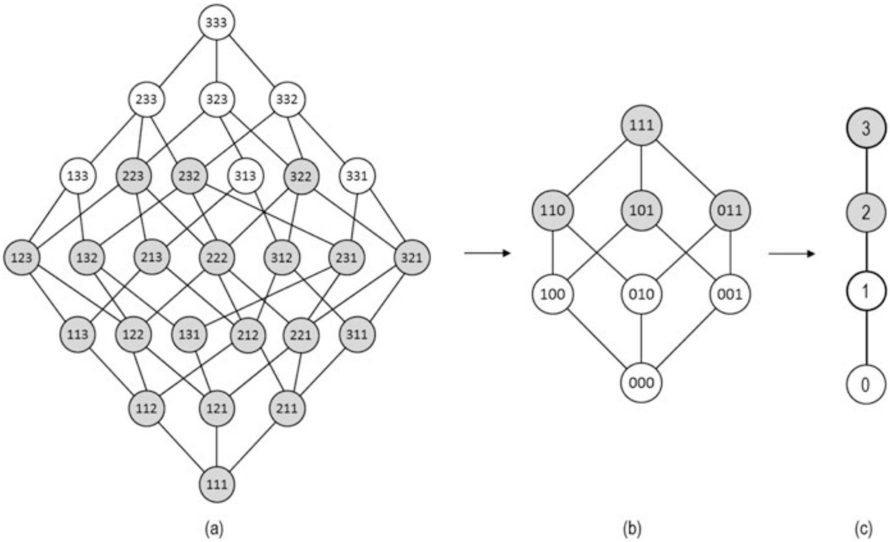


Fig. 8.8 Binarization process in the Alkire-Foster Counting Approach: (**a**) original ("full") achievement poset; (**b**) binarized achievement poset according to overall cutoff $c = 2$; (**c**) total order corresponding to deprivation counting. Grey circles identify deprived profiles, according to the cut-offs selected, in the full poset, in the reduced poset and in the final linear order of countings. Notice that in the reduced poset, deprivation increases moving upwards along chains (while in the full achievement poset, it increases moving downwards)

overcoming the classical composite paradigm: even if the "binarization" procedure is unsatisfactory under many perspectives, it partly supports the idea that ordinal variables cannot be treated as they were cardinals. Its wide application to multidimensional deprivation measurement may be a step forward, in view of more effective and powerful methodologies, such as that briefly illustrated in this chapter.

Computational Aspects

The procedures described in previous paragraphs require suitable algorithms to be implemented and applied. In principle, the exact computation of average ranks or evaluation functions requires the extraction of all the linear extensions of the achievement poset. In real posets, however, this is usually unfeasible, since the set of linear extensions is likely to be composed of a too much large number of elements. In practice, one samples linear extensions (Bubley and Dyer 1999), computing ranks and evaluation functions only in an approximated way. The sampling procedure and all of the facilities needed to compute average ranks and to run the evaluation procedure described above (and also to apply the Counting Approach, if needed) are collected in the R Core Team (2012) package PARSEC, freely available to R users. Some details on it and its performances may be found in Arcagni and Fattore (2014). Anyway, the package and its full documentation may be downloaded from the CRAN website and may also be accessed through R help facilities. Notwithstanding the use of sampling procedures, there are cases where the computation time may be too long. For this reason, simplified evaluation procedures are being developed (Fattore and Arcagni 2016), to dramatically reduce computational problems, allowing the applicability of poset-based procedures to larger multi-indicator systems (and posets in general). Other software packages for managing posets do exist, in particular the Reader may consider the PyHasse software (www.pyhasse.org).

Conclusion

In this chapter, we have shown how adopting the proper formal setting, it is possible to build evaluation procedures and synthetic indicators starting from multidimensional systems of ordinal data. At the heart of the procedure, there is the theory of partially ordered sets, a branch of discrete mathematics properly designed to treat ordinal data and order relations. A part from technicalities, the chapter shows that the road to synthesis need not pass through aggregations. This is not only fundamental when dealing with ordinal data that, even in principle, cannot be aggregated using classical composite indicators. It is of great importance whenever attributes, either ordinal or cardinal, are only weakly interdependent, so that dimension reduction tools prove ineffective, due to the intrinsic complexity of the latent traits of interest. The focus of the chapter is on the computation of synthetic indicators for evaluation purposes. But the use of Partial Order Theory in the social sciences is surely not confined to evaluation goals. Partial Order Theory may be also applied to perform sensitivity analysis, to identify relevant attributes (i.e those attributes that have the stronger impact on the structure of a partial order on a set of statistical units), to analyze patterns of deprivation (Annoni et al. 2011) or of other latent traits. Another relevant problem that can be addressed through partial order tools pertains to the measurement of inequality on multidimensional systems of ordinal indicators. The problem, which is very important in a “beyond GDP” perspective, is still open and different proposals to solve it may be found in literature, mainly based on aggregating inequalities over single attributes. In a poset setting, the measurement of multidimensional inequality is instead more naturally addressed in terms of inequality on linear extensions. All in all, Partial Order Theory represents a powerful “language” and a “rich toolbox” for social scientists. In particular, what is of interest here is to stress the possibility to overcome the composite indicator paradigm, in favor of new approaches that stick more tightly to the very nature of the problems at hand and of the available data. New concepts and new tools to try to properly address the description and the evaluation of societal phenomena, in complex societies where old statistical paradigms may fail.

References

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95*(7–8), 476–487.

Annoni, P., & Bruggemann, R. (2009). Exploring partial order of European countries. *Social Indicators Research, 92*, 471–487.

Annoni, P., Fattore, M., & Bruggemann, R. (2011). A multi-criteria fuzzy approach for analyzing poverty structure. *Statistica & Applicazioni – Special Issue, 2011*, 7–30.

Arcagni, A., & Fattore, M. (2014). PARSEC: An R package for poset-based evaluation of multidimensional poverty. In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *MultiIndicator systems and modelling in partial order*. Berlin: Springer.

Bruggemann, R., & Carlsen, L. (2014). Incomparable-what now? *MATCH Communications in Mathematical and in Computer Chemistry, 71*, 699–714.

Bruggemann, R., & Patil, G. P. (2010). Multicriteria prioritization and partial order in environmental sciences. *Environmental and Ecological Statistics, 17*, 383–410.

Bruggemann, R., & Patil, G. P. (2011). *Ranking and prioritization for multi-indicator systems – introduction to partial order applications*. New York: Springer.

Bruggemann, R., & Voigt, K. (2012). Antichains in partial order, example: Pollution in a German region by lead, cadmium, zinc and sulfur in the herb layer. *MATCH Communications in Mathematical and in Computer Chemistry, 67*, 731–744.

Bruggemann, R., Pudenz, S., Voigt, K., Kaune, A., & Kreimes, K. (1999). An algebraic/graphical tool to compare ecosystems with respect to their pollution. IV: Comparative regional analysis by Boolean arithmetics. *Chemosphere, 38*, 2263–2279.

Bubley, R., & Dyer, M. (1999). Faster random generation of linear extensions. *Discrete Mathematics, 201*, 81–88.

Carlsen, L., & Bruggemann, R. (2014). The "failed state index": Offers more than just a simple ranking. *Social Indicators Research, 115*, 525–530.

Cerioli, A., & Zani, S. (1990). A fuzzy approach to the measurement of poverty. In C. Dagum & M. Zenga (Eds.), *Income and wealth distribution, inequality and poverty*. Berlin: Springer.

Davey, B. A., & Priestley, B. H. (2002). *Introduction to lattices and order*. Cambridge: CUP.

Fattore, M. (2008). Hasse diagrams, poset theory and fuzzy poverty measures. *Rivista INternazionale di Science Sociali, 1*, 63–75.

Fattore, M. (2015). Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research, 128*(2), 835. doi:10.1007/s11205-015-1059-6.

Fattore, M., & Arcagni, A. (2016). A reduced posetic approach to the measurement of multidimensional ordinal deprivation. *Social Indicators Research* (to be published)

Fattore, M., & Maggino, F. (2014). Partial orders in socio-economics: A practical challenge for poset theorists or a cultural challenge for social scientists? In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order*. Berlin: Springer.

Fattore, M., Bruggemann, R., & Owsinski, J. (2011). Using poset theory to compare fuzzy multidimensional material deprivation across regions. In S. Ingrassia, R. Rocci, & M. Vichi (Eds.), *New perspectives in statistical modeling and data analysis*. Berlin: Springer.

Fattore, M., Maggino, F., & Colombo, E. (2012). From composite indicators to partial order: Evaluating socio-economic phenomena through ordinal data. In F. Maggino & G. Nuvolati (Eds.), *Quality of life in Italy: Research and reflections*, *Social indicators research series 48*. New York: Springer.

Lemmi, A., & Betti, G. (Eds.). (2006). *Fuzzy set approach to multidimensional poverty measurement*. New York: Springer.

Madden, D. (2010). Ordinal and cardinal measures of health inequality: An empirical comparison. *Health Economics, 19*, 243–250.

Neggers, J., & Kim, S. H. (1998). *Basic posets*. Singapore: World Scientific.

Patil, G. P., & Taillie, C. (2004). Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. *Environmental and Ecological Statistics, 11*, 199–228.

Qizilbash, M. (2006). Philosophical accounts of vagueness, fuzzy poverty measures and multidimensionality. In A. Lemmi & G. Betti (Eds.), *Fuzzy set approach to multidimensional poverty measurement*. New York: Springer.

R Core Team. (2012). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/

Schroeder, B. (2002). *Ordered set. An introduction*. Boston: Birkhauser.

Sen, A. (1992). *Inequality reexamined*. Cambridge: Harvard University Press.

Chapter 9
The Role of Extended IRT Models for Composite Indicators Construction

Michela Gnaldi, Simone Del Sarto, and Filomena Maggino

Introduction

A key purpose of measurement in social sciences is to identify, quantify and possibly explain the differences that exist between units of analysis, that is, individuals, students, Countries, patients, customers, etc., on the basis of some characteristics. These differences contribute to score variability and are the basis of much information. When scientists conduct studies of quality of life, development, or poverty, they are trying to identify and measure the differences that exist among units in order to understand the reasons of such differences. Similarly, when psychometricians study learning competences, intelligence, or happiness, they are attempting to detect differences from unit to unit (i.e., students, people, patients), and from time to time.

It is another matter of fact that in the social sciences, including psychological and educational sciences, unidimensional measurements are the norm. That is, the differences among units are mostly identified on the basis of a single variable or, alternatively, on the basis of a single dimension which synthesises a number of single variables. Although unidimensionality is the most common assumption in the analysis of social, educational and psychological issues, unidimensionality of data usually cannot be met because of the complexity inherent to society.

M. Gnaldi (✉)
Department of Political Sciences, University of Perugia, Perugia, Italy
e-mail: michela.gnaldi@unipg.it

S. Del Sarto
INVALSI (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione), Rome, Italy

F. Maggino
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_9

Using unidimensional measurement models can be problematic when the data at issue is complex and multifaceted. In these latest contexts, applying a unidimensional measurement to multidimensional data necessarily causes mismatch between the model and the data (Bonifay et al. 2015). This last side effect occurs whenever one tries to synthesise social phenomena through a single numeric value, as it happens when placing units of analysis in a unique position within a single ranking. By and large, building rankings means to reintroduce unidimensionality for the assessment of issues which are multidimensional instead.

There are many examples of multidimensional issues treated as unidimensional. For instance, ranking Countries on the basis of their (implied unidimensional) development; ranking students on the basis of their (implied unidimensional) mathematical ability; ranking universities on the basis of their (implied unidimensional) quality, and so forth. However, neither of them can be reasonably considered as a unidimensional issue.

Placing the single units of analysis into a single position within a rank has the obvious advantage of simplicity. However, it also means loosing information on the performances of the single units on the different dimensions which contribute to characterise a multidimensional phenomenon. We believe that a less stringent and more flexible method would lead us to a classification of the single units, instead of ranking them. Classifying, differently from ranking, is the process of identifying to which of a set of categories, or sub-populations, a unit belongs to, without placing it into a specific relative position in a ranking.

In this contribution we argue that *i*. retaining the multidimensional structure of the data is necessary to account for diversity, and that *ii*. classifying units in homogeneous groups is more flexible than ranking to identify and quantify the differences among units of interest. It follows that, when evaluating complex phenomena, both objectives have to be pursued. In other terms, it is both necessary to seek information on the different dimensions which contribute to characterise a multidimensional phenomenon, and contextually, classify the units in homogeneous classes as regards to these relevant latent dimensions. When assessing individual units' differences, seeking these two aims concurrently provides a much richer information than ranking units on a single metric.

In this Chapter, we claim that the most suited statistical models in the research contexts at issue are those which allow us to gain such a twofold purpose, that is, clustering (or grouping) units into homogeneous classes, by retaining at the same time the multidimensional facet of the analysed phenomena. One of such methods is the Latent Class (LC) multidimensional Item Response Theory (IRT) model (Bartolucci 2007), that we apply in the case study section of the present Chapter. In the application, we use data from the Italian Multipurpose Survey, and specifically the part of it reserved to peoples' habits as regards to their spare time and holidays. By applying the above mentioned method, we provide a classification of our sample units (i.e., individuals) in latent classes based on the posterior probabilities of belonging to each latent class, on one side; on the other side, we provide a clustering of the variables into multiple dimensions, which correspond to disjoint groups of variables, among those here considered within the Multipurpose survey. The main output of such an analysis is a cross tabulation which shows the estimated ability parameters for each latent class (of individuals) and dimension (of variables). Each of these estimates can be interpreted as the probability for an individual grouped in one of the latent classes to present a certain dimension of the analysed phenomenon, which in our case concerns a specific dimension of habits as regards to spare time and holidays. Therefore, this method allows us to characterise the different latent classes with respect to the selected dimensions. The method can be extended to other relevant applicative settings, when one aims at classifying (instead of ranking) other units of analysis based on some issues of interest.

Methodological Concerns in the Construction of Composite Indicators

Composite indicators (CIs) are the conventional approach to socio-economic evaluation. CIs combine a number of individual indicators, each assessing a dimension, into a single metrics. As such, they are agreed as being expressive of the performance of the single units considered (Munda et al. 2009; Zhou and Ang 2009). Likewise, CIs are a widespread communication and political tool for conveying summary performance information, as they combine a large amount of information in a format that is easily understood. However, many critical issues affect their computation. Specifically, in the construction of CIs, several methodological issues need to be carefully addressed if the results are not to be misinterpreted and manipulated (Gnaldi and Ranalli 2016). The OECD (2008) underlines that CIs development involves stages where a number of subjective judgements have to be made, from the selection of individual indicators, to the choice of normalisation methods, weighting schemes and aggregation models. Data selection is a key stage in the process of CI construction, since the quality of composite indicators largely derives from the quality of the underlying variables. Thus, individual indicators should be chosen on the basis of their relevance, analytical soundness, timeliness and accessibility. Detection of missing data is also important to guarantee the development of robust CIs (i.e., assessing the impact of the imputation on the composite indicator results, spotting the presence of outliers in the dataset). The underlying nature of the data, the existence of relationships among individual indicators have also to be accounted for through opportune multivariate techniques. Normalisation is required prior to any data aggregation, as the indicators in a dataset often have different measurement units. Thus, normalisation is needed to allow scale adjustments and to transform skewed indicators. When constructing composite indicators, particular attention has to be paid to the weighting process, which gives different importance to the single indicators forming the composite. A number of weighting techniques exists and weights can be chosen according to statistical methods and/or expert opinions. Aggregation methods also vary: while the linear aggregation method is useful when all individual indicators have the same measurement unit, geometric aggregations are better suited when non-compensability between individual indicators or dimensions is required.

It is argued (Saisana et al. 2005; Saltelli 2007; Saisana and D'Hombres 2008; Munda et al. 2009) that all these judgements (e.g., the selection of individual indicators, the choice of normalisation methods, weighting schemes, aggregation model, etc.) are potential sources of uncertainty that affect both the variance of the CIs and the variability of any ranking based on CIs. In other words, it is expected that the choice to include a particular indicator in the CI, or the choice to employ a normalisation scheme (rather than a different one) can have an impact on the rankings of the individual units within the composite (Gnaldi and Ranalli 2016).

The soundness of a CI depends not only on the methodological choices taken in the process of construction of a CI. A further issue on the use of CIs and rankings is related to the fact that they qualify the single units by means of a single and monotonic scale. If the concept being measured turns out to be actually unidimensional, computing a single composite indicator could be justifiable. But when concepts are truly multidimensional, then singling out just a composite indicator is questionable (Fattore et al. 2012).

Beyond methodological issues, the construction of composite measures and rankings raises a number of social and policy issues as well. For example, some authors (Amsler and Bolsmann 2012; Amsler 2014) discuss the social and policy limitations of rankings and league tables highlighting a tendency, in the predominant discourse on ranking, to consider rankings as natural and inescapable entities, to represent them as matters of fact (rather than matters of concern), and to celebrate the neoliberal values of individualised competition.

The Multidimensional Nature of Phenomena and the Need to Account for It

There is a shared acknowledgement of the difficulty of conveying into unidimensional indicators all the relevant information pertaining to complex phenomena. Several phenomena in the socioeconomic and related fields (i.e., psychometrics and behavioural fields) are, in fact, inherently complex and complexity implies multidimensionality. For instance, development is a complex key goal, which has been regarded as unidimensional until the 60s, when the multidimensional approach moved away from the traditional focus on the monetary single dimension as a sufficient proxy of human welfare (Alkire and Foster 2007). As another example in the field of higher education, university quality is acknowledged as being a multidimensional phenomenon. Higher education institutions are engaged in a variety of activities and target a number of different specific objectives, and this makes it difficult to condense the diversified work going on within them into a single numeric value, or position within a ranking. Further, the quality of education passes through, among other things, students achievement of the targeted competences at the relevant school levels. Students competences and abilities are, themselves, complex and multidimensional variables. For example, in the field of mathematics education, experts (Schoenfeld 1992; Bartolini Bussi et al. 1999; Douek 2006) identify in understanding, problem solving and reasoning the three main complex constructs involved in mathematics ability. This essentially means that mathematics proficiency is a complex target and students level of mastery of mathematics cannot be addressed through a single score.

The previous are just a few of the many examples of multidimensional issues taken from different disciplines, which might be taken as instances of the complex targets they mirror. Neither of the previously mentioned phenomena can be reduced to a single dimension. Development cannot be reduced to a single monetary dimension. Students mathematics ability cannot be forced into problem solving alone. University quality cannot be expressed as merely research quality. As a consequence, it is not reasonable to quantify and identify the differences among units on the above mentioned characteristics based on a single number, composite, or position within a rank.

Whatever the real-world ground of application, the need for multidimensional measurement of complex phenomena emerges from a range of imperatives (political, legal, etc.), largely overlapping across fields and ascribable to the need to retain diversity. When scientists conduct studies of quality of life, the demand for multidimensional work primarily results from a mandate to reduce inequality in populations in disadvantageous socioeconomic position, and from work on social exclusion that identifies deprived areas/populations for targeting funds and interventions (Alkire and Foster 2007). As another example, in the field of higher education, the development of multidimensional measures of university quality surfaces from the need to account for diversification among institutions as regards to the two main functions they are called upon to perform, that is, research and education. At least in Italy, there are several higher education institutions showing very different performances with respect to research, as opposed to teaching, so that focusing on a unidimensional measure for excellence would not fully reflect the reality and would risk to flatten deficits and surplus on the different dimensions, thereby reducing the diversity of the whole system (Gnaldi and Ranalli 2016).

A number of issues arises once the multidimensionality of complex phenomena is acknowledged. The first concerns the choice and selection of the relevant dimensions which should be accounted for when studying a multidimensional phenomenon. When the choice and selection of the type and numbers of dimensions is carried out by means of statistical techniques, one has a vast range of available tools for data reduction. The issue is then which is the most appropriate method to infer the dimensions, or factors, which contribute to characterise a complex phenomenon from observed single variables. As known, the most widespread techniques are principal component analysis and factor analysis, see for instance Hotelling (1933), Harman (1976), McDonald (1985). Both are variable reduction techniques and sometimes mistaken as the same statistical method. The first involves extracting linear composites of observed variables. The second is based on a formal model predicting observed variables from latent factors. In the literature, specifically on students' assessment through standardised tests, further approaches are proposed to evaluate the dimensionality of test items. For instance, item factor analysis (Wirth and Edwards 2007), among parametric approaches, generalises traditional factor analysis, based on continuous observed responses, to discrete observed responses. Bi-factor analysis, introduced by Holzinger and Swineford (1937), is another well-known and broadly used form of confirmatory factor analysis, which hypotheses a first general factor – which may be seen as the overall latent ability assessed by a test (i.e., the mathematical ability) – and a few group factors, which may be viewed as more specific sub-components of the overall ability assessed by the test (Golay and Lecerf 2011).

Overall, methods of data complexity reduction concerned with the structure of variables can be explorative and confirmative. Confirmative methods need to specify in advance the number and types of dimensions. Therefore, to apply them, a specific factor structure has to be specified in advance. Such aspecification may be difficult to find in practice, and this happens whenever one does not have prior information on the structure of dimensionality of the data at issue.

While all methodologies have both advantages and limitations, their discussion is out of the scope of this paper. In this contribution, we focus on a technique of variable complexity reduction developed within an Item Response Theory (IRT) framework, that is, a hierarchical clustering algorithm (Bartolucci 2007), which allows us to group variables measuring the same latent construct in the same cluster. As it will be specified further in this Chapter, this algorithm builds a sequence of nested models, starting with estimating the most general model (i.e., a model with a different dimension for each variable) and ending with the most restrictive model (i.e., a model with only one common dimension to all variables).

Ranking the Units or Classifying Them into Homogeneous Groups?

A second key issue concerns the use of the single variables and/or the relevant latent dimensions to compare units (i.e., Countries, schools, students, customers, and so on) in time and/or space in order to make evaluations. In fact, a key purpose of measurement in the social sciences is to identify, quantify and possibly explain the differences that exist among units of analysis. These differences contribute to score variability and are the basis of much information.

The most widespread approach to compare units and identify the differences among them is to combine the various indicators into a composite index. Composite indices have the advantage of allowing the ranking of units through a single number, which reflects the highest or lowest position of single units. The ranking method is simple to understand. Notwithstanding, its simplicity works to its disadvantage, because it implies loosing information on the performances of the single units on the different dimensions which contribute to characterise a multidimensional phenomenon. By and large, building composite indices means reintroducing unidimensionality for the assessment of phenomena which are intrinsically multidimensional.

A less stringent and more flexible method would lead us to a classification of the single units, instead of ranking them. Classifying, differently from ranking, is the process of identifying to which of a set of categories, or sub-populations, a unit belongs to, without placing it into a specific relative position in a ranking.

The classification approach drives us to methods of data complexity reduction concerned with the structure of the units, that is, to cluster analysis techniques. Clustering techniques aim at grouping a set of units in such a way that units in the same group (i.e., a cluster) are more similar to each other than to those in other groups, according to some criteria. There are many different types of clustering techniques, which reflect the many ways one can sort cases into groups. For instance, in connectivity models, hierarchical clustering builds models based on distance connectivity, while in centroid models, the *k*-means algorithm represents each cluster by a single mean vector. Overall, agglomerative clustering approaches are “bottom up” (i.e., each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy), while divisive approaches are “top down” (i.e., all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy).

Latent Class Analysis (LCA) is a further way of identifying subtypes of related cases (or latent classes) from multivariate categorical data, whose results can be used to classify cases into their most likely latent class. LCA is used in way analogous to cluster analysis. That is, given a sample of cases (subjects, objects, respondents, patients, etc.) measured on several variables, one wishes to know if there is a small number of basic groups into which cases fall. LCA is well suited to many application fields, for instance, in health, where one wishes to identify disease subtypes or diagnostic subcategories. Other common areas of application include marketing research, survey research, sociology, psychology and education.

Again, while all methodologies have both advantages and limitations, their discussion is out of the scope of this paper. In this contribution, we focus on a Latent Class approach to classify units within an IRT framework, as it will be discussed further in the following sections.

Combining Multidimensionality with Classification Instances. The Potentials of Extended IRT Methods for CIs Construction

In the previous two sections we highlight that multidimensionality is an inherent characteristic of complex phenomena, and we further stress the consequent need to handle such a complexity with methods of data complexity reduction concerned with *i*. the structure of the single variables on one side (section “The multidimensional nature of phenomena and the need to account for it”), and with *ii*. the structure of the units on the other side (section "Ranking the units or classifying them into homogeneous groups?"). The former methods are best suited for dimensionality assessment and are therefore useful for studying the latent structure of dimensionality of a complex phenomenon, and possibly retain it when making judgments. The latter methods are used for classification purposes, that is, to classify – instead of ranking – individual units into homogeneous classes with respect to certain characteristics.

On one side, retaining the multidimensional structure of the data is necessary to account for diversity, keeping the differential information on the several latent dimensions. On the other side, classifying units in homogeneous groups is less rigid and more flexible than ranking them through a crude single numeric value. It follows that when evaluating complex phenomena, both objectives have to be pursued. In other terms, it is necessary to seek information on the different dimensions which contribute to characterise a multidimensional phenomenon, and contextually classify the units in homogeneous classes as regards to these relevant latent dimensions.

Seeking these two aims at the same time provides a much richer information than ranking the single units on a single metrics. In fact, it allows us to have a picture of the way single units (Countries, students, regions within a Country, and so on) get together as regards the distinct latent dimensions which compose a complex phenomenon. Furthermore, such a key twofold goal is a good compromise between the need to retain multidimensionality and to classify units in a simple and accessible format for policy makers for communicating purposes. To gain such a twofold purpose a suitable statistical model is the latent class (LC) multidimensional IRT model (Bartolucci 2007), described in sections "The multidimensional LC IRT model" and "Dimensionality assessment", and applied in section "The multidimensional LC IRT in action: a case study".

Item Response Theory: Premises

Traditional Item Response Theory (IRT) models, which as known are born in the field of psychometrics, assume that the associations between the responses of an individual are fully accounted for by only one latent trait which, in the educational setting, is the students' ability. The characterisation of the latent person space in terms of a single unidimensional latent dimension implies that all variables of a measurement instrument (i.e., the items of a students' national test) are located on the same scale, contributing to measure a single latent trait (Bartolucci et al. 2015). However, measurement instruments, and among them students' assessment tests, are often composed by subsets of variables, or items, measuring different but potentially related constructs or content domains. In such later contexts, the traditional IRT assumption of only one underlying latent variable is inappropriate and restrictive for the data at issue. In fact, the unidimensional approach ignores the differential information on students' ability levels relative to several latent traits, which are confused in the same measurement (Camilli 1992; Embretson 1991; Luecht and Miller 1992).

To overcome these limitations, Bartolucci (2007) proposes a semi-parametric approach based on a class of multidimensional latent class IRT models. As it will be further described in the following section, such approach takes into account multidimensional latent traits (Reckase 2009) and more general item parameterisations than those of Rasch-type models (Rasch 1961), that is, the two-parameter logistic (2PL) model introduced by Birnbaum (1968).

Moreover, the model at issue represents abilities by a random vector with a discrete distribution common to all subjects. Representing the ability distribution through a discrete latent variable is more flexible than representing it by means of a continuous distribution, as it allows to classify individuals in homogeneous classes having very similar latent characteristics.

The Multidimensional LC IRT Model

The class of multidimensional LC IRT models developed by Bartolucci (2007), and applied in this paper, presents two main differences with respect to classic IRT models: *i*. the latent structure is multidimensional and *ii*. it is based on latent variables that have a discrete distribution, meaning that the population under study is made up by a finite number of classes, with subjects in the same class having the same ability level (Lazarsfeld and Henry 1968; Lindsay et al. 1991; Formann 1995); see Bacci et al. (2014) for a more general formulation for polytomously-scored items. In this paper we consider in particular the version of these models, based on the two-parameter (2PL) logistic parameterisation of the conditional response probabilities (Birnbaum 1968).

Let n denote the number of units in the sample and suppose that they answer to r dichotomous test items, which measure s different latent traits, or dimensions. Besides, let J_d, $d = 1,\ldots,s$, be the subset of $J = \{1,2,\ldots,r\}$containing the indices of the items measuring the latent trait of type d and let r_d denoting the cardinality of this subset, so that$r = \sum_{d=1}^{s} r_d$. The subsets J_d are disjoint as each item measures only a latent trait. On the other hand, we assume that the latent traits may be correlated.

The 2PL parameterisation implies that

$$\text{logit}\left[p\left(Y_{ij} = 1|V_i = v\right)\right] = \gamma_j\left(\sum_{d=1}^{s} \delta_{jd}\xi_{vd}^{(V)} - \beta_j\right), \ \ i = 1, \ldots, n, \ \ j = 1, \ldots, r, \tag{9.1}$$

where Y_{ij} is the response to item j provided by unit i ($Y_{ij} = 0,1$ for wrong or right response, respectively), β_j is the difficulty level of item j and γ_j is its discrimination level. Besides, V_i is a latent variable indicating the latent class of the subject i, v is one of the possible realisations of V_i, and δ_{jd} is a dummy variable equal to 1 if index j belongs to J_d (and then item j measures the dth latent trait) and to 0 otherwise. Finally, each random variable V_i has a discrete distribution with support $1,\ldots,k_V$ corresponding to the k_V latent classes in the population.

Associated to subjects in latent class v, there is a vector $\boldsymbol{\xi}_v^{(V)}$ with elements $\xi_{vd}^{(V)}$ corresponding to the ability level of subjects in latent class v with respect to dimension d. Note that, when $\gamma_j = 1$ for all j, then the above 2PL parameterisation reduces to a multidimensional Rasch parameterisation. At the same time, when the elements of each support vector $\boldsymbol{\xi}_v^{(V)}$ are obtained by the same linear transformation of the first element, the model is indeed unidimensional even when $s > 1$.

The assumption that the latent variables have a discrete distribution implies the following *manifest distribution* of the full response vector $\boldsymbol{Y}_{\mathrm{i}} = (Y_{i1}, \ldots, Y_{ir})^{\mathrm{T}}$:

$$p(\boldsymbol{y}_i) = p(\boldsymbol{Y}_i = y_i) = \sum_{v=1}^{k_V} p_v(\boldsymbol{y}_i)\pi_v^{(V)}, \tag{9.2}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ir})^{\mathrm{T}}$ is a realisation of $\boldsymbol{Y}_{\mathrm{i}}$, $\pi_v^{(V)} = p(V_i = v)$ denotes the weight, or *a priori* probability of the vth latent class, with $\sum_v \pi_v^{(V)} = 1$ and $\pi_v^{(V)} > 0$ for $v = 1, \ldots, k_V$. Moreover, the *local independence assumption* which characterises all IRT models, implies that.

$$p_v(\boldsymbol{y}_i) = p(\boldsymbol{Y}_i = \boldsymbol{y}_i | V_i = v) = \prod_{j=1}^{r} p\left(Y_{ij} = y_{ij} | V_i = v\right), \quad v = 1, \ldots, k_V.$$

Dimensionality Assessment

The specification of the multidimensional LC 2PL model, based on the assumptions illustrated above, univocally depends on: *i*. the number of latent classes (k_V), *ii*. the number of latent dimensions (s), and *iii*. the way items are associated to the different latent dimensions. As to the number of latent classes (k_V), it can be chosen through a statistical approach (i.e., an information criterion such as BIC or AIC), or according to subjective choices based on specific objectives or previous research. Once the number of latent classes k_V is chosen for the 2PL model expressed by Eq. (9.1), the next step is the assessment of the dimensionality of the test. For this purpose, we test the hypothesis that the r items of the questionnaire measure s–1 instead of s dimensions: the s–1 dimensions are specified by collapsing two dimensions of the s initial ones into one, and then grouping the corresponding items.

To cluster items into a reduced number of groups, we adopt a hierarchical algorithm which allows us to group items measuring the same ability in the same cluster. The algorithm builds a sequence of nested models. It starts with estimating the most general model, that is, a multidimensional LC 2PL IRT model with a different dimension for each item – corresponding to the classic LC model – and ends with the most restrictive model, that is, a model with only one common dimension to all items – corresponding to a unidimensional LC 2PL IRT model. At each step of the procedure, the algorithm estimates a multidimensional LC 2PL IRT model and reduces the dimensionality of the test of a dimension, by collapsing two items in the same group (or two groups of items in the same group). Specifically, at each step, the following Likelihood Ratio (LR) test is performed for every pair of possible aggregations of items (or groups of items):

$$\mathrm{LR} = 2\sum_{\boldsymbol{y}} n(\boldsymbol{y})\log\left[\frac{\widehat{p}(\boldsymbol{y})}{\widehat{p}_0(\boldsymbol{y})}\right], \tag{9.3}$$

where $\widehat{p}(\boldsymbol{y})$ and $\widehat{p}_0(\boldsymbol{y})$ are the estimated probability of configuration y under the model with s and s–1 dimensions, respectively.

The LR test is thus used to compare models which differ only in terms of their dimensional structure, all the rest keeping constant. This type of statistical test allows us to evaluate the similarity between a general model and a restricted model, i.e., a model which is obtained by the general one by imposing one constraint (i.e., so that the restricted model is nested in the general one). More precisely, the LR test evaluates, at a given significance level, the null hypothesis of equivalence between the two nested models at issue. If the null hypothesis is not rejected, the restricted model is preferred, in the interest of parsimony. If the null hypothesis is rejected, the general model is preferred.We remind that in our framework, the most general model is the one with a dimension for each item, whereas the most restricted model is used when all items belong to the same dimension.

The output of the above clustering algorithm may be displayed through a dendrogram that shows the deviance between the initial (s-dimensional) LC model and the model selected at each step of the clustering procedure. As known, the results of a cluster analysis based on a hierarchical procedure, and the consequent choice of the number of dimensions of a test, depend on the adopted rule to cut the dendrogram, which may be chosen according to several criteria. Bartolucci (2007) proposed a criterion based on the LR test statistic: the dendrogram is cut at the level corresponding to the first aggregation for which the test is significant. However, such an approach can be misleading for large samples, because it leads to overestimate the dimensionality of the latent structure. A more suitable rule which may be adopted for large samples, as ours, is based on a suitable information criterion, such as the Bayesian Information Criterion (Schwarz 1978), defined by:

$$\mathrm{BIC} = -2\widehat{l} + (\log n)m,$$

where $\widehat{l}$ is the maximum of the log-likelihood for a given model, whose number of parameters is equal to m, while n is the sample size. In particular, we select the model (then the number of dimensions) for which the difference between its BIC and that of the LC model becomes positive.

The Multidimensional LC IRT in Action: A Case Study

In this section an application to the Italian Multipurpose Surveys on Household is presented, in particular referred to the sample survey "Aspects of daily life" collected in 2010. From the whole set of items of the questionnaire at issue, only the 21 items specifically related to spare time and holidays are kept. As outlined in section "Introduction", the objective is to classify individuals (i.e., Italian citizens) in latent classes and to provide a clustering of the items in disjoint groups, each corresponding to a different dimension of the phenomenon under study.

By following the steps described in section "Dimensionality assessment", the number of latent classes (k_V) has to be chosen. We recall that this is the number of groups in which the population under study is divided: each group contains individuals having similar characteristics in terms of habits about spare time and holidays. In this application we do not use a statistical criterion to get k_V, and choose $k_V = 3$ as a suitable number of latent classes.

Then, in order to assess the dimensionality of the phenomenon considered here, we apply the hierarchical clustering algorithm described in section "Dimensionality assessment" to our sample of 1500 subjects. Such procedure aims at clustering items that contribute to measure the same dimension. The dendrogram associated to the clustering procedure is reported in Fig. 9.1.The clustering algorithm highlights three groups of items and each group contributes to measure a different dimension ($s = 3$), that we interpret as follows: *i*. developing one's growth, education, social commitment; *ii*. reinforcing one's family relationships and friendships; *iii*. engaging in sport activities.

The average probabilities of presenting each dimension $d = 1, 2, 3$, given the latent class $v = 1, 2, 3$, are reported in Table 9.1 and denoted with $\overline{\lambda}_{d|v}$. Such latest probabilities are obtained from a 2PL model (1) and considering the three clusters of items (dimensions) obtained above. Then, they are averaged over the items that contribute to measure each dimension, given each latent class.

By inspecting Table 9.1 we can provide a characterisation of each class of individuals ($v = 1, 2, 3$), based on people's involved dimensions of spare time habits ($d = 1, 2, 3$). We can observe that the probability of having each dimension increases if we move from the first latent class to the third one, for all three dimensions. Hence, overall, the third sub-population (i.e., the class of individuals in $v = 3$) includes individuals mostly dedicated to leisure, while the first class (i.e., the class of individuals in $v = 1$) includes people who are relatively least involved in any of the three dimensions of spare time. The class $v = 2$ is a class which collects individuals with an intermediate level of engagement in spare time. Furthermore, we can observe that the dimension which mostly contributes to characterise class $v = 3$ is "Dimension 2: reinforcing one's family relationship and friendship" ($\lambda^{-}_{2|3} = 0.664$). Differently, class $v = 2$, that is, the class of people who show an intermediate level of engagement in leisure activities, is made of people mostly involved in "Dimension 3: engaging insport activities" and "Dimension 1: developing one's growth, education, social commitment". Finally, the third class collects

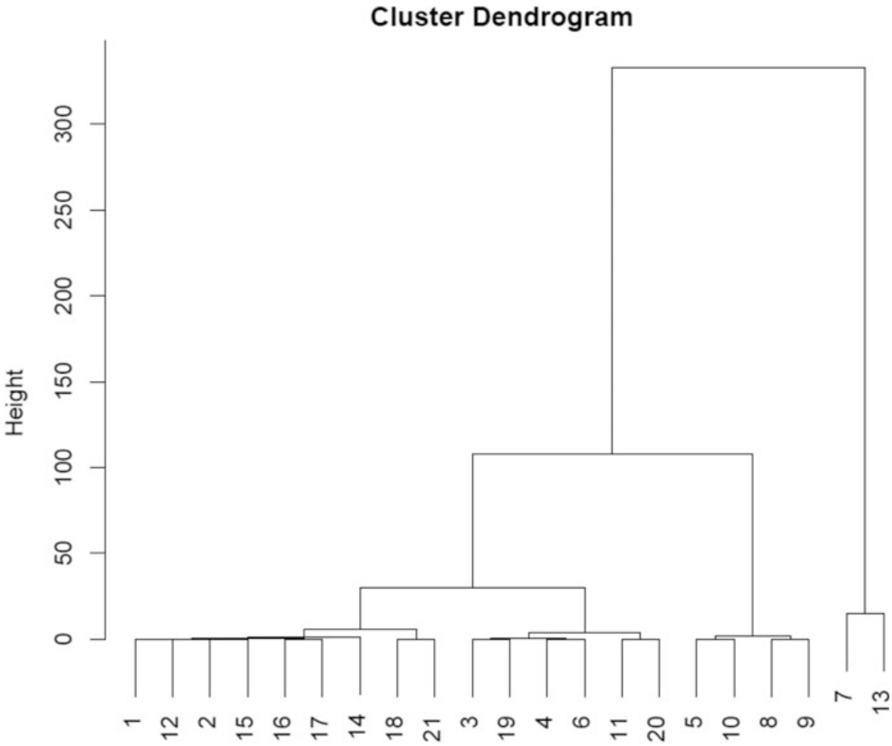


Fig. 9.1 Dendrogram associated to the hierarchical clustering procedure applied to data on habits about spare time and holidays from the Multipurpose Surveys on Household (2010): on the x-axis the item labels are reported, while the deviance between the initial LC model (with 21 dimensions) and the model selected at each step of the clustering procedure is reported on the y-axis

Table 9.1 Estimated averaged probability $\overline{\lambda}_{d|v}$ for individuals grouped in latent class $v = 1, 2, 3$ to present dimension $d = 1, 2, 3$

	Dimension d		
Latent class v	1	2	3
1	0.071	0.010	0.311
2	0.353	0.022	0.496
3	0.410	0.664	0.575

people who do not care much of spare time and mostly devote it to sport activities (Dimension 3).

To end up, the analyses show that the habit as regards to one's spare time is multidimensional. Such a habit goes from a basic level of engagement which involves sport activities, to a maximum level where a person's wish to reinforce family relationships and friendships comes also at issue. Therefore, singling outjust one, albeit composite, indicator for such a habit would be questionable because it would imply to remove as noise all the relevant dimensions conflicting with unidimensionality.

With this application, we also show that, when synthesising complex phenomena, it is possible to combine the call for condensed information with multidimensional instances. In fact, we provide a classification of our units of analysis in latent classes (instead of ranking them) characterised on the three highlighted dimensions of spare time. Clearly, the method can be extended to other applicative settings, when one aims at identifying (and possibly explaining) the differences among units of interest (such as Countries, instead of individuals) and, through this aim, at a classification of them based on other issues of interest – such as development, quality of higher education, learning competences, etc., instead of spare time habits.

References

Alkire, S., & Foster, J. (2007). Counting and multidimensional poverty measures (Vol. 7 of OPHI working paper series). Oxford: University of Oxford Poverty and Human Development Initiative.

Amsler, S. (2014). University ranking: A dialogue on turning towards alternatives. *Ethics in Science and Environmental Politics, 13*, 155–166.

Amsler, S., & Bolsmann, C. (2012). University ranking as social exclusion. *British Journal of Sociology of Education, 33*, 283–301.

Bacci, S., Bartolucci, F., & Gnaldi, M. (2014). A class of multidimensionallatent class IRT models for ordinal polytomous item responses. *Communication in Statistics Theory and Methods, 43*, 787–800.

Bartolini Bussi, M. G., Boni, M., Ferri, F., & Garuti, R. (1999). Early approach to theoretical thinking: Gears in primary school. *Educational Studies in Mathematics, 39*, 67–87.

Bartolucci, F. (2007). A class of multidimensional IRT models for testingunidimensionality and clustering items. *Psychometrika, 72*, 141–157.

Bartolucci, F., Bacci, S., & Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*. New York: Chapman &Hall/CRCHall.

Birnbaum, A. (1968). Some latent trait models and their use in inferring anexaminee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Reading: Addison-Wesley.

Bonifay, W. E., Reise, S. P., Scheines, R., & Mei, R. R. (2015). Whenare multidimensional data unidimensional enough for structural equationmodeling? An evaluation of the detect multidimensionality index. *StructuralEquation Modeling: A Multidisciplinary Journal, 22*, 504–516.

Camilli, G. (1992). A conceptual analysis of differential item functioning interms of a multidimensional item response model. *Applied Psychological Measurement, 16*, 129–147.

Douek, N. (2006). Some remarks about argumentation and proof. In P. Boero (Ed.), *Theorems in school: From history, epistemology and cognition to classroom practice*. Rotterdam: Sense Publishers.

Embretson, S. E. (1991). A multidimensional latent trait model for measuringlearning and change. *Psychometrika, 56*, 495–515.

Fattore, M., Maggino, F., & Colombo, E. (2012). From composite indicators to partial orders: Evaluating socio-economic phenomena through ordinaldata. In F. Maggino & G. Nuvolati (Eds.), *Quality of life in Italy: Research and reflections, number 48 in social indicators research series* (pp. 41–68). Dordrecht/Heidelberg/New York/London: Springer.

Formann, A. K. (1995). Linear logistic latent class analysis and the Raschmodel. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations,recent developments, and applications, pages* (pp. 239–255). NewYork: Springer.

Gnaldi, M., & Ranalli, M. G. (2016). Measuring university performance bymeans of composite indicators. A robustness analysis of the composite measure used for the benchmark of Italian universities. *Social Indicators Research, 129*, 659. doi:10.1007/s11205-015-1116-1.

Golay, P., & Lecerf, T. (2011). On higher order structure and confirmatoryfactor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological Assessment, 23*, 143–152.

Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.

Holzinger, K., & Swineford, S. (1937). The bi-factor method. *Psychometrika, 47*, 41–54.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principalcomponents. *Journal of Educational Psychology, 24*(7), 417–441.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lindsay, B., Clogg, C., & Greco, J. (1991). Semiparametric estimation inthe Rasch model and related exponential response models, including a simplelatent class model for item analysis. *Journal of the American StatisticalAssociation, 86*, 96–107.

Luecht, R. M., & Miller, R. (1992). Unidimensional calibrations and interpretations of composite traits for multidimensional tests. *Applied Psychological Measurement, 16*, 279–293.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale: Erlbaum.

Munda, G., Nardo, M., Saisana, M., & Srebotnjak, T. (2009). Measuringuncertainties in composite indicators of sustainability. *International Journal of Environmental Technology and Management, 11*, 7–26.

OECD. (2008). *Handbook on constructing composite indicators: Methodologyand user guide*. Paris: OECD Publishing.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *In proceedings of the IV Berkeley symposium on mathematical statistics and probability* (pp. 321–333). Berkeley: University of California Press.

Reckase, M. (2009). *Multidimensional item response theory*. New York: Springer.

Saisana, M., & D'Hombres, B. (2008). Higher education rankings: Robustness issues and critical assessment. EUR Report 23487 EN, EuropeanCommission. JRC-IPSC EC Publishing.

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of compositeindicators. *Journal of the Royal Statistical Society. Series A, 168*, 307–323.

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research, 81*, 65–77.

Schoenfeld, A. (1992). Learning to think mathematically: Problem solving,metacognition, and sense making in mathematics. In D. Grows (Ed.), *Handbook for research on mathematics teaching and learning*. New York: Macmillan.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Wirth, R., & Edwards, M. (2007). Item factor analysis: Current approachesand future directions. *Psychological Methods, 12*(1), 58–79.

Zhou, P., & Ang, B. W. (2009). Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure. *Social Indicators Research, 94*, 83–96.

Part IV
Particular Experiences

Chapter 10
Synthesis of Indicators Through Weighting: The Experiences of Quality of Life Measures

Chang-ming Hsieh

Introduction

Development and construction of synthetic indicators for socioeconomics inevitably face the very same concern confronting many quality of life (QOL) measures for individuals: how to account for potential societal, cultural and individual differences in values associated with different sectors, facets or domains represented by different indicators. In QOL research, a common approach is to use weighting to reflect the potential individual differences in values associated with different facets or domains of life. Also known as importance weighting in QOL literature, researchers capture individual differences in values, using relative importance of different facets or domains of life with the assignment of weights. Although using importance weighting to capture relative importance of various life facets or life domains appears straightforward, there are a number of conceptual and methodological issues that remain unsettled. In this chapter, major issues related importance weighting in QOL measures are discussed to provide implications for developing and constructing synthetic indicators.

Similar to synthetic indicators, measures of QOL are often developed and constructed to represent an abstract and multidimensional construct. For QOL, its definition is often elusive and encompasses different meanings in different disciplines (e.g., Cummins et al. 1994; Michalos 2004). Happiness, life satisfaction, and other well-being measures have all been considered indicators of QOL (e.g., Campbell et al. 1976; Cummins et al. 1994; Michalos 2004). QOL measures constructed based on the so-called "bottom-up" approach (e.g., Hsieh 2004) closely resemble the construction of synthetic indicators in that domain-specific indicators are viewed as determining or causing the overall QOL construct (Bollen and

C.-m. Hsieh (✉)
University of Illinois at Chicago, Chicago, IL 60607, USA
e-mail: chsieh@uic.edu

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_10

Lennox 1991; Chin and Newsted 1999; Cohen et al. 1990). Debates and issues regarding the need as well as the justification for weighting in QOL measures following the bottom-up approach, therefore, offer invaluable insights to the development of synthetic indicators.

Issues related to importance weighting in QOL measures can be roughly divided into conceptual and methodological. Conceptual issues involve questions and theoretical justifications for importance weighting. Methodological issues involve the adequacy of how weighting is done. Besides conceptual and methodological issues, there are issues related to how importance weighting is assessed. More specifically, these issues concern the procedures by which the need for and the benefit of importance weighting are evaluated. The following is a summary of the conceptual, methodological issues and issues regarding assessing importance weighting.

Conceptual Issues: Justified or Unjustified

Importance Weighting: Justified

Importance weighting is commonly used to serve the purpose of capturing the potential uneven contributions of various life domains to global QOL. The concept of uneven contributions of various life domains to global QOL was recognized decades ago. Although named differently, domain importance (Campbell et al. 1976), value priority (Inglehart 1978), and psychological centrality (Ryff and Essex 1992) all represent the similar concept of valence in QOL measures (Hsieh 2012). Therefore, arguing for importance weighting is not without conceptual foundation. In addition, some QOL measures are conceptualized and defined specifically to account for importance of various life domains. For example, the Ferrans and Powers' Quality of Life Index (QLI) defined QOL as "a person's sense of well-being that stems from satisfaction or dissatisfaction with the areas of life that are important to him/her" (Ferrans 1990, p. 15). Based on the definition of QOL of the Ferrans and Powers' QLI (Ferrans and Powers 1985), weighting domain satisfaction using domain importance can be justified.

A main challenge with the conceptual justification for importance weighting, like psychological centrality, is the fact that it does not provide any specific guidance regarding *how* weighting is to be constructed. In other words, the conceptual support for importance weighting merely suggests that more important life domains may contribute more to global QOL, but it does not offer any insight regarding the extent to which how much more a more important life domain should contribute to global QOL than a less important life domain. Without any guidance or insight, it is not clear how much contribution a life domain that is perceived as "not important" should have in one's global QOL. Similarly, if domain importance is measured using a Likert-type rating scale, it is unclear how much more contribution should be given for each one point increase in a five-point (or seven-point) importance rating scale.

Importance Weighting: Unjustified

There have been two major claims arguing against importance weighting in QOL measures (Hsieh 2013). One claim used to oppose importance weighting is that not all life domains can be included and domains chosen in any QOL measure are clearly important. Given domains are selected because they are important, it is unnecessary to consider domain importance again. Importance weighting, therefore, is unnecessary as well (e.g., Trauer and Mackinnon 2001). The other claim that opponents of importance weighting in QOL measures (see Wu 2008a, b; Wu and Yao 2006a, b, 2007) often cite is Locke's "range-of-affect" hypothesis (Locke 1969, 1976, 1984). More specifically, the claim against importance weighting is an extension of the range-of-affect hypothesis: the concept of implicit weighting (e.g., McFarlin et al. 1995; McFarlin and Rice 1992). The range-of-affect hypothesis postulates that satisfaction with specific domains is dependent on the importance that individuals perceive for the domains. Given that domain satisfaction has already considered domain importance (implicit weighting), importance weighting would be redundant.

The claim that life domains included in any QOL measure are clearly important so it is unnecessary to consider domain importance again is somewhat misleading (Hsieh 2004). Assuming the claim that domains included in QOL measures are selected based on their importance is true, there is no evidence to suggest that all the selected domains would be *equally* important. That is, the fact that QOL measures select life domains based on the importance of the domains does not preclude the possibility that some of the selected life domains are more important than others. Unless the claim that "QOL measures select life domains based on the importance of the domains" actually means that "all the life domains selected are equally important," arguing that it is unnecessary for importance weighting is a leap from the claim (Hsieh 2013).

Using Locke's range-of-affect to argue against importance weighting in QOL measures is also controversial (Hsieh 2012). In fact, the main argument of Locke's range-of-affect hypothesis (Locke 1969, 1976) spoke to the relationships between (the so-called have-want) discrepancy, importance and satisfaction within individual domains (or facets). Regarding the relationships between global satisfaction and domain satisfaction, what was postulated, based on the range-of-affect hypothesis, was that (1) it would be redundant to weight domain satisfaction with domain importance (the concept of implicit weighting) and (2) global (job) satisfaction should just be the sum of domain satisfactions (Locke 1969, 1976). Although the rationale, based on the range-of-affect hypothesis, for the argument that importance has already been included in satisfaction evaluations was pretty straightforward, the rationale for the additive relationship between domain satisfactions and global satisfaction was less than clear. Specifically, it does not seem that the hypothesis – global satisfaction is the sum of domain satisfactions – can be derived from the concept of implicit weighting without assuming that the relationship between domain satisfactions is a linear and consistent one across all domains. The notion that importance has been included in satisfaction evaluations *within* a domain does not speak to the relationship of satisfaction *between* or across domains. By arguing that global satisfaction is a sum of domain satisfactions, one assumes that, among others, a consistent relationship that has the nature of perfect substitution between satisfaction across all domains (see Rojas 2006 for a detailed discussion). However, this linear relationship of satisfactions across domains appeared to be more of an assumption than a hypothesis since the literature on range-of-affect hypothesis has not offered any justification for it. Empirical evidence from job satisfaction literature, however, seemed to suggest against this simple linear relationship assumption. For example, Rice et al. (1991) showed that there was a significant curvilinear (quadratic form) relationship between global job satisfaction and domain specific satisfaction in two of the job domains studied, while the same curvilinear relationship between global job satisfaction and domain specific satisfaction was not found in other domains studied. Since the relationship between global (job) satisfaction and domain satisfaction did not appear consistent across domains, the assumption that global satisfaction was the simple sum of satisfaction across domains became arbitrary, if not questionable.

Using the concept of implicit weighting, derived from the range-of-affect hypothesis, to argue against domain importance weighting in the life satisfaction context may appear reasonable. After all, if domain importance is implicitly reflected in domain satisfaction, then it would seem redundant to weight domain satisfaction with importance repeatedly. Unfortunately, like in the job satisfaction literature, the concept of implicit weighting, in the life satisfaction context, centers on the relationship between discrepancy, importance and satisfaction within individual life domains, not between or across life domains. When it comes to the relationship between global life satisfaction and the composite of life domain satisfactions, the proposed simple linear additive relationship of the range-of-affect hypothesis (similar to the job satisfaction literature) relies on an assumption without justification. On the other hand, the relationship between global life satisfaction and the composite of domain satisfactions is the focus point of investigation of domain importance weighting. The recent study by Rojas (2006) detailed the limitations and implications of this assumed linear relationship (and three other alternative specifications) between global life satisfaction and domain satisfaction. As Rojas (2006) suggested, assuming a linear relationship between global life satisfaction and domain satisfaction would result in a loss of many features of the relationship and would clearly restrict our understanding of life satisfaction.

Methodological Issues: Weighting or Multiplicative Scores

Importance weighting is, in many instances, accomplished by multiplying satisfaction rating by importance rating scores (e.g., Hsieh 2003). Specifically, some QOL measures ask respondents to provide both importance and satisfaction ratings for various life domains and then multiply satisfaction rating by importance rating for each life domain before summing or averaging across domains to obtain global QOL scores (e.g., Trauer and Mackinnon 2001). This method of scoring, known as multiplicative scores, has been criticized for a lack of conceptually defined meaning, inadequate psychometric property and no empirical support (e.g., Hsieh 2003, 2004; Russell and Hubley 2005; Russell et al. 2006; Trauer and Mackinnon 2001). Since the product of satisfaction rating and importance rating, a multiplicative score, cannot be clearly defined, the extent to which importance weighting can be accomplished by multiplicative scores is questionable.

It is important to note that there is much more to importance weighting than multiplicative scores. Maggino and Zumbo (2012) suggested that weighting could be accomplished by statistical methods (such as principal component analysis, data envelopment analysis), multiattribute methods (such as analytic hierarchy process and conjoint analysis), and subjective methods. According to Wu and Yao (2014), subjective methods remained most popular in QOL measures. Multiplicative scores are only one type of subjective methods used for weighting. Other subjective methods of weighting are clearly possible. For example, instead of multiplying, Hsieh (2003, 2004) proposed various possible weighting methods that accomplish importance weighting without conceptual ambiguity. As Campbell et al. (1976) pointed out decades ago, importance weighting can be implemented in a variety of ways. Land (2014) identified two major approaches to weighting schemes: (1) the weighted average model of subjective judgments and (2) the weighted product model. Unfortunately, importance weighting in QOL measures has seldom considered options beyond multiplicative scores (e.g., Hsieh 2003, 2004, 2012, 2013). In other words, importance weighting cannot and should not be considered synonymous to multiplicative scores, given all the other possible weighting methods. Arguing against importance weighting completely based on only issues/problems related to multiplicative scores is clearly over-generalizing.

Issues Regarding Assessing Importance Weighting

Given that importance weighting in QOL measures involves both conceptual and empirical aspects, assessing the performance of importance weighting becomes a critical matter in supporting arguments for and against importance weighting. According to Hsieh (2012, 2013), most empirical research investigating the performance of importance weighting in QOL measures follow a line of logic in research methodology as follows. First, choose popular QOL measures as criteria and then compare the association between the criteria and unweighted scores and the association between criteria and weighted scores. If results from the comparison indicate that weighted scores do not show stronger association with the criterion than the unweighted scores, then the conclusion is to argue against the use of importance weighting in QOL measures, or even to call for the complete abandonment of importance weighting. The line of logic, although straightforward, is based on two critical assumptions: (1) the criteria chosen must be appropriate; and (2) the weighting method(s) used must be correct.

Unfortunately, neither of the above two assumptions has been closely examined. In fact, the majority of research assessing the performance of importance weighting in QOL measures simply accepts the assumptions as given and has three major characteristics (Hsieh 2012, 2013). First, only a limited number and choice of QOL measures is used as criterion variables. Second, methods used to construct importance weighting are based on multiplicative scores. Third, only a limited number of domains of life are included in the measures being assessed. These characteristics call into question the appropriateness of the two underlying assumptions upon which the research methodology of assessing importance weighting in QOL measures is based. More specifically, there is no clear evidence to support that criteria chosen is appropriate nor is there evidence to support the correctness of weighting methods applied. In addition, these characteristics bring forth the following three unresolved issues:

Issue One: The Choice of Criteria

A few measures that are developed to assess overall QOL (i.e., global QOL measures), such as the Satisfaction with Life Scale (Diener et al. 1985), and single-item overall QOL questions, have been used in previous studies as criterion variables to assess the performance of importance weighting, There are certainly many more global QOL measures that exist (e.g., Evans 1991; Hagerty et al. 2001). Although it is reasonable to expect that scores from all global QOL measures reflect global QOL, there does not appear to be any specific standard for the selection of global QOL measures to serve as criteria in evaluating the performance of importance weighting. All measures of global QOL should be expected to overlap to some extent. However, there is no reason to presume that all QOL measures must capture identical constructs (Russell and Hubley 2005; Russell et al. 2006). Given that there is no single agreed-upon "best" global QOL measure, the extent to which any chosen QOL measures can be regarded as the "true" criterion is unclear.

As demonstrated in previous research (e.g., Hsieh 2003, 2004; Russel et al. 2006), performance of importance weighting is heavily dependent on the criterion variables chosen. The choice of global QOL measures as criterion variables is, therefore, consequential. When empirical evidence is used to argue for importance weighting, it is important to acknowledge that similar performance results of importance weighting may or may not necessarily hold if other criterion variables are used. However, it is even *more* important to acknowledge that similar performance results of importance weighting may or may not necessarily hold if other criterion variables are used when empirical evidence is used to argue against importance weighting. Considering the number of existing global QOL measures, it is very unlikely, if not impossible, for any single study to select all of them as criterion variables. Under the circumstances, caution is needed in interpreting study results arguing against importance weighting, in order to avoid overgeneralization.

Issue Two: Weighting Methods

This issue of weighting methods is closely related to the methodological issue regarding importance weighting discussed earlier. In the literature on importance weighting in QOL measures, most weighting methods applied so far have been based on the assumption of a linear function of domain importance scores (Hsieh 2013). More often than not, multiplicative scores (multiplying satisfaction by importance scores) are used to represent the weighting function (e.g., Russell et al. 2006; Trauer and Mackinnon 2001). As Hsieh (2003, 2004) showed, the performance of importance weighting was dependent upon, among others, (1) how importance was measured and (2) how importance was weighted. The popular practice of using rating scales to measure domain importance and using multiplicative scores as weighting is only one of the many possible ways to measure importance weighting. As Kaplan et al. (1993) indicated, the use of rating scales was by no means the only method for assessing relative importance. Other methods, including utility assessment, such as the standard gamble, time trade-off and person trade-off methods, as well as economic measurements of choice should not be ignored (Kaplan et al. 1993). In addition to different ways of measuring domain importance, there can be many other options other than simply multiplying satisfaction scores by importance scores to the construction of weighting. Campbell et al. (1976), for example, suggested various types of weighting function of domain importance, such as “hierarchy of needs,” “threshold,” and “ceiling.” It should be noted that the actual weighting function of domain importance is still unclear and is not necessarily limited to the functions discussed by Campbell et al. (1976). Guardiola and Picazo-Tadeo (2014), for example, used data envelopment analysis (DEA) and multi-criteria-decision-making (MCDM) techniques to construct weighted-domain composite indices of life satisfaction. Based on the so-called “benefit-of-the-doubt” principle (Cherchye et al. 2007), these weighting techniques went beyond the typical multiplicative scores and allowed weights to differ across life domains and individuals (Guardiola and Picazo-Tadeo 2014).

Since importance weighting is by no means limited to only multiplicative scores, it is critical not to over-generalize the results based on only multiplicative scores in determining the adequacy of importance weighting. In particular, one must be cautious not to argue against importance weighting completely using only evidence based only on multiplicative scores. It is also important to recognize that even if there is no evidence to support multiplicative scores or other linear function of domain importance based on rating scale scores, it is unclear what the evidence would show with other types of weighting function of domain importance and different types of measurement of domain importance. In sum, there is no reason to equate the concept domain importance weighting with one particular type of measurement of domain importance or weighting method.

Issue Three: Limited Number of Domains Included

According to Cummins (1996), there are at least 173 different domain names that have been used in the QOL literature, and it is unlikely that 173 domain names cover all possible domains comprehensively. For any single QOL measure to cover all possible domains of life is extremely unlikely, if not impossible. Measures of global QOL typically follow either the "bottom-up" or "top-down" approach (e.g., Diener 1984; Feist et al. 1995; Headey et al. 1991; Lance et al. 1989). For example, the Satisfaction with Life Scale by Diener et al. (1985) used the "top-down" approach, while the Quality of Life Index by Campbell et al. (1976) followed the "bottom-up" approach. QOL measures that follow the bottom-up approach are consistent with the "causal-indicator" model, and QOL measures that follow the top-down approach are consistent with the "effect-indicator" model (Chin and Newsted 1999; Cohen et al. 1990; Hsieh 2004). In a causal-indicator model, indicators are viewed as determining or causing the construct. In an effect-indicator model, indicators are viewed as determined by the construct (Bollen and Lennox 1991; Cohen et al. 1990; Chin and Newsted 1999). In the context of QOL research, Hsieh (2004) provided a conceptual framework for incorporating importance weighting in a causal-indicator model.

Consequences of not covering all domains comprehensively vary, depending on the measurement model. For QOL measures that follow the reflective-indicator model in which measure items are considered interchangeable, whether or not domains are covered comprehensively is of no major concern. On the other hand, for QOL measures that follow the causal-indicator model, items in the measure may not be interchangeable. The extent of domains covered in the measure can be a cause for major concern, given that not including items can alter the measurement construct (e.g., Bollen and Lennox 1991; Diamantopoulos 2006; Diamantopoulos et al. 2008; Hardin et al. 2011; Jarvis et al. 2003; MacKenzie et al. 2005).

The performance of importance weighting is typically assessed with QOL measures following the bottom-up approach, or causal-indicator model (e.g., Campbell et al. 1976; Hsieh 2003; Wu 2008b; Wu et al. 2009). Given that it is unlikely for any single measure using the bottom-up approach to cover all possible domains of life comprehensively, the consequences of not including all possible domains comprehensively in causal-indicator QOL measures should be investigated. There has not been any established standard regarding the extent of comprehensiveness or exhaustiveness of the domains for causal-indicator global QOL measures. As a result, the number of domains covered in different QOL measures varies, depending on the focus of the measure. Causal-indicator global QOL measures that cover different subsets of domains are unlikely to measure identical QOL constructs.

Consequently, finding any global QOL measures as criterion variables to assess the performance of importance weighting in causal-indicator QOL measures becomes a complex task. As discussed earlier, it is unrealistic to expect a causal-indicator global QOL measure to cover all possible domains comprehensively. It is, therefore, hard to justify why the causal-indicator QOL measures being assessed (in relation to importance weighting) and the criterion variables must measure the same concept of QOL. It does not seem reasonable to expect a causal-indicator global QOL measure covering five domains and one covering 20 domains to measure the same concept of QOL. In assessing the performance of importance weighting, it is important to consider the fact that causal-indicator global QOL measures are unlikely to cover all possible domains comprehensively. Specifically, the selection of criterion variables should take into account the scope of domains covered in the causal-indicator QOL measures. In the case when evidence does not support importance weighting, it is important to carefully examine whether the lack of support is a result of inappropriate criterion variables before coming to a concluding argument against importance weighting.

Recently, Hsieh and Kenagy (2014) discussed a major characteristic associated with formative-indicator global QOL measures, which is that formative-indicator global QOL measures normally do not cover all possible life domains comprehensively. This characteristic is critical to the investigation of domain importance weighting for at least two reasons. First, it is necessary to recognize the potential variations in the concept of QOL that different formative-indicator global QOL measures intend to capture. Second, the true nature of the relationship between global QOL and domain-specific satisfaction may not be so easily discerned with formative-indicator global QOL measures, using the popular analytical approaches for assessing domain importance weighting. Therefore, results of prior studies on domain importance weighting should be interpreted with caution. More specifically, since the extent to which the criterion variables used in previous studies accurately captured the concept that the formative-indicator QOL measures intended to measure was unclear, whether or not the commonly used criterion variables could or should be used to evaluate the performance of domain importance weighting with formative-indicator global QOL measures, therefore, became questionable.

Equally, if not more, alarming is the fact that both of the two most popular methods of evaluating the performance of domain importance weighting in QOL measures, correlation and moderated regression analysis, could produce misleading results with formative-indicator global QOL measures. As the findings of Hsieh and Kenagy's (2014) study showed, even if an individual's global QOL score could be obtained using a formative-indicator QOL measure that included a full range of life domains, the correlation between this "true" score and that of the score with a subset of domains might not be a good indicator for assessing domain importance weighting. Also evident from the findings of Hsieh and Kenagy's (2014) study was that moderated regression analysis could lead to the rejection of domain importance weighting when analyzing scores from formative-indicator measures with subsets of domains even if the full-domain scores reflected domain importance weighting. Findings from Hsieh and Kenagy's (2014) study suggest the following. First, our understanding of the topic of domain importance weighting remains quite limited. In the area of domain importance weighting at the individual level, there are issues with the way in which domain importance weighting has been assessed. In addition to issues that have been discussed previously, such as the selection of criterion variables and weighting methods (Hsieh 2012), identifying effective analytical approaches to assess the performance of domain importance weighting should be added as a major hurdle that needs to be overcome. As shown in Hsieh and Kenagy's (2014) study findings, neither of the two most popular approaches of assessing the performance of domain importance was robust enough to discern potential weighting functions of domain importance. It is important for future studies to re-examine the logic or rational underlying these approaches and develop effective approaches to clarify the relationship between global QOL and domain-specific satisfaction. Second, in light of the poor performance of the two most popular approaches of assessing domain importance weighting shown in Hsieh and Kenagy's (2014) study, findings from previous studies that used similar approaches (e.g., Hsieh 2003, 2004; Russell et al. 2006; Wu 2008a, b; Wu and Yao 2006a, b, 2007) may need to be re-interpreted. More specifically, the major characteristic of formative-indicator global QOL measures, which is not covering all possible life domains comprehensively, must be considered in interpreting the findings. That is, existing evidence is far from conclusive in determining whether domain importance weighting must be incorporated in linking domain satisfaction to global QOL. Third, results of Hsieh and Kenagy's (2014) study showed that a simple linear weighting function using domain importance scores produced weighted full-domain scores that were highly correlated with the unweighted scores. These results suggest that if the simple linear weighting function is correct, the unweighted (simple sum of satisfaction) scores are not an unreasonable choice to approximate the weighted scores (see Hagerty and Land 2007 for more). In fact, as demonstrated by Hagerty and Land (2007), in the absence of good information on domain importance, an equal-weighting method could be used and it was a so-called minimax estimator, as it minimized extreme disagreements on a composite QOL index. Given what we know, not having evidence to support importance weighting should only be interpreted as not having evidence to support importance weighting, which is far from having definitive evidence against importance weighting.

The Often Overlooked Issue of Statistical Power in Assessing the Performance of Importance Weighting

In addition to the three issues discussed above, assessing the performance of importance weighting has also been greatly affected by the analytical methods. The common analytical methods used to assess the performance of importance weighting have been correlation- or regression-based (such as correlation, moderated regression, and partial least squares regression), with moderated regression analysis being most popular (e.g., Hsieh 2012; Mastekaasa 1984; Russell et al. 2006; Wu and Yao 2006a, b). More specifically, as Hsieh (2015) recently pointed out, even in a hypothetical situation where all the issues discussed above are resolved, the use of moderated regression analysis can still be problematic if consideration of statistical power is not given.

For moderated regression analysis, the popular analytical method used to assess the performance of domain importance, statistical power is dependent upon, among other factors, sample size (Aberson 2010). A close look at the empirical evidence used for or against importance weighting would suggest that most studies had sample sizes that were small or moderate at best. For example, analyzing data from 130 undergraduate students, Wu and Yao (2006a) found that domain importance did not moderate the relationship between domain satisfaction and overall life satisfaction and argued against importance weighting. Russell et al. (2006) investigated the topic of importance weighting based on a sample of 241 subjects. More recently, Phillip et al. (Philip et al. 2009) argued against importance weighting based on data from 194 cancer patients.

Sample size affects statistical power. With a limited sample size, statistical power is likely to be limited if the effect size (Cohen 1988) to be detected is not large. Hsieh (2015) showed that in order to accurately assess the performance of importance weighting in QOL measures, the sample size required for reasonable statistical power might be larger than many of the previous studies could offer. The actual sample size required for any study to evaluate the performance of importance weighting will depend on the criterion variable(s) selected and the number of domains included. However, if the criterion variable(s) selected and the "true" scores of importance weighting function have only a moderate level of correlation (for example, $r = .60$ or less), a sample size of 200 or less, which is easily found in many previous studies, is unlikely to reach reasonable statistical power to assess the performance of importance weighting. Without adequate statistical power, chances of a type II error (false negative decision) occurring are likely to increase. Importance weighting studies that do not have adequate statistical power are prone to fail to reject the null hypothesis of no significant weighting effect when the null hypothesis is, in fact, false.

According to Shultz and Whitney (2004), the majority of validity coefficients that have been used to assess concurrent validity in the literature do not reach 0.5. In other words, it would be unrealistic to expect high associations/correlations between different measures of QOL. Under the circumstances, unless there is clear evidence to suggest otherwise, it would be prudent not to assume that any criterion variable selected to assess importance weighting will represent the "true" global QOL with a high degree of overlap. Since the association or correlation between any criterion variable and the "true" global QOL is unknown, selection of criterion variables is quite subjective. Treating the criterion variables as absolute standards for assessing importance weighting without considering whether or not as well as the extent to which these criterion variables really represent the "true" global QOL is by no means ideal. For the popular analytical method of moderated regression analysis to be effective in assessing importance weighting, statistical power must be considered. It would be helpful for future studies to include statistical power information, especially when analysis results show no significant weighting effect, so the potential false positive rate can be understood.

Weighting in the Construction of Synthetic Indicators: Implications from QOL Measures

The conceptual and empirical issues regarding weighting in QOL measures above discussed offer implications as well as lessons for the topic of weighting in the construction of synthetic indicators. Similar to QOL measures, synthetic indicators are constructed to represent concepts that are abstract and multidimensional, such as human development, or economic and social development. The conceptual and empirical debates of weighting in QOL measures offer important insight for the role of weighting in synthetic indicators. Specifically, when it comes to the topic of weighting in development and construction of synthetic indicators, at least three major issues must be carefully considered. These issues are: conceptual justification, weighting methods, and performance of weighting.

Justifications for Weighting in Synthetic Indicators

Before any consideration can be given regarding how to use weighting in synthetic indicators, reasonable justification should be offered regarding the rationale for weighting. Like psychological constructs/theories that have been applied as conceptual justifications for the need (or no need) for weighting in QOL measures, it would be important to provide conceptual/theoretical justification if weighting is to be incorporated in constructing synthetic indicators. In particular, a clear definition of what the synthetic indicators are used to represent must first be stated. Based on the definition, justification(s)/rationale for weighting (or not weighting) should be provided. Justifications can be based on specific disciplinary theories, concepts and/or logic. The literature on weighting in QOL measures points to the importance and necessity of not only providing conceptual/theoretical justifications but also acknowledging assumptions, potential limitations and consequences associated with the justifications (and/or rationale) for weighting in synthetic indicators. Adequacy and utility of synthetic indicators cannot be fully assessed if conceptual/theoretical justifications are not provided. Similarly, Adequacy and utility of synthetic indicators cannot be fully assessed without assumptions, potential limitations and consequences associated the justifications. For example, the Human Development Index (HDI) weights life expectancy, education and income dimensions (indicators) equally (Sayed et al. 2015), which assumes each of these dimension contributes to a nation's human development equally and human development can be correctly captured by just these three equally weighted dimensions. Clearly, adequacy and utility of HDI cannot be fully understood without considering these assumptions.

It must be noted that many may not consider the practice of a simple additive approach a type of importance weighting. In fact, a simple additive approach means to apply equal weights to all indicators included, and therefore, should be considered as a case of subjective weighting. This simple additive approach assumes, among others, that a constant marginal rate of substitution (it is always the same and independent of specific indicator levels), and perfect substitution between different indicators (Rojas 2006). Conceptual and practical implications of these assumptions associated with equal weights should be fully explored if the equal weight approach is to be used for developing synthetic indicators.

Weighting Methods in Synthetic Indicators

Given that weighting can be accomplished in a variety of ways, the choice of a specific weighting method for synthetic indicators should be reasonably justified (e.g., Saisana 2014; Saltelli 2007). In other words, there should be clear links between the "how" and the "why" regarding weighting. Equally, if not more, imperative is the justifications of the assumptions (and properties) of the weighting methods in relation to the purpose of the synthetic indicators. Assumptions and properties of various weighting methods should be considered carefully.

When incorporating a subjective weighting method (expert opinion, for example), it is important to consider the weights assigned in relation to substitution/compensation among indicators. If indicator A is assigned a weight coefficient of 0.2 and indicator B is assigned a weight coefficient of 0.1, it is necessary to recognize that indicator A has not only more but twice more weight than indicator B. Justifications for weights need to go beyond the general principle of assigning more weights for more "important" indicators to cover the extent/magnitude of differences in weights; otherwise, it would be necessary to acknowledge the subjective/arbitrary nature of the assigned weights. Limitations and implications associated with the subjective weight assignment should be investigated and addressed.

When using a statistical weighting method (such as principal component analysis), correlation among indicators should be carefully considered. Should indicators with low correlation with others be included, and why? Justifications for the appropriateness of the statistical method in meeting the purpose of synthetic indicators should be provided; otherwise, it would be necessary to acknowledge the limitations of the statistical method chosen. Implications stemming from the limitations should be assessed.

When using multiattribute method of weighting (such as analytic hierarchy process), it is important to recognize the subjective aspect of the method. Specifically, it is important to acknowledge the possible variations of results due to different decision makers (judges, or respondents) involved. Justifications for the selection of subjects/people should be offered. Potential bias in relation to cultural/societal values reflected by subjects involved should also be explored, especially when synthetic indicators are developed for cross-nation comparisons.

Different weighting methods certainly involve different considerations. It is likely that new and blended methods of weighting can be used for synthetic indicators. The main attention should be given to establish the conceptual/theoretical foundation between the method selected and the purpose of the synthetic indicators. For example, Sayed and colleagues (Sayed et al. 2015) recently developed a weighting method that used Meta-Goal Programming (MGP) to calculates weights for composite indicators. By building on the "Benefit-of-the-Doubt" (BoD) approach, the MGP-BoD method enhances discriminating power (reducing ties) and possesses the property that weights adding up to one. Although the features of enhanced discriminating power and weights adding up to one may be attractive, a method like MGP-BoD has properties that can be difficult to interpret given it produces results that are completely dependent on the data. In other words, if the same set of indicators are to be used to rank nations (on human development, for example) over time, the MGP-BoD method will produce different weight for the same indicator from 1 year to another. Similarly, the MGP-BoD method will produce results with very different ranking patterns for full data (the whole world, for example) from partial data (just Asia, for example). Implications of these properties must be carefully examined before practical utility can be established. Assuming the purpose is to develop a weighting method to compare human development across nations, then how should the rankings of nations be interpreted from year to year, given the MGP-BoD method produce different weights for same indicators from year to year? Similarly, how should the rankings of nations be interpreted if, given exactly the same indicators (national data), nation A ranks higher than nation B in a global (full-data) context but ranks lower than nation B in a regional (subgroup-data) context? Without carefully examining the meanings and implications of assumptions and properties associated with a weighting method, it would be difficult to evaluate the adequacy of the weighting method. It would also be difficult to derive practical applications of the weighting method to comprehend the meaning of synthetic indicators constructed using the weighting method. In short, conceptual justifications, assumptions and implications for synthetic indicators developed using weighting (including subjective weighting methods, statistical methods and/or multiattribute methods) must be provided and discussed, so adequacy of weighting can be assessed and true meanings of synthetic indicators constructed using weighting can be understood.

Performance of Weighting

In addition to conceptual/theoretical justification, consideration of weighting methods, performance of weighting should also be evaluated when incorporating weighting for synthetic indicators. The main implications that the literature on the topic of weighting in QOL measures offer include: (a) standards used to assess the performance of weighting must be reasonable, (b) analytical methods used to assess the performance of weighting must be justifiable, and (c) performance of weighting of synthetic indicators must be assessed in clear relation to the purpose of the synthetic indicators.

Performance Standards Similar to the criterion variables used in the literature on weighting in QOL measures, performance of weighting in synthetic indicators is dependent upon the assessment standards selected. It is important to provide a clear rationale for selecting specific standards for assessing the performance of weighting. It is also critical to recognize that (similar to the literature in QOL measures) it is unlikely to find existing index/indicators that will capture or represent exactly the same concept that newly developed synthetic indicators intend to capture or represent. Caution, therefore, should be given in interpreting any empirical results on the performance of weighting in synthetic indicators.

Analytical Methods The literature on weighting in QOL measures points to the important role of analytical methods in assessing the performance of weighting. Similarly, adequate and reasonable analytical methods should be developed and adopted to assess the performance of weighting in synthetic indicators. It is likely that for synthetic indicators, like for QOL measures, the performance of weighting depends on the analytical methods chosen. Adequacy of the analytical methods, including the rationale, statistical power, sensitivity and specificity, if applicable, should be provided and addressed.

Purpose of the Synthetic Indicators Just like the performance of weighting in QOL measures cannot be assessed without considering the purpose of the measure, performance of weighting in synthetic indicators should also take into account the purpose of the synthetic indicators. If the synthetic indicators are meant to serve multiple purposes, it may be necessary to prioritize these purposes should the incorporation of weighting results in difficulty in serving the purposes. For example, many synthetic indicators are developed for the purpose of cross-nation comparisons. If, with the application of a specific weighting method, the synthetic indicators produce many ties among these nations, weighting in this case may not necessarily be desirable. Or, if, for the purpose of increasing discriminating power (reducing ties), a weighting method (such as the MGP-BoD method mentioned earlier) produces very different ranking patterns of nations with exactly the same national data, depending on the number of nations included in the weighting procedures, the policy and practical utility of the synthetic indicators may become confusing and unclear.

Conclusion

Incorporating weighting in synthetic indicators can be desirable. The literature on weighting in QOL measures offers a number of critical implications that should be considered regarding incorporating weighting in synthetic indicators. When it comes to the topic of weighting in synthetic indicators, at least three major areas deserve attention: conceptual justification, weighting methods, and performance of weighting. Careful consideration should be given in all three areas in the development and construction of synthetic indicators. Incorporation of weighting in synthetic indicators needs to be conceptually reasonable, methodological adequate and well performed.

References

Aberson, C. L. (2010). *Applied power analysis for the behavioral sciences*. New York: Psychology Press.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–314.

Campbell, A., Converse, P. E., & Rogers, W. L. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. New York: Russel Sage.

Cherchye, L., Moesen, W., Rogge, N., & van Puyenbroek, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research, 82*, 111–145.

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. H. Hole (Ed.), *Statistical strategies for small sample research* (pp. 307–341). Thousand Oaks: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hilsdale: Lawrence Erlbaum Associates.

Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equations casual models. *Applied Psychological Measurement, 14*, 183–196.

Cummins, R. A. (1996). The domains of life satisfaction: An attempt to order chaos. *Social Indicators Research, 38*, 303–328.

Cummins, R. A., McCabe, M. P., Romeo, Y., & Gullone, E. (1994). The comprehensive quality of life scale: Instrument development and psychometric evaluation on tertiary staff and students. *Educational and Psychological Measurement, 54*, 372–382.

Diamantopoulos, A. (2006). The error term in causal measurement models: Interpretation and modeling implications. *Journal of Modeling in Management, 1*, 7–17.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1201–1302.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542–575.
Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life sacle. *Journal of Personality Assessment, 49*, 71–75.
Evans, M. G. (1991). The problem of analyzing multiplicative composites: Interactions revisited. *American Psychologist, 46*, 6–15.
Feist, G. J., Bodner, T. E., Jacobs, J. F., Miles, M., & Tan, V. (1995). Integrating top-down and bottom-up structural models of subjective well-being: A longitudinal investigation. *Journal of Personality and Social Psychology, 68*, 138–150.
Ferrans, C. E. (1990). Development of a quality of life index for patients with cancer. *Oncology Nursing Forum, 17*(3), 15–19.
Ferrans, C., & Powers, M. (1985). Quality of life index: Development and psychometric properties. *Advances in Nursing Science, 8*, 15–24.
Guardiola, J., & Picazo-Tadeo, A. J. (2014). Building weighted-domain composite indices of life satisfaction with data envelopment analysis. *Social Indicators Research, 117*, 257–274.
Hagerty, M. R., & Land, K. C. (2007). Constructing summary indices of quality of life: A model for the effect of heterogeneous importance weights. *Sociological Methods and Research, 35*, 455–496.
Hagerty, M. R., Cummins, R. A., Ferris, A. L., Land, K. C., Michalos, A. C., Peterson, M., et al. (2001). Quality of life indexes for national policy: Review and agenda for research. *Social Indicators Research, 55*, 1–96.
Hardin, A. M., Chang, J. C., Fuller, M. A., & Torkzadeh, G. (2011). Formative measurement and academic research: In search of measurement theory. *Educational and Psychological Measurement, 71*, 281–305.
Headey, B., Veenhoven, R., & Wearing, A. (1991). Top-down versus bottom-up theories of subjective well-being. *Social Indicators Research, 24*, 81–100.
Hsieh, C. M. (2003). Counting importance: The case of life satisfaction and relative domain importance. *Social Indicators Research, 61*, 227–240.
Hsieh, C. M. (2004). To weight or not to weight: The role of domain importance in quality of life measurement. *Social Indicators Research, 68*, 163–174.
Hsieh, C. M. (2012). Should we give up domain importance weighting on QOL measures? *Social Indicators Research, 108*, 99–109.
Hsieh, C. M. (2013). Issues in evaluating importance weighting in quality of life measures. *Social Indicators Research, 110*, 681–693.
Hsieh, C. M. (2015). The often overlooked issue of statistical power: This and other issues regarding assessing importance weighting in quality of life measures. *Social Science Research, 50*, 303–310.
Hsieh, C. M., & Kenagy, G. P. (2014). Measuring quality of life: A case for re-examining the assessment of domain importance weighting. *Applied Research in Quality of Life, 9*, 63–77.
Inglehart, R. (1978). Value priorities, life satisfaction, and political dissatisfaction among western publics. *Comparative Studies in Sociology, 1*, 173–202.
Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*, 199–218.
Kaplan, R. M., Feeny, D., & Revicki, D. A. (1993). Methods for assessing relative importance in preference based outcome measures. *Quality of Life Research, 2*, 467–475.
Lance, C. E., Lautenschlager, G. J., Sloan, C. E., & Varca, P. E. (1989). A comparison between bottom-up, top-down, and bidirectional models of relationships between global and life facet satisfaction. *Journal of Personality, 57*, 601–624.
Land, K. C. (2014). Weighting schemes. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 7028–7029). New York: Springer.
Locke, E. A. (1969). What is job satisfaction? *Organizational Behavior and Human Performance, 4*, 309–336.

Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297–1349). Chicago: Rand McNally.
Locke, E. A. (1984). Job satisfaction. In M. Gruneberg & T. Wall (Eds.), *Social psychology and organizational behavior* (pp. 93–117). London: Wiley.
MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*, 710–730.
Maggino, F., & Zumbo, B. D. (2012). Measuring the quality of life and construction of social indicators. In K. C. Land, A. C. Michalos, & M. J. Sirgy (Eds.), *Handbook of social indicators and quality of life research* (pp. 201–238). New York: Springer.
Mastekaasa, A. (1984). Multiplicative and additive models of job and life satisfaction. *Socical Indicators Research, 14*, 141–163.
McFarlin, D. B., & Rice, R. W. (1992). The role of facet importance as a moderator in job satisfaction processes. *Journal of Organizational Behavior, 13*, 41–54.
McFarlin, D. B., Coster, E. A., Rice, R. W., & Coopper-Alison, T. (1995). Facet importance and job satisfaction: Another look at the range of affect hypothesis. *Basic and Applied Social Psychology, 16*, 489–502.
Michalos, A. C. (2004). Social indicators research and health-related quality of life research. *Social Indicators Research, 65*, 27–72.
Philip, E. J., Merluzzi, T. V., Peterman, A., & Cronk, L. B. (2009). Measurement accuracy in assessing patient's quality of life: To weight or not to weight domains of quality of life. *Quality of Life Research, 18*, 775–782.
Rice, R. W., Gentile, D. A., & McFarlin, D. B. (1991). Facet importance and job satisfaction. *Journal of Applied Psychology, 76*, 31–39.
Rojas, M. (2006). Life satisfaction and satisfaction in domains of life: Is it a simple relationship? *Journal of Happiness Studies, 7*, 467–497.
Russell, L. B., & Hubley, A. M. (2005). Importance ratings and weighting: Old concerns and new perspectives. *International Journal of Testing, 5*, 105–130.
Russell, L. B., Hubley, A. M., Palepu, A., & Zumbo, B. D. (2006). Does weighting capture what's important? Revisiting subjective importance weighting with a quality of life measure. *Social Indicators Research, 75*, 146–167.
Ryff, C. D., & Essex, M. J. (1992). The interpretation of life experience and well-being: The sample case of relocation. *Psychology and Aging, 7*, 507–517.
Saisana, A. (2014). Composite indicators. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1156–1161). New York: Springer.
Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research, 81*, 65–77.
Sayed, H., Hamed, R., Ramadan, M. A.-G., & Hosny, S. (2015). Using meta-goal programming for a new human development indicator with distinguishable country ranks. *Social Indicators Research, 123*, 1–27.
Shultz, K. S., & Whitney, D. J. (2004). *Measurement theory in action: Case studies and exercises*. Thousand Oaks: Sage.
Trauer, T., & Mackinnon, A. (2001). Why are we weighting? The role of importance ratings in quality of life measurement. *Quality of Life Research, 10*, 579–585.
Wu, C. H. (2008a). Examining the appropriateness of importance weighting on satisfaction score from range-of-affect hypothesis: Hierarchical linear modeling for within-subject data. *Social Indicators Research, 86*, 101–111.
Wu, C. H. (2008b). Can we weight satisfaction score with importance ranks across life domains? *Social Indicators Research, 86*, 468–480.
Wu, C. H., & Yao, G. (2006a). Do we need to weight satisfaction scores with importance ratings in measuring quality of life? *Social Indicators Research, 78*, 305–326.
Wu, C. H., & Yao, G. (2006b). Do we need to weight item satisfaction by item importance? A perspective from Locke's range-of-affect hypothesis. *Social Indicators Research, 79*, 485–502.

Wu, C. H., & Yao, G. (2007). Examining the relationship between global and domain measures of quality of life by three factor structure models. *Social Indicators Research, 84*, 189–202.

Wu, C. H., & Yao, G. (2014). Scoring weights. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 5729–5731). New York: Springer.

Wu, C. H., Chen, L. H., & Tsai, Y. M. (2009). Investigating importance weighting of satisfaction scores from a formative model with partial least squares analysis. *Social Indicators Research, 90*, 351–363.

Chapter 11
The Role of Normalisation in Building Composite Indicators. Rationale and Consequences of Different Strategies, Applied to Social Inclusion

Ludovico Carrino

JEL Classification C43 • C83 • D63 • I32

Introduction

Although there is quite a consensus on the need for broadening the scope of the analysis of Well-being beyond the monetary dimension (see, e.g., the influential report by Stiglitz et al. (2010)), there is not equal agreement on how such an ambitious task should be operationalized. It is well known that subjectivity and arbitrariness exist with respect to the choice of the dimensions to be included in the composite index, the normalisation of the variables, and the characterisation of the aggregation function (see, e.g., Ravallion (2012a), Decancq and Lugo (2013), Martinetti and von Jacobi (2012)).[1] The socio-economic literature highlighted

[1]The act of synthesizing a composite latent phenomenon encompasses methodological issues that have economic, philosophical (as well as psychological) and political connotations. Indeed, these issues arise from a fundamental mismatch between the kind of multiplicity inherent in the latent concept and the multiplicity characterizing the forged measure (the result of the researcher's work). In a sense, the latent multidimensional concept (e.g., Well-being or Social Inclusion) is an un-synthesized multiplicity, in that it is composite by nature and perceived as a whole by the human sensibility. Since the phenomenon is unmeasurable per se, the researcher is forced to separate it, operationally, in numerous measurable components, in order to aggregate them back to provide a proxy of the latent phenomenon. In other words, building a synthetic index of Well-being

L. Carrino (✉)
Department of Global Health and Social Medicine, King's College London, The Strand, London WC2R 2LS, UK

Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, F.ta S.Giobbe, Venice 30121, Italy
e-mail: ludovico.carrino@kcl.ac.uk

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_11

that no unanimous method exists to perform such choices, pointing out numerous theoretical issues (Stiglitz et al. 2010; Ravallion 2011, 2012a; Klugman et al. 2011; Maggino and Nuvolati 2012; Decancq and Lugo 2013), testing empirical robustness (Kasparian and Rolland 2012; Lefebvre et al. 2010; Saisana et al. 2005; Ravallion 2012b). Yet, although there may be no "absolute cure" for multidimensional evaluations, a good practice could consist in enhancing methodological transparency (Sen and Anand 1997).

In this paper we focus on the crucial phenomenon of "enlarged" poverty conditions in Europe, referred to as Social Inclusion, for which the need for a synthetic measure has been often expressed by the European Commission. We rely on the theoretical work done by (Atkinson et al. 2002) and the European Commission (European Council 2000, 2001) and we build a composite measure of Social Inclusion, whose components are life expectancy at birth, early school-leaving, long-term unemployment and poverty rate, for 63 European administrative regions for 2012, using data from EUROSTAT.[2] Social inclusion is, indeed, multidimensional in that it encompasses income poverty, unemployment, access to education, information, childcare and health facilities, living conditions, as well as social participation. It is also multi-layered, as its causes can be identified at the national, community, household or individual levels.

While the major focus of the recent literature on composite measures has been devoted to the choice of the dimensions' weights, few studies have concentrated on the role played by normalisation in influencing the final results (Lefebvre et al. 2010; Saisana et al. 2005). Our contribution highlights that, in fact, normalisation is a crucial stage where an "early" implicit weighting takes place, which can strongly affect the overall results of the multidimensional analysis. Therefore, the unavoidable arbitrariness regarding the choice of the normalisation function should be made as transparent as possible.

In this analysis, we adopt a baseline linear aggregation model where the normalised components have equal weights and we look at what happens to the aggregate measure of Social Inclusion when the sole normalisation function changes. In particular, we apply two widely used *data-driven* normalisation strategies, min-max and Z-score, whose parameters depend on the available data. These data-driven functions generate implicit trade-offs (between the index' components) and shadow prices with weak economic justification. We also propose a novel strategy, an *expert-based* min-max function, whose parameters are grounded on the responses to a survey conducted on a population of 150 professors of Economics or Management at the Ca' Foscari University of Venice.

requires that the indeterminate nature of multiplicity is made determinate through a specification of its contents, and of their relationship.

[2]Through this analysis, we do not aim at providing efficiency index for the Welfare States, which would require a much more structured set of information. We, rather, limit ourselves at evaluations of performances, as suggested by Pestieau (2009) and Lefebvre et al. (2010).

Our results indicate that, even within a simple-average framework, changing the normalisation function substantially affects the relative relevance of each component of the aggregate measure. As a consequence, significant differences emerge in the levels and rankings of regional Social Inclusion in Europe, leading to very different policy implications. The data-driven strategies soften the heterogeneities *within* and *between* European countries by putting a substantial weight on the longevity variable rather than on educational and economic statuses. As a result, the European regional distribution of Inclusion appears to be uni-modal around the mean. Conversely, the expert-based normalisation emphasises the unemployment and the school-dropouts variables, and returns a bi-modal distribution of Social Inclusion. We, thus, discuss how the normalization stage should be made as transparent as possible to the reader, and how the different premises of the two strategies characterise the interpretation of the results: the data-driven approaches allows for a positive interpretation of the index, while the survey-driven approach allows for a normative one. In other words, if the index' intrinsic trade-offs are grounded on statistical terms, its results should be interpreted accordingly.

The remaining of the chapter is organized as follows. Section "Social inclusion, definition and sample selection" briefly describes the concept of Social Inclusion and the data. Section "Aggregation framework" sets a standard framework for multidimensional aggregation and details the baseline model. Section "Normalization framework" introduces the normalisation strategies, while section "Implicit trade-offs from normalization" discuss the implicit trade-offs resulting from applying the aforementioned normalisation functions on the baseline model. Section "Results" details the results of the Social Inclusion indices, section "Weights and interpretation of results" concludes.

Social Inclusion, Definition and Sample Selection

Social Inclusion (as its corresponding opposite, Social Exclusion)[3] is one of the five priorities selected by the European Commission in the context of the Europe 2020 Strategy. A definition of Exclusion was already drawn in December 1992 by the Commission of the European Communities (European Communities Commission 1992): "Social Exclusion is a multidimensional phenomenon stemming from inadequacies or weaknesses in the services offered and policies pursued in these various

[3]The concept of Social Inclusion/exclusion should not be confused with the variable 'at risk of poverty or social exclusion' in the Eurostat database, which defines an individual as at risk of poverty or social exclusion when at least one of the following conditions hold: (a) equivalent household income below 60% of national median; (b) households with at least 4 of the following 9 issues: (i) impossibility to bear unexpected expenses, (ii) cannot afford a week holiday, (iii) issues with the mortgage, rent, bills; (iv) cannot afford a proper meal every 2 days; (v) not able to adequately heat the house; (vi) not able to afford a washing machine (vii) a color TV (viii) a phone (ix) an automobile; (c) living in families whose members aged 18–59 work less than a fifth of their time.

policy areas. Such insufficiencies and weaknesses often combine to affect both people and regions via cumulative and interdependent processes of such a nature that it would be futile to try to combat exclusion by tackling only one of its dimensions. More clearly than the concept of poverty, (. . .) it states out the multidimensional nature of the mechanisms whereby individuals and groups are excluded from taking part in the social exchanges, from the component practices and rights of social integration and of identity".

The Laeken European Council in 2001 has developed a set of unanimously agreed indicators that could capture the multifaceted aspects and outcomes of Social Inclusion, thus providing reliable and comparable data to monitor the social and economic conditions of European citizens (European Council 2001), through the Open Method of Coordination. In particular, four basic dimensions have been identified: the level and distribution of income, the performance in the labour-market, education and health areas. For each of them, a set of primary indicators were adopted: income (Poverty rate (after social transfers), Persistent risk-of-poverty rate, Relative median at risk-of-poverty gap, Inequality of income distribution); labour market (Long-term unemployment, Regional cohesion, Persons living in jobless households); education (early school leavers); health (life expectancy at birth, Self-defined health status by income level).[4]

The target of this paper is to build an aggregate index of Social Exclusion at administrative-regional level in Europe. We choose administrative regions as the main territorial unit of this analysis, with the aim of capturing higher variability than it can be inferred from aggregate national data. Data-availability is often mentioned as a serious constraint for analyses which focus on a wide set of countries for a long time-period (Lefebvre et al. 2010; Martinetti and von Jacobi 2012). In the context of Social Exclusion at administrative regional level, we are able to gather data for four out of the 10 aforementioned indicators, one per dimension: poverty-rate, long-term unemployment, early school-leavers and life expectancy at birth. Our data-source is the on-line Eurostat Regional Database 2015, and the most recent data are available for all the four variables for the year 2012, for 63 administrative regions in five countries (Belgium, Denmark, Germany, Italy and Spain). For other countries (e.g., Greece, France, Czech Republic and Norway), data were either not available for all the indicators, or they are available for statistical-regions, but not for administrative regions.[5]

As argued in Lefebvre et al. (2010), "these indicators cover the most relevant concerns of a modern welfare state, also reflecting aspects that people who want to enlarge the concept of GDP to better measure social welfare generally take into

[4]We refer to (Atkinson et al. (2002), (2004)), as well as to European Commission (2009), (2010) for further details on the rationale of Social Inclusion indicators and on the issues related to their measurement.

[5]As a robustness check we enlarged the sample with data for statistical-regions for Czech Republic, Greece, Norway and the Netherlands, without any significant change to the results of the analysis. Besides, as stated in the introduction, the purpose of this chapter is to offer a methodological discussion that can be applied to composite analyses in various fields and from various data-selections.

Table 11.1 Variable definitions

Variable	Definition
Poverty rate	Share of individuals living in households with an income below 60% national median equivalised disposable income.
Long-term unemployment rate	Total long-term unemployed population (≥12 months; ILO definition) as proportion of total active population.
Early school-leavers	Share of total population of 18–24-year olds having achieved ISCED level 2 or less and not attending education or training.
Life expectancy at birth	Number of years a person may be expected to live, starting at age 0.

Table 11.2 Descriptive statistics for four indicators of Social Inclusion, year 2012

	Longevity	Early school leaving	Long-term unemployment	At-risk-of-poverty rate
BE	80.5 (1.3)	11.8% (4.8)	3.4% (3.2)	15.3% (9.3)
DE	81 (0.6)	10.6% (2.4)	2.5% (1.4)	15.1% (3.4)
DK	80.2 (0.4)	24.5% (4.7)	2.1% (0.2)	13.1% (1.1)
ES	82.5 (1)	1% (0.7)	11.2% (2.9)	22.2% (7.6)
IT	82.4 (0.7)	17.5% (3.8)	6.1% (3.6)	19.6% (11.6)

Source: Eurostat Regional Database 2015. Country labels correspond the ISO 3166–1 alpha-2 standard

account". The latter referenced paper discusses, as do Atkinson et al. (2004), the limitations of these data and the necessary simplifying assumptions that have to be done when translating a complex multidimensional phenomenon like Social Exclusion in empirical terms. Table 11.1 provides a brief definition for our four variables:

The following table and figure report descriptive statistics on the four indicators in our sample (Table 11.2).

Appendix A includes further descriptive statistics on correlations and data-distribution of the selected variables in the sample-data.

Aggregation Framework

Let us consider *m dimensions* (hereinafter also *variables*, *attributes, components*) of Social Inclusion, observed for n regions. For a generic region *i* we can therefore build the vector $\mathbf{x}^i = \left(x_1^i, \ldots, x_m^i\right)$, while $\mathbf{X} \in \mathbb{R}^{n \times m}$ is the distribution matrix of *m* attributes for *n* regions. To retrieve an aggregated measure for region *i*, we consider the function *F* defined as:

$$F^i\left(v\left(\mathbf{x}^i\right)\right) = \left[w_1 v_1\left(x_1^i\right)^\beta + \cdots + w_m v_m\left(x_m^i\right)^\beta\right]^{1/\beta} \tag{11.1}$$

which is often referred to as a CES (constant elasticity of substitution) function, or a generalized mean of order β. Its arguments are the elements $v_1, \ldots, v_m$ which are transformations of the original variables $x_1,\ldots,x_m$ (defined hereafter). The function F is non-decreasing, separable, weakly scale-invariant and homogenous of degree-one in its arguments v; we refer to Blackorby and Donaldson (1982) and (Decancq and Lugo (2009), 2008)) for an analytic characterization of these properties.

The parameters $w_1,\ldots,w_m$, the weights of the normalised dimensions v, are non-negative and sum to one.

Provided that a choice of the m dimension has been performed, the main methodological task is now the selection of the set of functions $v_1,\ldots,v_m$, of parameters $w_1,\ldots w_m$, as well as of β.

Baseline Linear Aggregation Model

The parameter β in (11.1) determines the elasticity of substitution $\varepsilon_{k,j}$ between any pair (v_k, v_j). In the CES function, the elasticity between any pair k, j is constant and equal to $1/1$-β. The elasticity of substitution determines the percentage change in v_j/v_k, which would result from a percentage change in the slope along a level-set (the marginal rate of substitution, MRS, along an indifference curve). The parameter β must be lower than one to generate iso-inclusion contours convex to the origin in the two-dimensional region of the space of attributes (Bourguignon and Chakravarty 2003). The smaller is β, the higher is the increase in dimension v_j needed to keep constant the overall index after a one-unit decrease in dimension v_k.

Since the focus of this chapter is on the normalisation choices, let us adopt a standard aggregation framework by setting $\beta = 1$ in (11.1), therefore obtaining a linear weighted average with linear indifference curves, constant MRS and infinite elasticity of substitution between pairs of normalised dimensions. We also let the weights w_j be equal, i.e., $w_1 = \ldots = w_m = 1/m = 1/4$ (since $m = 4$ in our case study). The resulting model will be, for a generic region i (time subscripts are omitted) an aggregation function L, as in *linear*, such as:

$$L^i\left(\nu\left(\mathbf{x}^i\right)\right) = \frac{1}{m}\nu_1\left(x_1^i\right) + \cdots + \frac{1}{m}\nu_m\left(x_m^i\right) \tag{11.2}$$

The arbitrary choice of setting equal weights is a widely adopted strategy in the literature of multidimensional measurement. As Hoskins and Mascherini (2009) and Decancq and Lugo (2013) highlight, this approach is often justified with the argument that all the dimensions are equally important (Atkinson et al. 2002) or, conversely, that there is insufficient knowledge for setting a more detailed weighting scheme (sometimes referred to as an "agnostic view"). Although being frequently described as a *simple* and relatively *neutral* strategy, "equal weighting" does not mean "no weighting", because it involves an implicit judgment on the weights being equal, and because it often applies just on the normalised dimensions of the index.[6]

In the following Sections, we will investigate how original attributes contribute to the overall measure, and what characterizes the relationship between attributes within the linear framework. In general, such effects can be retrieved via the partial derivative of the aggregate measure L with respect to variable x_j (region-specific indices are dropped for convenience), as follows:

$$\frac{\partial L(v(\mathbf{x}))}{\partial x_j} = w_j v_j'(x_j) \tag{11.3}$$

From (11.3) we can identify two main drivers that determine how the aggregate measure L reacts at small changes in the j-th real-valued dimension x_j. First, the higher is the weight of the normalised j-th dimension, the higher will be the marginal variation in the L. Second, the steeper is the normalisation function, the higher will be the effect of a change in the j-th dimension on the aggregate measure.

Within the linear aggregation function L, the marginal rate of substitution between a pair of observed-indicators x_j and x_k is:

$$MRS_{x_k,x_j} = -\frac{\mathrm{d}x_j}{\mathrm{d}x_k} = \frac{\frac{\partial L(v(\mathbf{x}))}{\partial x_k}}{\frac{\partial L(v(\mathbf{x}))}{\partial x_j}} = \frac{w_k}{w_j}\frac{v_k'(x_k)}{v_j'(x_j)} \tag{11.4}$$

Both the marginal contribution of the j-th attribute and its MRS depend on the shape of the normalisation function v_j. If, however, the transformation function is the identity function ($v_j(x_j) = x_j$), the effect of a change in x_j can be uniquely determined by its weight w_j, while the MRS between a pair of dimensions j and k is determined by the ratio between their weights.[7]

Normalization Framework

Raw variables are usually observed and measured with different measurement units. The component $v_i(x_j)$ is a weakly monotonic and continuous normalisation function that maps the values of the j-th variable x_j on the closed interval [0,100],

[6]In the words of Martinetti and von Jacobi (2012), the implicit assumption for equal weighting is that "in absence of any objective mechanism for determining the relative importance of the considered dimensions, the most neutral method is assigning an equal weight to each of them". Indeed, both Chowdhury and Squire (2006) and Nguefack-Tsague et al. (2011) provide evidence in favour of equal weighting after collecting expert preferences.

[7]The MRS between two observed dimensions will be equal to their "weights" also if the derivatives of their normalisation functions are equal, i.e., if $v_k'(x_k)/v_j'(x_j) = 1$

i.e., $v_j(x_j) \in [0,100]$. Moreover, attributes might be, alternatively, positively or negatively related to the latent phenomenon, i.e., their may have a positive or negative *polarity*. Hence, in order to ensure comparability and monotonicity of any aggregation function, each variable must be normalised such that better performances in the *j-th* dimension correspond to non-lower values of $v_j(x_j)$ and therefore of the aggregated value L. In other words, each normalised variable should have a positive polarity. The normalisation function thus ensures that L is bounded between 0 and 100 when the weights w sum to one. In what follows we will briefly present three normalisation strategies, and we refer the reader to Giovannini et al. (2008) for their detailed description. Two out of three are "data-driven" strategies, that is, transformations whose characteristics are entirely determined by the data at hand. The third one, conversely, is defined through the elicitation of explicit value judgements.

Two Data-Driven Normalization Functions

The Min-Max Function

The *min-max normalisation function* is widely used in the literature of multidimensional measures (see, e.g., Cherchye et al. (2007), Silva and Ferreira-Lopes (2013), Pinar et al. (2014), Mazziotta and Pareto (2015)), as well as in the Human Development Index (Anand and Sen 1994) and in the OECD Better Life (Boarini and D'Ercole 2013).

For each attribute x observed in region i at a time t (we drop the previously used attribute-specific j index to ease readability), the corresponding normalised value is defined as:

$$\begin{aligned}
&\nu_+^{i,t}(x_+) = 100^* \frac{x_+^{i,t} - b_+\min(x_+)}{b_+\max(x_+) - b_+\min(x_+)}\\
&\nu_-^{i,t}(x_-{}^{i,t}) = 100^* \frac{b_-\max(x_-) - x_-{}^{i,t}}{b_-\max(x) - b_-\min(x)}\\
&\nu_+^{i,t}(x_+^{i,t}) = 0 \text{ if } x_+^{i,t} \le b_+\min(x_+)\\
&\nu_-^{i,t}(x_-^{i,t}) = 0 \text{ if } x_-^{i,t} \ge b_-\max(x_-)\\
&\nu_+^{i,t}(x_+^{i,t}) = 100 \text{ if } x_+^{i,t} \ge b_+\max(x_+)\\
&\nu_-^{i,t}(x_-^{i,t}) = 100 \text{ if } x_-^{i,t} \le b_-\min(x_-)
\end{aligned} \tag{11.5}$$

where ν_+ is used when x has positive polarity (i.e., it is a "good") and ν_- is used when x has negative polarity (i.e., it is a "bad").

The coefficients bmin_i and bmax_i are the highest and lowest values to be used as benchmarks for the x variable for region i. Regardless on how the benchmarks are defined, it is straightforward that, when x has positive polarity, b_+max corresponds to a more desirable performance than $b_{+}min$, while the opposite is true when x has negative polarity. The *min-max* strategy rescales indicators into an identical range [0,100].[8] E.g., for x "good", 0 is given to values lower or equal to $b_{+}min$, while 100 is given to those higher or equal to $b_{+}max$. The values within these benchmarks are proportionally converted into the 0–100 scale. Hence, ν is a stepwise continuous function.

The data-driven min-max normalisation (11.6) defines the benchmarks *min* and *max* as the best and worst observed performance among selected regions (Lefebvre et al. 2010; Silva and Ferreira-Lopes 2013; Murias et al. (2012)), and across a time-series, in order to take into account the evolution of indicators and offer time-comparability (Giovannini et al. 2008). In our case study, this corresponds to assigning a value of 0 to the region which reports the worst-observed performance in the period from 2004 to 2012, while assigning a value of 100 to the "best-observed" one.

For each region i where an attribute x is observed at a time t, the corresponding normalised value $\nu_{dM}^{i,t}(x^{i,t})$, where the subscript *dM* stands for "data-driven min-max", is determined as:

$$\begin{aligned} \nu_{dM+}^{i,t}\left(x_{+}^{i,t}\right) &= 100 * \frac{x_{+}^{i,t} - \min\limits_{t \in T} \min\limits_{i}\left(x_{+}^{t}\right)}{\max\limits_{t \in T} \max\limits_{i}\left(x_{+}^{t}\right) - \min\limits_{t \in T} \min\limits_{i}\left(x_{+}^{t}\right)} \\ or\ \ \nu_{dM-}^{i,t}\left(x_{-}^{i,t}\right) &= 100 * \frac{\max\limits_{t \in T} \max\limits_{i}\left(x_{-}^{t}\right) - x_{-}^{i,t}}{\max\limits_{t \in T} \max\limits_{i}\left(x_{-}^{t}\right) - \min\limits_{t \in T} \min\limits_{i}\left(x_{-}^{t}\right)} \end{aligned} \tag{11.6}$$

where x_{+} and x_{-} have the usual meaning of a "good" and a "bad" attribute, respectively.

Table 11.3 displays the data-driven thresholds for our sample of regions[9]:

[8]The choice of multiplying by 100 eases readability of the results in the remaining of the paper, and does not affect any result.

[9]The autonomous cities of Ceuta and Medilla, located on the Mediterranean coast of Morocco but belonging to Spain since fifteenth century, are substantially different from other Spanish regions. Given that their values for school-dropouts, long-term unemployment and poverty rate are sensibly higher than the rest of the sample, we prudently decided to treat them as outliers and exclude them from the computation of the thresholds. This decision has no significant consequences on the results of the paper, nor on its implications. Including them in the sample would raise the maximum values for early school-leaving rate to 54.2% (Ceuta 2005), for long-term unemployment to 18.2% (Ceuta 2012), and for poverty rate to 48.9% (Ceuta 2008). A graphical distribution of the data used for the min-max normalisation is reported in Fig. 11.6 (Appendix A)

Table 11.3 Data-driven benchmarks

	Observed minimum	Observed maximum
Longevity *(good)*	77.5 (Région wallonne)	84.2 (Comunidad de Madrid 2012)
Early school leaving *(bad)*	5.4% (Thüringen 2009)	42.8% (Murcia 2004)
Long-term unemployment *(bad)*	0.3% (Midtjylland 2008)	15.3% (Canarias 2012)
At-risk-of-poverty rate *(bad)*	5.2% (Valle d'Aosta 2006)	44.3% (Sicilia 2011)

Note: data from the Eurostat Database 2015 (2004–2013). Data for autonomous cities of Ceuta and Medilla (Spain) are excluded (see footnote 9)

Standardisation (Z-score)

Among the normalization functions alternative to the min-max, there are some which do not force the normalized variables within a predetermined range, but rather limit their variability. Here we will considered the widely used Z-score standardisation (Mazziotta and Pareto 2015; Hoskins and Mascherini 2009). For each attribute x observed in region i at a time t, the corresponding normalised value is defined as:

$$v_{Z\pm}^{i,t}\left(x^{i,t}\right) = 100 \pm \frac{x^{i,t} - \bar{x}}{\sigma(x)} \tag{11.7}$$

where the sign in front of the ratio depends on whether x has positive or negative polarity, and where $\bar{x}$ and $\sigma(x)$ are, respectively, the average and the standard deviation computed across all the regions and all the years (i.e., between 2004 and 2012).[10]

What characterizes the Z-score standardization is that all normalized variables have the same average value (here set at 100) and a unitary standard deviation across regions. Moreover, the actual minima and maxima of the normalized variables across regions are not bounded, conversely to what happens in the min-max procedure.

The following table reports the average values and standard deviations adopted for the standardization (Table 11.4).

[10] A common specification for the Z-score normalisation in time-dependent studies (as detailed, e.g., in OECD and European Commission (2008)) adopts as references the averages and the standard deviations across countries for a given reference year. We chose to compute both the references across countries *and* time, to be consistent with strategy followed in designing the data-driven min-max (where the same aforementioned OECD report suggests to adopt minima and maxima across countries and time). As a robustness test, we also computed the Z-score values using as references the averages and the standard deviations across countries for the year 2012. Such change has no consequences on the results and implications of our analysis.

Table 11.4 Data-driven benchmarks for the standardisation

	Average	Standard deviation
Longevity *(good)*	81	1.2
Early school leaving *(bad)*	18.1%	8.1
Long-term unemployment *(bad)*	4.3%	3
At-risk-of-poverty rate *(bad)*	17.5%	8.2

Note: data from the Eurostat Database 2015 (2003–2012). Both statistics are computed across regions and time

A Novel Expert-Based Normalisation Function

Consistently with what is often debated with respect to the aggregation function, the parameters of the normalisation function can either reflect a predetermined choice by the researcher herself (e.g., through a *data-driven* strategy), or be elicited from some stakeholders group, e.g., field-experts, members of institutions, citizens (Kim et al. (2015) and Decancq and Lugo (2013) produce a recent review of elicitation strategies).

In a simple linear model with $\beta = 1$ and $w_j = 1/m$ for each *j-th* indicators, the crucial determinant of a dimension's relevance relies heavily on the normalisation function, as visible from (11.3). As we will discuss in the next Section, if the function's parameters are data-driven, then their implications in terms of dimensions' weights and MRS are to be interpreted under a mathematical perspective, yet it is harder to determine what do they reflect in economic terms (Lefebvre et al. 2010). As an example, in the data-driven min-max, a variable with transformed-value equal to "0" just implies it being "the last one", or "the worst one" observed among the available data, which does not necessarily corresponds to an undesirable condition of Well-being. A similar reasoning, with opposite meaning, can be done for normalised values of "100".

An alternative to the data-driven normalisation would require to incorporate some value judgments in the normalisation (e.g., goalposts, see Bertin Carrino and Giove (2016) and Mazziotta and Pareto (2015)). This translates to linking the extreme values "0" and the "100" with, e.g., a certain definition of desirability, thus making the normalisation independent from the data. When an indicator lies above or below such fixed bounds, further variations do not contribute to the latent variable under study (see e.g., the discussion in Anand and Sen (1994), Klugman et al. (2011), Ravallion (2012b), Lefebvre et al. (2010). A major example of fixed threshold is the Human Development Index that, since 1994, adopted "goalposts" as minimum and maximum values in the normalisation function. The interpretation behind these fixed thresholds relies on the belief that objective upper and lower bounds can be identified and defined as "subsistence" minimum or "satiation" points, beyond which additional increments would not contribute to the expansion of capabilities.

Contrary to Human Development, Social Exclusion's concept has been developed with reference to advanced industrialized economies, as are those of the

European Union members. Therefore, rather than on "subsistence", its focus is posed on the "unacceptability" and "undesirability" of living conditions, as in an enlarged definition of poverty. Accordingly, a positive threshold for each of our four social-inclusion attributes would refer to a "certainly desirable and favourable condition of Well-being", to which a normalised value of 100 would correspond. Conversely, a negative threshold would refer to a "certainly undesirable and harmful conditions of Well-being", corresponding to a normalised value of 0.

In order to select the actual thresholds, we chose to elicit expert preferences through a survey, rather than to pre-determine them in a top-down fashion. To the best of our knowledge, this is a strategy rarely applied to normalisation stage, and mostly adopted for the aggregation phase instead.

Following Chowdhury and Squire (2006) and Hoskins and Mascherini (2009) (who, indeed, both elicit weights on aggregation rather than on normalisation), we intended to involve informed opinions and therefore selected the population of professors and researchers in the Departments of Economic and Management of the Ca' Foscari University of Venice. Specifically, our population consisted of 149 professors (57 + 38 full or associate professors of Economics and Management, respectively; 29 + 25 assistant professors, *ricercatore universitario*, of Economics and Management, respectively).[11] As for any expert sample, issues could be raised on our group's capability of ensuring all values of efficiency, equity and democracy in the elicitation process. As Kim et al. (2015) pointed out, there is no elicitation method that can ensure all the aforementioned problems. Moreover, being concerned with democratic representativeness, one could argue that greater citizen participation were required; nevertheless, such strategy would likely cause loss of efficiency and quality of the elicitation, together with a lower degree of representativeness (given the resources' constraints). Conversely, we selected a narrow population with specific characteristics but with a working experience that is, at least, partially related with the issues involved in Social Inclusion. Moreover, thanks to an adequate response rate, we are able to statistically represent it.

The survey was worded in Italian and conducted in electronic-form with the QUALTRICS software, a web-based tool that enables users to build custom surveys and distribute them via email.[12] Participants were invited with an email including a link to take part to the on-line questionnaire on an anonymous basis. Appendix A includes further details on the survey pages and wording.

[11]Although, in principle, it would be of interest to widen the Survey population to professors of other Departments (Asian and North African Studies, Environmental Sciences, Humanities, Linguistic, Molecular Sciences and Philosophy), we were led by time and resources constraints to focus on those Faculty more specifically connected to the issues of Social Inclusion and to the disciplines related to the four indicators over which a judgment was asked.

[12]For further details, please refer to http://www.qualtrics.com/

Table 11.5 Survey-elicited benchmarks

	Median elicited minimum	Median elicited maximum
Longevity	73 years	83 years
Early school leaving	10%	20%
Long-term unemployment	3%	9%
At-risk-of-poverty rate	5%	20%

Note: details on the survey's results are available in Appendix A

The Expert-Based Thresholds and the Min–Max Normalisation Function

We implement a min-max normalisation as in (11.5), where the benchmarks corresponds to the median values elicited through the Qualtrics survey (Table 11.5). In particular, the favourable threshold for life expectancy is chosen at 83 years old, while the negative threshold is 73 years old. Early school-leaving's range lies between 10% (which corresponds to the EUROPE 2020's target for members of the European Union). A rate of 9% (or higher) of long-term unemployment denotes a median certainly undesirable condition, while the positive threshold is determined at 3%. As for poverty rate, a certainly harmful level has its median value at 20%, while desirability corresponds to 5% (or lower) share of population below the poverty line set by the Eurostat.

The interquartile ranges are always relatively small, except for the negative threshold for early-school-leaving (15%–25%). Nevertheless, we are aware that no "true values" exist, with respect to these thresholds. In the words of Mascherini and Hoskins (2008), "the judgment of one of the outline may be correct, and those who share a consensus view may be wrong".

A quick comparison of Tables 11.3 and 11.5 suggests that "certain desirability" and "certain undesirability" largely differ from observed minimum or maximum achievements. Indeed, the lowest observed level of longevity (77.4 years) is considered to be "certainly undesirable" just by a small fraction of respondents (Fig. 11.8 in Appendix A). Similarly, any rate of long-term unemployment beyond 9%, or of school dropouts higher than 20%, or of poverty-rate beyond 20%, is regarded as certainly negative, while the actual observed maximums are quite higher. A capping on the positive threshold occurs for those regions which report long-term unemployment lower than 3% or early school leaving rates lower than 10%.[13]

For each region i where an attribute x is observed at a time t, the corresponding normalised value $\nu_{sM}^{i,t}(x^{i,t})$, where the subscript sM stands for "survey-driven min-max", is determined as:

[13]No territories in our sample reach 5% poverty-rate or 73 years in longevity-at-birth, so no "positive" capping occurs.

$$\nu_{sM+}^{i,t}\left(x_{+}^{i,t}\right) = 100 * \frac{x_{+}^{i,t} - \text{lowest threshold}}{\text{thresholds}' \text{ range}}$$
$$or \;\; \nu_{sM-}^{i,t}\left(x_{-}^{i,t}\right) = 100 * \frac{\text{highest threshold} - x_{-}^{i,t}}{\text{thresholds}' \text{ range}}$$
with
$$\nu_{sM+}^{i,t}\left(x_{+}^{i,t}\right) = 0 \text{ if } x_{+}^{i,t} < \text{lowest threshold and } \nu_{sM+}^{i,t}\left(x_{+}^{i,t}\right) = 100 \text{ if } x_{+}^{i,t} > \text{highest threshold}$$
$$\nu_{sM-}^{i,t}\left(x_{-}^{i,t}\right) = 0 \text{ if } x_{-}^{i,t} > \text{highest threshold and } \nu_{sM-}^{i,t}\left(x_{-}^{i,t}\right) = 100 \text{ if } x_{-}^{i,t} < \text{lowest threshold} \tag{11.8}$$

where x_+ and x_- have the usual meaning of a "good" and a "bad" attribute, respectively.

The expert-based normalisation is closer in nature to a "social value-functions", in that the rescale is performed according to how much a value fulfils a "desirability" requirement. Moreover, the normalisation function may become weakly monotonic (instead of being *strongly monotonic* as the data-driven min-max), when the elicited constraints are binding for some observed variable. Indeed, there are regions having attributes with observed performances outside the elicited boundaries, which will receive a normalised value of 100 or 0. Therefore, as we will discuss in the next Section, when an attribute's value lies outside the thresholds, its marginal contribution to the aggregate measure is zero.[14]

Implicit Trade-offs from Normalization

When implementing the min-max normalisation, both in the data-driven (11.6) and in the survey-driven (11.8) setup, and the Z-score standardisation (11.7) in the baseline linear model with "equal weighting" (11.2) we obtain three aggregation functions: LD (linear, data-driven min-max), LZ (linear, Z-score standardisation), LS (linear, survey-driven min-max). For a generic region *i*, such aggregation functions take the following form:

$$\begin{aligned} &LD^{i}(\nu_{dM}(\mathbf{x}^{i})) \\ &= 100\left(0.25 * \frac{x_1 - 77.4}{83.2 - 77.4} + 0.25 * \frac{41.2 - x_2}{41.2 - 6.5} + 0.25 * \frac{13 - x_3}{13 - 0.3} + 0.25 * \frac{23.1 - x_4}{23.1 - 9.5}\right) \end{aligned} \tag{11.9}$$

[14] The min-max normalisation function can be smoothed, in order to avoid the step-wise shape (see, e.g., the discussion in Ravallion (2012b), Lefebvre et al. (2010), Martinetti and von Jacobi (2012), Meyer and Ponthière (2011) and Pinar et al. (2014))

$$LZ^i(v_Z(\mathbf{x}^i)) = 100\left(0.25 * \frac{x_1 - 80.3}{1.1} + 0.25 * \frac{14.7 - x_2}{7.3} + 0.25 * \frac{3.2 - x_3}{2.7} + 0.25 * \frac{15.3 - x_4}{3.4}\right) \tag{11.10}$$

$$\begin{aligned} LS^i(\nu_{sM}(\mathbf{x}^i)) &= 100\left(0.25 * \frac{x_1 - 73}{83 - 73} + 0.25 * \frac{20 - x_2}{20 - 10} + 0.25 * \frac{9 - x_3}{9 - 3} + 0.25 * \frac{20 - x_4}{20 - 5}\right) \\ &\text{with} \\ &\nu_{sM}(x_1) = 0 \text{ if } x_1 < 73 \text{ and } \nu_{sM}(x_1) = 100 \text{ if } x_1 > 83 \\ &\nu_{sM}(x_2) = 0 \text{ if } x_2 > 20 \text{ and } \nu_{sM}(x_2) = 100 \text{ if } x_2 < 10 \\ &\nu_{sM}(x_3) = 0 \text{ if } x_3 > 9 \text{ and } \nu_{sM}(x_3) = 100 \text{ if } x_3 < 3 \\ &\nu_{sM}(x_4) = 0 \text{ if } x_4 > 20 \text{ and } \nu_{sM}(x_4) = 100 \text{ if } x_4 < 5 \end{aligned} \tag{11.11}$$

Before implementing such models on the sample data, it is useful to highlight the implicit economic and statistical mechanisms acting beyond these aggregation functions, through the normalisation stage. The most direct way to do it is to investigate the "relative importance" (marginal contribution) that each dimension is given in each of the three aforementioned models of Social Inclusion. Indeed, the aggregation function is kept constant and it is characterised by *equal* weighting to the normalised dimensions. Nevertheless, since no such things as normalized-longevity or normalized-unemployment rates exist in reality, it is particularly useful to focus on how observed-attributes contribute to the overall measure of social inclusion, and what characterize the relationship between raw-variables within the aggregation framework.

The marginal contribution of each *j-th* raw-indicator to the overall synthetic measure can be determined by computing the partial derivative as in (11.3). Given that the selected aggregation model has $w_j = 0.25$, the magnitude of the marginal contribution is entirely determined by the steepness of the adopted normalisation function. Indeed, for a generic linear aggregation model L, and for any normalisation function v, it holds that:

$$\frac{\partial L(v(\mathbf{x}))}{\partial x_j} = 0.25 \cdot v_j'(x_j) \tag{11.12}$$

The derivative v' represents the link implicitly imposed, when normalising data, between the original variable x and its counterpart $\nu(x)$. The partial derivatives in (11.13), (11.14), and (11.15) illustrate such link for the data-driven min-max (MM), the Z-score (Z), and the survey-driven min-max (sM), respectively:

$$\frac{\partial \nu^i_{\mathrm{MM}\pm}(x^i)}{\partial x^i} = \pm \frac{25}{\max\limits_{t \in T} \max\limits_{i} (x^t_+) - \min\limits_{t \in T} \min\limits_{i} (x^t_+)} \tag{11.13}$$

$$\frac{\partial \nu^i_Z(x^i)}{\partial x^i} = 0.25 \cdot \frac{1}{\sigma(x)} \tag{11.14}$$

$$\frac{\partial \nu^i_{\mathrm{sM}\pm}(x^i)}{\partial x^i} = \pm \frac{25}{[\text{thresholds}' \text{ range}]} \text{ if } x^i \in [\text{thresholds}' \text{ range}] \qquad (11.15)$$
$$= 0 \; if \; x^i \leq \text{lowest threshold} \vee x^i \geq \text{highest threshold}$$

We can immediately notice that, in all of the cases, the effect of a one-unit increment in x on the transformed variable $\nu(x)$ is constant. This is due to the transformation functions being linear, and the benchmarks b*max*, b*min*, $\sigma(x)$, being fixed (they are either extracted from the data, or elicited from experts).[15] In particular, the higher the range or the standard deviation of a raw variable, the lower its unitary marginal contribution to the normalised one. Hence, unless all the attributes have very similar distributions and comparable units of measurement (which would make the normalisation itself of secondary importance), all the normalisations are performing a preliminary, and unequal, weighting of the original variables, regardless of the choice of the aggregation function.

A partial exception must be highlighted for the survey-driven min-max (11.15), for which the usual *non-satiation* hypothesis (more of a "good" is always preferred to less) is maintained in a weaker form. Indeed, more of a "good" is *non-ill favoured* with respect to less of it, after a certain performance is reached (and, conversely, more of a "bad" is *non-preferred* to less of it). As a rough realisation of the diminishing sensitivity hypothesis, the effects of a change in variables' score on the social utility is zero after the thresholds are crossed. From a policy-implication point of view, this suggests to focus on those dimensions whose performances lie farther away from the "desirability" level.

To help further clarifying the aforementioned observations, Fig. 11.1 reports a graphical visualisation of the two versions of the min-max transformation implemented on the selected data. The heterogeneity of the functions' steepness both *within* and *between* normalisation frameworks reflects the differences in each variable's thresholds' range. In particular, the steep of the survey-based functions for unemployment and school-dropouts is higher with respect to the data-driven version because of the shorter min-max range imposed by the experts. The opposite is true for life expectancy, which has a steeper normalisation under the data-driven strategy. To make some examples, a long-term unemployment rate of 3% is normalised to 100 under the expert thresholds, to around 80 under the data-driven thresholds. A life expectancy of 80 years old results in a transformed value around 70 under the expert-function, whereas it is around 40 in the data-driven normalisation.

We can now compute, by the means of partial derivations, the marginal contribution of each indicator with respect to the three synthetic measures LD, LS, LZ, that is, the impact that a unitary change in the original attribute has on the overall measure of Social Inclusion. Table 11.6 illustrates the results.

[15]The linearity hypothesis of the min-max can be relaxed by imposing a non-linear shape (convex, concave or s-shaped, see, e.g., Martinetti and von Jacobi (2012) and Meyer and Ponthière (2011)). Such alternatives were tested and do not in any way alter the implications of this paper.

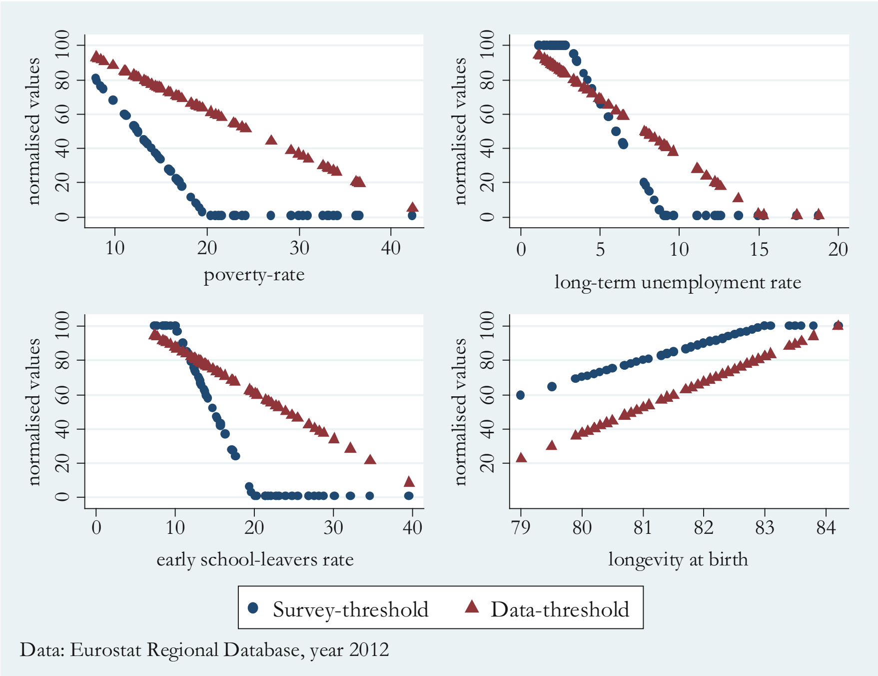


Fig. 11.1 Min-max normalisation: survey-driven vs data-driven benchmarks

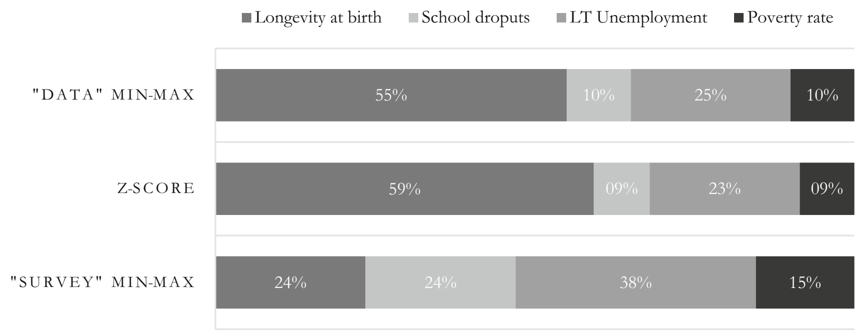
Table 11.6 Dimensions' relative weights in linear data-driven model under different normalisations

	Longevity	Early school leavers	Long-term unemployment	Poverty-rate
Data-driven min-max	3.7	−0.7	−1.6	−0.6
Survey-based min-max[a]	2.5	−2.5	−4	−1.65
Z-score	0.21	−0.031	−0.083	−0.03

Note: Data from the Eurostat Database 2015 (2003–2013)
[a]Derivatives refer to unitary increments of the original variable from a starting value within the interval [max(x_j)–min(x_j)] as detailed in Table 11.5. For all the values outside the boundaries, a unitary variation would produce no zero change in the Index

Since these marginal contribution coefficients cannot be easily compared across normalisation-methods, we normalised them row-wise, so that their sum is always 100. This allows us to interpret the results in terms of "relative weights", i.e., how much weight (in percentage) is given to a specific raw variable. Such normalised weights are shown in Fig. 11.2.

Although the weights were set as equal for each normalised attribute, those related to the actual indicators are highly un-balanced, regardless of the transformation adopted.



	"Survey" Min-max	Z-score	"Data" Min-max
Longevity at birth	24%	59%	55%
School droputs	24%	09%	10%
LT Unemployment	38%	23%	25%
Poverty rate	15%	09%	10%

Fig. 11.2 Relative weights in different normalisation strategies

In particular, the longevity dimension is assigned a predominant role in the aggregation (a relative weight higher than 55%) under the data-driven min-max. Indeed, this is the variable for which the data-driven min-max exhibits the highest slope (Fig. 11.1). Much lower effects derive from a decrease of one unit in long-term unemployment, and an even lower one from reductions in school-dropouts and poverty-rate.

The Z-score standardisation returns relative weights which are fairly similar to those of the data-driven min-max: the relative relevance of the variables included in the index is inversely related to their variability. Given that longevity has by far the lowest observed standard deviation, each unitary increase in its value leads to a much higher increase of the normalised indicator (and therefore of the aggregate index) than for the remaining variables.

Such trade-offs change significantly when the expert-based min-max is adopted. The dimensions' weights appear slightly more homogeneous: longevity and school-dropouts have equal relative relevance (24%), while poverty-rate accounts for 15.4% of the weight and the unemployment indicator being the one with the highest marginal effect on the aggregate measure (37.6%).

Moreover, using (11.4), the marginal rates of substitution between any pairs of indicators x_j, x_k can be computed for each of the three aggregation models. Results are reported in the following tables (Tables 11.7, 11.8, and 11.9).

As expected, the marginal rates of substitution mirror the heterogeneity in the relative weights, and yet convey a more pragmatic evidence on the relevance of the hidden, and partially unintended, trade-offs lying behind the apparently simple and neutral aggregation framework adopted. Just to make an example, in the data-driven min-max model, one additional year of longevity increases the synthetic

Table 11.7 Marginal rates of substitution in the linear model with data-driven min-max

		II. can be compensated by the following change in...			
		Longevity years	% points early school leavers	% points l.t. unemployment	% points poverty rate
I. An increase of...	1 year in longevity		5.3	2.3	6.2
	1% point in early school leavers	0.2		−0.4	−1.2
	1% point in l.t. unemployment	0.4	−2.3		−2.7
	1% point in poverty rate	0.2	−0.8	−0.4	

Table 11.8 Marginal rates of substitution in the linear model with Z-score standardisation

		II. can be compensated by the following change in...			
		longevity years	% points early school leavers	% points l.t. unemployment	% points poverty rate
I. An increase of...	1 year in longevity		6.77	2.53	7
	1% point in early school leavers	0.15		−0.37	−1.03
	1% point in l.t. unemployment	0.39	−2.67		−2.76
	1% point in poverty rate	0.14	−0.96	−0.36	

Table 11.9 Marginal rates of substitution in the linear model with the survey-driven min-max

		II. can be compensated by the following change in...			
		longevity years	% points early school leavers	% points l.t. unemployment	% points poverty rate
I. An increase of...	**1 year in longevity**		1	0.62	1.5
	1% point in early school leavers	1		−0.62	−1.5
	1% point in l.t. unemployment	1.6	−1.6		−2.4
	1% point in poverty rate	0.66	−0.66	−0.41	

Note: these MRS are only valid when the variables take values within the boundaries [max(xj)-min (xj)]. Elseways, the MRS would be either zero, or infinite, or indeterminate, according to whether the numerator, the denominator, or both, in the ratio $v_k'(x_k)/v_j'(x_j)$ in (11.4) is zero index of Social Inclusion as it would a reduction of at least 5.3% points in school dropouts, around 2.5% points in long-term unemployment, and around 6.2 points in poverty rate. In the survey-based min-max model, the marginal rates of substitution for a unitary increase in life-expectancy are much lower: namely, 1% point change in early-school leavers, 0.62 points of long-term unemployment, 1.5 points of poverty rate.

Discussion: Positive Vs Normative Analysis

The aforementioned heterogeneous relative weights and trade-offs, both *within* and *between* the aggregation models, arise because of the (differences in the) adopted normalisation strategies, and will strongly influence the resulting indices, as section "Results" will show. Moreover, this happens in the context of aggregation frameworks granting "equal weights" to their components. As already stated, such label can be partially misleading, since the equal weighting pertains just to the normalised attributes. Indeed, to the extent to which rescaling is a requirement for composite measures, the actual aggregation concerns the transformed variables, in place of the observed performances, and yet there is an unavoidable and intrinsic difference between the interpretation of original and normalized performances. The transformed unit of measurement (e.g., between zero and one, if the min-max rescaling is adopted) can be interpreted as a sort of degree of fulfilment of some criterion. Whether this criterion should be purely statistical (e.g., being far or close to the observed minimum or maximum achievements), or whether it should encompass some informed value judgements related to the topic at hand (as in the expert elicitation or in the adoption of policy benchmarks), relies on the researcher's choice.

What we would like to stress at this point is not whether such trade-offs are acceptable, but rather that (i) they are inevitable, (ii) the ground on which they are justified can differ greatly, depending on the normalisation strategy adopted, and, therefore, (iii) the resulting aggregate indices should be interpreted accordingly.

When a data-driven approach is selected, debating on the acceptability of the underlying marginal rates of substitution is a marginal issue: the justification, and therefore the interpretation, of such coefficients is inherently statistical. It follows that the interpretation of the resulting composite indices should be of the same nature, that is, statistical, which constitutes a strong and solid ground for a *positive* analysis of a composite phenomenon. That being said, the absence of value-judgements in the construction of a data-driven index does not neutralise the hidden trade-offs shown in the previous tables. It is still true that, with a data-driven min-max and with a data-selection as described in section "social inclusion, definition and sample selection", life-expectancy carries a weight which is more than twice what is assigned to the remaining three variables, and that one additional year of longevity is implicitly made equivalent to, e.g., 6.77% points of school dropouts.

It follows that, by construction, such relative weights and marginal rates of substitution are sensitive to the choice of the data-sample and to distribution of the original variables. Indeed, and especially for the data-driven min-max, the presence of outliers in the data would stretch the range over which the normalisation is performed, therefore altering the original variable's marginal contribution to the overall index (as noted, we prudently excluded the extremely high values for the Spanish autonomous cities of Ceuta and Medilla in computing the data-driven benchmarks). As an additional warning, such transformations – again, especially the min-max – are not stable when data for new years or new regions become available, which could sensibly affect the distribution of data (Lefebvre et al. 2010). Similarly, a shift in the territorial dimension of the analysis (e.g., from a national to a regional or provincial level) will cause similar changes, since the provincial data are likely to exhibit higher variability than the regional ones.

Under the strategy of expert-elicitation of the transformation parameters, standard properties as strong non-satiation and continuity of the normalisation function are not guaranteed (indeed, in our example, the min-max becomes weakly monotonic when the elicited constraints are binding for some observed variable). Moreover, the elicitation method suffers from the arbitrariness embedded in any survey exercise, e.g., choice of the population, bias in the framing of questions, and is by definition sensitive to such choices.

Dimensions' trade-offs reflect the preferences of an actual group of experts, and are therefore independent from the selection of data and from the territorial dimension of the analysis. Such trade-offs are to be interpreted under an economic perspective, in terms of social desirability. Therefore, the resulting aggregate measure constitutes a tool for *normative* analysis.

Results

We start by commenting the results at country-level, for practical purposes. We, then, will switch to the analysis of social-inclusion values for specific regions. Summary results at the country level for the three Social-Inclusion indices are reported in Table 11.10 (regional indices are weighted by population size). Full results at regional level are available in the Appendix (Table 11.16).

Given the methodological purposes of this contribution, the results should be read both *within* and *between* columns, with respect to *levels* (especially for the two min-max based models, which share the same metric), *variability* (standard deviation) and *ranking*.[16] It is also important to recall that the conceptual premises of the two strategies are substantially different, and influence the interpretation of the

[16] Atkinson et al. (2004) argue that the ultimate concern of the policy-maker should be casted on performance levels, since rankings might conceal the actual distances between territorial units, thus leading the reader to misleading conclusions.

Table 11.10 Social Inclusion measure and coefficients of variation, baseline model with data-driven and survey-driven normalisations

	Survey-driven min-max		Data-driven min-max		Z-score	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Belgium (BE)	67.3	30.9	70.1	18.3	100.23	0.91
Germany (DE)	73.8	10.7	74.8	7.2	100.47	0.36
Denmark (DK)	79.2	1.8	74.4	1.8	100.43	0.92
Spain (ES)	30.3	12	52.1	15.3	99.39	0.76
Italy (IT)	54.5	23	66.2	17.4	100.08	0.86
Standard dev.	*23.5*	–	*15.7*			

Note: Country averages by with population weights

results. The data-driven models result in indices which do not embed any *normative* perspective: both levels and distances between regions and countries stem from the characteristics of the data. E.g., when the min-max transformation is adopted, the closer the Inclusion index gets to 0, the more we can conclude that a territorial unit is exhibiting the *lowest* performance (among the observed performance in the sample) in each of the four dimensions. In the survey-driven model, the extreme levels of 0 and 100 have a *normative* interpretation, since they directly reflect desirable/undesirable welfare thresholds. The closer the Inclusion index gets to 0, the more a region's dashboard is characterised by an *undesirable* welfare level in each of the four dimensions.

Let us start with the similarities emerging from the models. Germany and Denmark stand out as the top-two countries for inclusion levels, regardless of the adopted normalisation strategy. Nevertheless, their ranking is reversed when switching from the data-driven approaches (data-driven min-max or Z-score), where Germany has a small advantage, to the survey-driven strategy, where the index for Denmark is considerably higher. Belgium is third-placed under every specification. Yet, its relative Inclusion-loss with respect to both Germany and Denmark is much lower under the two data-driven strategies. Similarly, average values for Italy and Spain are, respectively, fourth and fifth placed. Nevertheless, the spread between them and the remaining countries is extremely higher when the survey-based normalisation is implemented.

As the coefficients for standard deviation suggest, country averages conceal a substantial degree of heterogeneity between regions. Spain, Italy and Belgium, in particular, exhibit high levels of variability, with their standard deviation being close to 50% of national averages in the survey-driven model. Furthermore, such coefficients suggest that the choice of the normalisation strategy strongly affects the extent of such heterogeneity. This leads us to conclude that looking at national averages cannot provide a fully informative tool to evaluate the phenomenon at study. The graphical representations in Fig. 11.3 allow us to visualise the regional-dimension of the indices and to draw further valuable insights on both levels and variability between and within countries and models. Figure 11.4 summarizes the rankings obtained for each region under each normalisation strategy. Territories are sorted by their ranking in the linear data-driven model ("X" marker). Furthermore, the Z-score ranking is characterized by "triangle" symbols, while the survey-driven

Fig. 11.3 baseline models of regional Social Inclusion with different normalisations, by country, year 2012

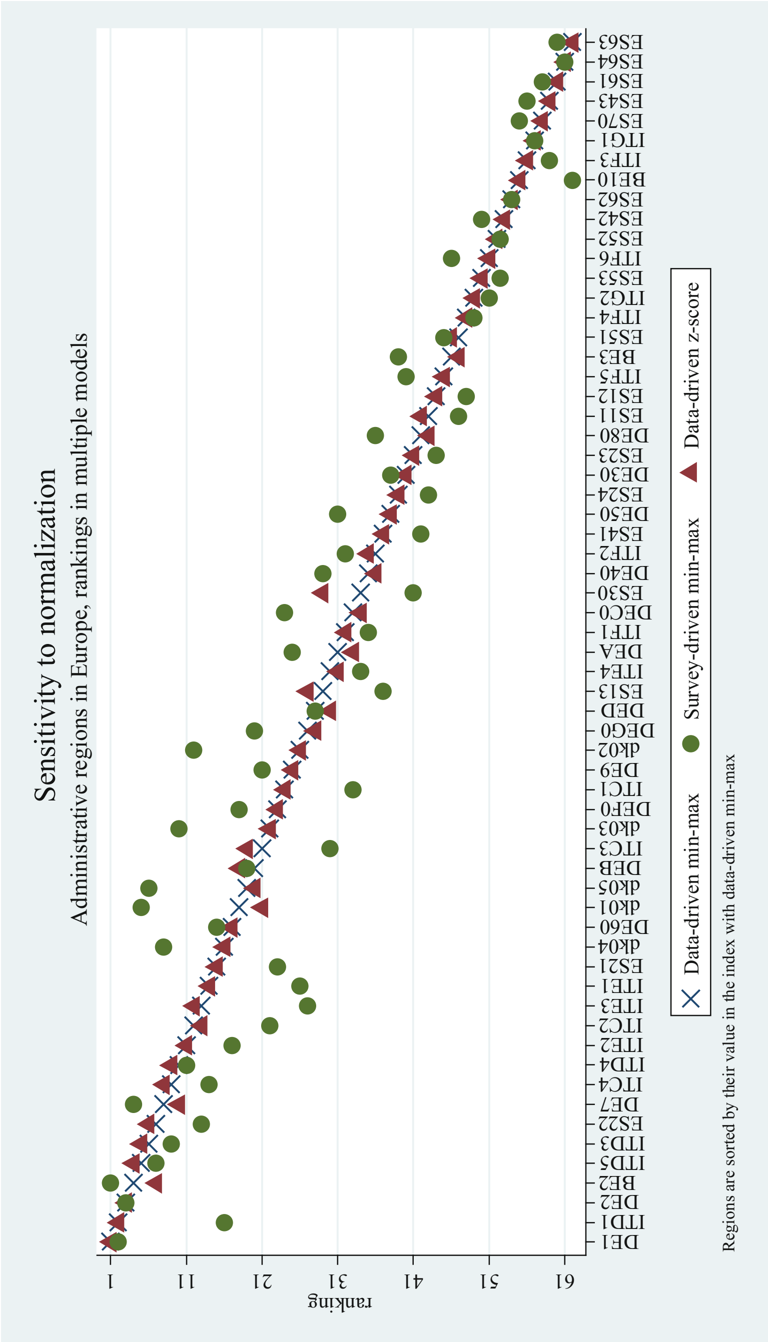


Fig. 11.4 Rankings' sensitivity

ranking has a "circle" mark. The labels on the X-axis report the NUTS code for the administrative regions, whose first two letters identify the country (e.g., "ES01" refers to a Spanish region). The correspondence between labels and region names is illustrated in Appendix A (Table 11.13).

A comparison of the first and the second graph in Fig. 11.3 illustrates how, switching from the data-driven to the survey-based model: (i) Spanish and Italian regions suffer from a substantial drop in Social Inclusion, particularly with respect to Germany and Denmark; (ii) Danish territories are relatively better off; (iii) all countries, except for Spain, exhibit a higher degree of regional variation.

The case of Spain is of particular interest for our methodological approach. Indeed, even though at country level the ranking is consistent, the Spanish picture conveyed by the survey-driven model is more troublesome: only three regions appear to have levels of Inclusion comparable with Germany and Italy (Navarra, Pais Vasco, Cantabria), with the others lying far below on the metric scale, placed (almost exclusively) in the bottom-20 of the ranking. This denotes a notable welfare loss with respect to the remaining territories, as well as a distinctive skewedness of the distribution. Both of these features are completely absent from the data-driven results, where Spain exhibits a high degree of heterogeneity, with some territories performing relatively well, others relatively badly, and a group lying in between. In particular, under these frameworks, a substantial group of ten regions (therefore, a majority share) appears to be roughly in line with the German and Italian distribution of the index (Navarra, Pais Vasco, Cantabria, Madrid, Castilla y León, Aragón, La Rioja, Galicia, Asturias, Cataluña). Evidently, the survey-driven model conveys a much stronger early-warning message than the data-driven ones, which would potentially lead to very different policy implications.

We can exploit Fig. 11.4 to spot the normalisation-induced variation in Spanish territorial rankings: e.g., focusing on the X and the circle marks, it is relatively easy to identify the regions which are "penalized" by the data-driven min-max normalisation with respect to the survey-driven one (i.e., the "x" lies below the "circle"), as well as the opposite (when the "x" lies above the "circle"). There are eight Spanish regions among the last eleven under all of the three specifications, and their ranking is basically normalisation-invariant (Valencia, Castilla-la Mancha, Murcia, Canarias, Extremadura, Andalucía, Melilla, Ceuta).[17] Conversely, almost all of the remaining regions' rankings are strongly affected by normalisation choices, and numerous substantial rank-reversals occur, as we will show by making few examples. The Comunidad Foral de Navarra (ES22) ranks 7th under the data-driven min-max and 6th under the Z-score. However, it drops to 13th place under the survey-based model. Meanwhile, the German Land Hessen, ranked just below Navarra in the data-driven models (8th and 10th), reaches a much higher rank in the survey-driven model (4th). An even more dramatic set of rank-reversals occurs when comparing Navarra with all of the five Danish regions: although the latter

[17] Still, we know from Fig. 11.3 that their distance from the more virtuous territories is lower under the data-driven approaches

group appears at far distance in data-driven models (the best ranked Danish region being Midtjylland at 16th), these territories largely outrank Navarra in the survey-based framework. Incidentally, the exact same reversal affects also the region of País Vasco (ES21). Spanish regions with lower positions in the table are similarly outranked by other European territories when the analysis is performed through survey-driven normalisation: e.g., Cantabria, Madrid, Castilla y León, overtaken by, among others, Nordrhein-Westfalen, Molise, Bremen.

As well as it is for Spain, the relative performance of Italian regions changes distinctly when comparing methods (Fig. 11.3). Under both the data-min-max and the Z-score, the Italian regional picture appears extremely heterogeneous, with some territories performing worse, and the remaining others being equal, or better off, with respect to regions in Germany and Denmark. Moreover, the levels and distribution of Italian Social Inclusion appear very similar to those in Spain, when excluding the two "outlier" autonomous cities of Ceuta and Medilla. Switching to the survey-based normalisation makes the picture change radically, and in a different fashion with respect to Spain. Italy appears as a highly severed country, at least in term of Social Inclusion. The number of low-performing territories is sensibly higher, and a much wider distance separate these territories with the remaining relatively-good performing ones, which are – in turn – worse placed with respect to both Germany and Denmark than they were under the data-driven strategies. Nevertheless, the overall Italian picture looks substantially better than the Spanish one, again, contrarily to what could be inferred in the data-driven models. Figure 11.4 helps is in identifying the numerous rank-reversals affecting Italian, German and Denmark regions, which happen almost exclusively to the benefit of the latter two countries. A notable example is the northern region of Trentino-Alto Adige (ITD1), ranked 2nd in both the data-driven approaches, which drops to 17th under the survey-based approach, being overtaken by German, Danish and Italian regions, as well as by Madrid. Similarly, the central regions Toscana (ITE1) and Marche (ITE3) as well as the industrial north-west region Piemonte (ITC1), who all rank mid-high in the data-driven models, lose numerous positions once the normalisation switches.

Belgium is affected by the aforementioned "Italian effect", yet to a lower extent. Although the ranking of its three regions appears extremely robust, both the within-country heterogeneity and the between-countries relative rankings are substantially modified by the methodological choices. The Inclusion level for Flanders, as well as its ranking, is constantly very high, and is increased when the survey-normalisation is adopted. The French-speaking region Wallonia, however, has a significantly lower performance under the survey-based model, which results in a wider gap with the Flemish region. Finally, the Bruxelles region is already among the worst-ranked in the data-driven models, and yet it drops to bottom of the ranking under the survey-model. As a consequence, the emerging Belgian picture conveys much more heterogeneity in the survey-based framework.

A similar pattern can be found for Germany: although its overall levels of Social Inclusion are higher than Belgium and Italy, the survey-driven approach returns a degree of within-country variability that is absent in the data-driven models.

The distribution of the Social Inclusion index for Denmark is characterised by a low degree of dispersion, regardless of the adopted normalisation. As for levels and

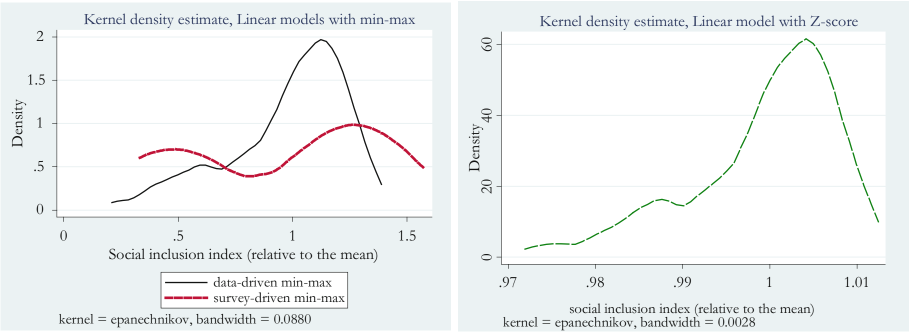


Fig. 11.5 Distributions of inclusion indices

rankings in the data-driven models, Danish regions score relatively high values. Yet, their best performing territory is overcome by one Belgian, nine Italian, three German, and two Spanish regions. Conversely, in the survey-driven framework their ranking significantly improves: e.g., the worst performing Danish region is overcome by just three regions from Germany, three from Italy, and one from Belgium.

Moreover, we show that changing the normalisation strategy has large implications on the distribution of the measure of Social Inclusion in our sample. To this purpose, in Fig. 11.5 we plot a kernel density estimation for each of the three indices, after having normalised them with respect to their respective mean (using population weights). The left-hand side graph refers to the indices based on min-max normalisation (data-driven in black, survey-driven in red-dashed), whereas the right-hand side figure refers to the Z-score model. Both the data-driven kernels have a density peak around the average value, denoting a rather uni-modal distribution. Conversely, the survey-driven dashed line denotes a much higher heterogeneity, and a distinct bi-modal distribution, confirming the key point already conveyed in this Section: important differences in Social Inclusion exist across Europe between areas with very low and very high levels (relative to the average). Again, such polarisation in the distribution may lead to very distinct conclusions in terms of policy implication to assuage and prevent the Exclusion phenomenon.

Finally, in order to test whether the three specifications convey similar rankings, we perform a Kendall's tau[18] test between the rankings from the data-driven and the survey-driven models. Results are reported in Table 11.11. Coefficients' magnitude indicate that rakings' correspondence is far from perfect, yet we can always reject (significance at 99%) the null-hypothesis of no correlation between the models' rankings.

[18]The Kendall-τ test is a non-parametric method that allows to measure the degree of correspondence between two rankings. In particular, the Kendall-τ b allows for the possibility of ties in the rankings. Command in STATA: ktau. A resulting test-value of zero would indicate that no correlation exist between the two rankings, while a value of 1 would indicate perfect correlation. Conversely, negative values (down to a minimum of -1) would indicate that rankings are inverted.

Table 11.11 Kendall's τ correlation coefficients

	Survey-driven min-max	Z-score
Data-driven min-max	0.79	0.97
Survey-driven min-max	–	0.77

Weights and Interpretation of Results

It is useful to briefly recap how different normalisation strategies can lead to substantially different scenarios of Social Inclusion in Europe. The key factors to consider are: (i) the relative-advantage that each country has in a specific dimension; (ii) as well as the *within-country* heterogeneity between the performance of the four selected raw-variables (Table 11.2). Indeed, Spain and Italy present, on average, a strongly unbalanced dashboard: the longevity dimension is particularly high, relative to the other countries, while education and socio-economic variables show much worse relative-values. Since the two normalisation strategies (data-driven vs survey-driven) give opposite relative weights to these dimensions (a prevalent weight to longevity in the data-driven models; a more equal set of weights in the survey-driven one, with some prevalence to unemployment) this leads to the aforementioned difference in-levels, with many Spanish and Italian regions falling to lower-ranked positions. Countries like Denmark and Germany are less affected by the change, given that their dashboard of indicators is uniform. Moreover, heterogeneity *within* countries is very different: Spain, Italy and Belgium show coefficients of standard deviations much higher than Denmark and Germany in each of the included variables. This means that the Inclusion-mix can differ greatly between regions, within the same country. This kind of heterogeneity is somehow softened in the data-driven models, given that a single dimension gets such a large relative weight. In the survey-driven model, conversely, because of the higher weight given to the remaining dimensions other than longevity, such heterogeneity is enhanced.

This, in turn, explains the numerous rank reversals discussed in this Section. The Italian region Trentino-Alto Adige, for instance, scores slightly better than Belgium's Flanders under both the data-driven models. Although both regions admittedly exhibit virtuous performance in each of the four raw variables, the Italian region is relatively better off in longevity-at-birth (83.6 vs 81.4 years) while Flanders has an edge in unemployment (1.5 vs 1.6), education (8.7 vs 15.8) and poverty (9.8 vs 12.5). In both data-driven models, Trentino's losses in three over four dimensions are more than compensated by the gain in life-expectancy, so that the two territories end up having very similar ranking and levels (83.8 for Trentino, 82.4 for Flanders, in the data-driven min-max). Conversely, the survey-based normalisation implies larger weights to the three dimensions where Trentino trails Flanders, thus enhancing the score and the ranking of the Flemish region, while depressing Trentino's ones (88 for Flanders, 73 for Trentino).

Another example comes from the comparison of Spanish País Vasco and Danish Syddanmark. In the data-driven min-max model they rank 15th (76.9) and 22nd (73.7), respectively. Under the survey-based framework the Danish region climbs to 10th place (77.6), while País Vasco drops to 23rd (69.4). Again, the reason for this shift relies on the heterogeneity in the two regions' dashboards. País Vasco performs very well in life-expectancy (83.1, well in school dropouts (11.5%, close to the 9% of Syddanmark), yet it loses ground in long-term unemployment (6.4%, while Syddanmark's rate is just 2.4%).

Conclusion

The unavoidable subjectivity of composite measures of Well-being are cause of controversies in this field of economic analysis. In this chapter, we argued that the lack of transparency on methodological choices can turn out to be more troublesome than subjectivity per se: specifically, we focused on the choices of the normalisation function. In the context of building a synthetic Index of Social Inclusion for 63 European regions in 2012, we showed the consequences of adopting different normalisation methods while keeping constant the (linear) aggregation model, with equal weights allocated to the normalised dimensions.

To the extent to which rescaling is a requirement for composite measures, the actual aggregation involves the transformed variables, rather than the observed performances. There is an unavoidable and intrinsic difference between the interpretation of original and normalised performances, and yet social researchers are ultimately interested in the contribution of the *original* variables to the aggregate index. The rescaled unit of measurement (e.g., between zero and one-hundred) can be interpreted as a sort of degree of fulfilment of some criterion. Whether this criterion should be purely statistical (e.g., being far or close to the observed minimum or maximum observed achievements), or whether it should encompass some informed value judgements related to the topic at hand (as in the expert elicitation or in the adoption of policy benchmarks), relies on the researcher's choice.

In the former case, the *agnostic* choice of "letting the data talk", standard properties as strong non-satiation and continuity of the normalisation function are guaranteed, yet dimensions' trade-offs are hard to interpret in economic terms or from a social desirability perspective. Therefore, the resulting aggregate measure would be characterized as a tool for mainly "positive" analysis.

In the latter case, conversely, the normalisation function may become weakly monotonic, when the elicited constraints are binding for some observed variable. Moreover, the elicitation method suffers from the arbitrariness embedded in any survey exercise (choice of the population, biases in the framing of questions). Moreover, dimensions' trade-offs reflect the preferences of an actual group of experts, thus are independent from the data-selection, and allow to characterize the final measure with a "normative" connotation.

The main result of our analysis is that neither method is neutral, and that the implicit trade-offs embedded in each model should be made transparent to the reader. Indeed, data-driven normalisation softens the differences between territories, since it puts a consistent weight on the longevity variable which follows dynamics that are only partially related to socio-economic contingencies.

Conversely, the survey-driven normalisation emphasizes the worse performance of Italy and Spain in long-term unemployment, thus producing a bi-modal picture of Inclusion in Europe, with a cluster of region scoring very high and another scoring very low levels of Inclusion. As a result, numerous rank-reversals occur between regions when switching the normalisation methods.

Acknowledgement The author wishes to thank Michele Bernasconi, Giovanni Bertin, Eric Bonsang, Agar Brugiavini, Stefano Campostrini, Roberto Casarin, Koen Decancq, Silvio Giove, Filomena Maggino, Sergio Perelman, Pierre Pestieau, Dino Rizzi, Maurizio Zenezini, for their valuable comments to previous versions of this paper. The paper benefited as well from comments by participants to seminars at Ca' Foscari University of Venice, University of Trieste, as well as to the conference "Complexity in Society" at University of Padova and "Data Science and Social Research" at University of Napoli Federico II. The author acknowledges the financial support of Fondazione Ca' Foscari.

Appendices

Appendix A: Data-Description

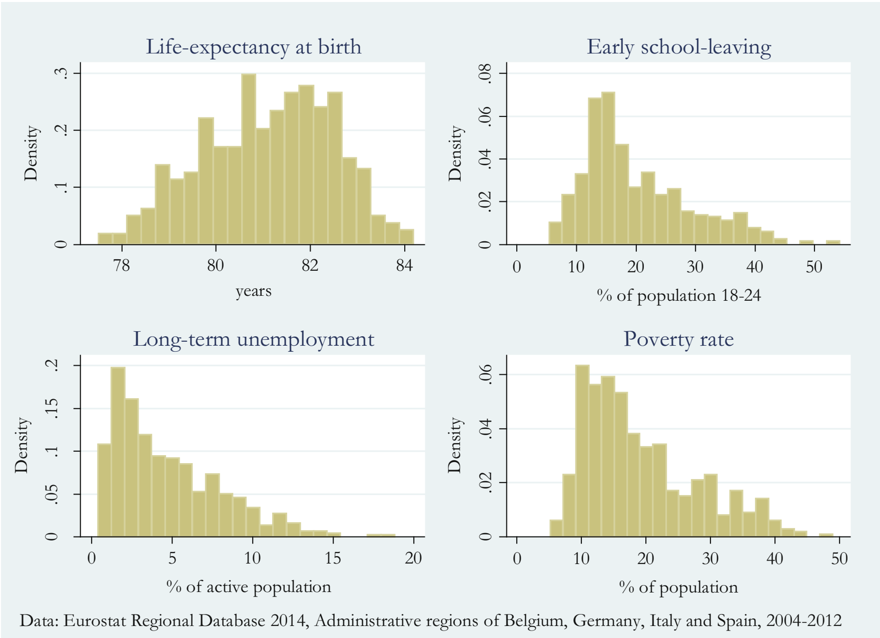


Fig. 11.6 Variables' distribution for data-driven normalisation for Belgium, Denmark, Germany, Italy and Spain from 2004 to 2012

Table 11.12 Correlation between the four variables of Social Inclusion

	Longevity	Early school leaving	Long-term unemployment	At-risk-of-poverty rate
Longevity	1			
Early school leaving	0.28	1		
Long-term unemployment	−0.02	0.14	1	
At-risk-of-poverty rate	0.1	0.53	0.54	1

Data: Eurostat Database 2015, year 2012

Table 11.13 Administrative regions, names and NUTS codes

Country	NUTS	Administrative region
BE	BE10	Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest
BE	BE2	Vlaams Gewest
BE	BE3	Région wallonne
DE	DE1	Baden-Württemberg
DE	DE2	Bayern
DE	DE30	Berlin
DE	DE40	Brandenburg
DE	DE50	Bremen
DE	DE60	Hamburg
DE	DE7	Hessen
DE	DE80	Mecklenburg-Vorpommern
DE	DE9	Niedersachsen
DE	DEA	Nordrhein-Westfalen
DE	DEB	Rheinland-Pfalz
DE	DEC0	Saarland
DE	DED	Sachsen
DE	DEE0	Sachsen-Anhalt
DE	DEF0	Schleswig-Holstein
DE	DEG0	Thüringen
DK	dk01	Hovedstaden
DK	dk02	Sjælland
DK	dk03	Syddanmark
DK	dk04	Midtjylland
DK	dk05	Nordjylland
ES	ES11	Galicia
ES	ES12	Principado de Asturias
ES	ES13	Cantabria
ES	ES21	País Vasco
ES	ES22	Comunidad Foral de Navarra
ES	ES23	La Rioja
ES	ES24	Aragón

(continued)

Table 11.13 (continued)

Country	NUTS	Administrative region
ES	ES30	Comunidad de Madrid
ES	ES41	Castilla y León
ES	ES42	Castilla-la Mancha
ES	ES43	Extremadura
ES	ES51	Cataluña
ES	ES52	Comunidad Valenciana
ES	ES53	Illes Balears
ES	ES61	Andalucía
ES	ES62	Región de Murcia
ES	ES63	Ciudad Autónoma de Ceuta (ES)
ES	ES64	Ciudad Autónoma de Melilla (ES)
ES	ES70	Canarias (ES)
IT	ITC1	Piemonte
IT	ITC2	VDA
IT	ITC3	Liguria
IT	ITC4	Lombardia
IT	ITD1	TAA
IT	ITD3	Veneto
IT	ITD4	FVG
IT	ITD5	ER
IT	ITE1	Toscana
IT	ITE2	Umbria
IT	ITE3	Marche
IT	ITE4	Lazio
IT	ITF1	Abruzzo
IT	ITF2	Molise
IT	ITF3	Campania
IT	ITF4	Puglia
IT	ITF5	Basilicata
IT	ITF6	Calabria
IT	ITG1	Sicilia
IT	ITG2	Sardegna

Appendix B: Description of the Survey

The Survey was structured as follows:

- An introductory section discussed the topics, the purpose and the contents of the survey.
- Respondents were asked to select the variables (amongst the four described in section "Social Inclusion, definition and sample selection") for which they would be willing to perform an evaluation.

- A randomization led the respondent to a page devoted to one of the selected variables. All pages were homogeneously designed with a consistent phrasing.
- The EUROSTAT definition of the variable at hand was offered, and descriptive statistics were shown through a bar graph, for 25 European countries (years 2000 and 2012).
- The main task of the survey was then detailed. Respondents should identify, according to their own opinion, two main thresholds for the variable at hand: a negative threshold, defined as a "value of the selected variable which conveys a certainly undesirable and problematic condition", and a positive defined as "a value conveying a certainly desirable and virtuous condition". The threshold had to be chosen by dragging a slider (using the mouse left-click) on a predetermined discrete interval of values,[19] and releasing it to identify the preferred value (see Fig. 11.7 for a snapshot of the negative-threshold choice for life expectancy).
- An example involving a mock variable "X" explained how to deal with the Qualtrics layout in order to identify the thresholds.
- After choosing the positive and the negative thresholds, a confirmation was required by clicking on "confirm and proceed" button, which would lead the respondent to the next variable-specific page, or to the last section of the survey (if no variables were left).
- The last section of the survey included questions on respondents' age, gender and affiliation (either Economics or Management).

As an example, let us consider the survey-page devoted to the life-expectancy-at-birth indicator. First, a definition of life expectancy was provided. Then, data for 25 European countries (years 2000 and 2012) were shown. At this point, respondents are faced with the summary of what they will be asked to do, i.e., identifying both a favourable and a harmful threshold for life-expectancy-at-birth, according to their own opinion. The harmful threshold is defined as a "level of longevity which represents a certainly negative and undesirable condition". The favour threshold is defined as a "level of longevity which represents a certainly positive and desirable condition". Before reaching the actual question, a full example was provided with a generic variable "X". Respondents had, then, to determine the harmful threshold by dragging a slider on an interval of values (with the left mouse-click), and dropping it at the point that corresponded to their view of a certainly undesirable level of longevity. Figure 11.7 illustrates the choice that respondents were facing for the harmful threshold of longevity. The choice was not entirely free, since we constrained respondents to select a level of life expectancy within a predetermined interval ranging from 60 to 90 years old, in order to avoid extremely implausible choices (like 0 years old). Similar steps characterized the choice of the favourable

[19]Fixed intervals of values were imposed in order to avoid extreme and implausible choices (like 0 years old of longevity as "harmful" threshold). The predetermined intervals were: [90–60 years] for longevity; [0%, 50%] for early-school-leaving; [0%, 50%] for long-term unemployment; and [0%, 50%] for poverty-rate. No respondents chose one of the non-zero extremes as their preferred threshold.

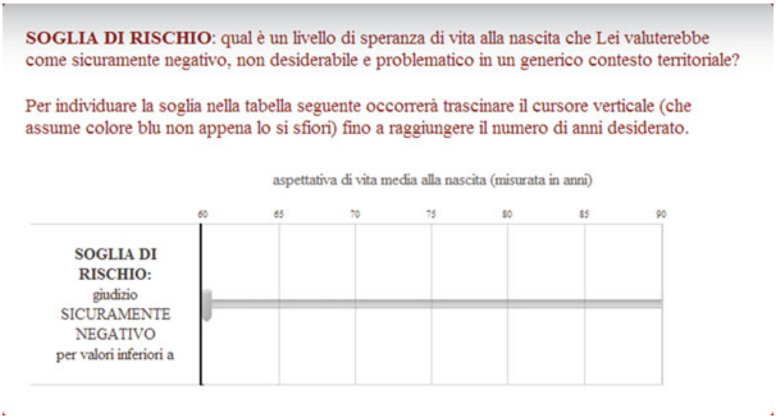


Fig. 11.7 Choice of the negative threshold for life expectancy

Table 11.14 Descriptive statistics on the survey' sample

	Economics	Management	Overall
Respondents	59	28	88
Age:			
Less than 40	36.1%	13%	28.6%
Between 40 and 49	29.9%	43.5%	34.3%
Between 50 and 59	23.4%	26.1%	24.3%
60 or more	10.6%	17.4%	12.6%
Female respondents	40.4%	34.8%	38.6%

threshold, where respondents had to select their answer in the same interval between 60 and 90 years old. A cautionary disclaimer was emphasized at this point, stressing the fact that the favourable threshold should, by construction, be higher than- or equal to- the harmful threshold previously selected.[20]

Out of 149 invitations, we received 88 responses. 59 were faculty members of the Department of Economics, 29 from the Department of Management. The following table provide brief descriptive statistics on our sample.

Median responses and interquartile range are reported in Table 11.15,[21] while Fig. 11.8 illustrates the histograms for the responses' distribution. The blue thick-dashed lines represent the answers for the favourable thresholds.

[20]The disclaimer aimed at avoiding inconsistent choices, e.g., a respondent who would choose, say, 81 years old as a harmful threshold, and subsequently choose 80 as a favourable threshold. No such patterns occurred.

[21]We chose the median response as a measure of central tendency to summarize a representative answer, as often done in the literature (e.g., Hoskins and Mascherini (2009)) because of its lower sensitivity to outliers, especially when the sample size is small.

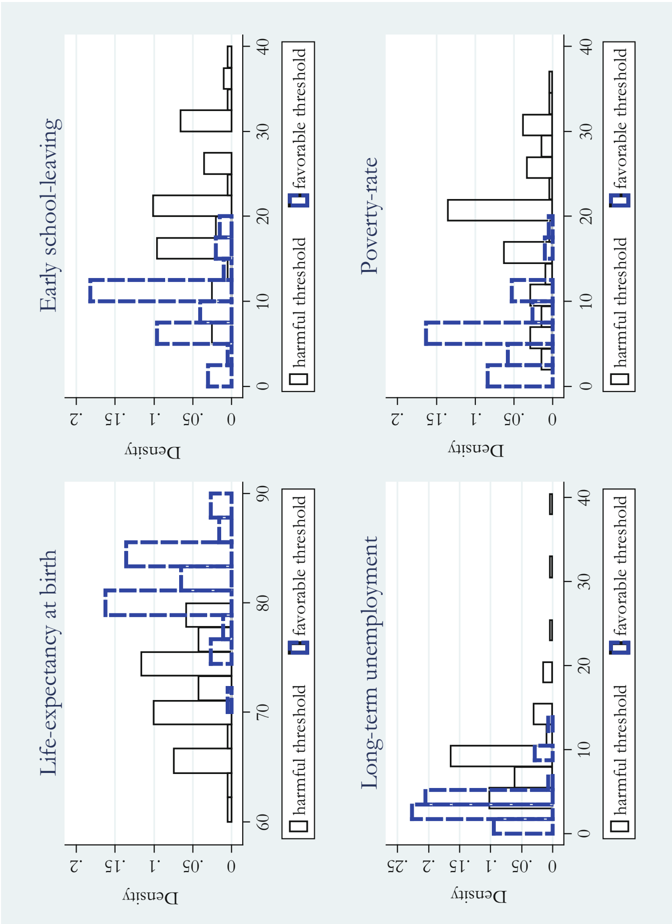


Fig. 11.8 Distribution of survey responses

Table 11.15 Survey-elicited benchmarks

	Median elicited minimum (25–75p)	Median elicited maximum (25–75p)
Longevity	73 years (70–75)	83 years (80–85)
Early school leaving	10% (5–10)	20% (15–25)
Long-term unemployment	3% (2–4)	9% (5.25–10)
At-risk-of-poverty rate	5% (3–7)	20% (17–21.5)

Note: on-line survey (QUALTRICS software), 88 responses from professors in Economics or Management at the Ca' Foscari University of Venice

Appendix C: Results for Administrative Regions

The following coefficients are obtained by implementing the LD model (11.9)

Table 11.16 Aggregate measure of Social Inclusion, baseline model with data-driven normalisation

Rank-country-region			Survey min-max	Rank-country-region			Data min-max	Rank-country-region			Data Z-score
1	BE	Vlaams Gewest	88.0	1	DE	B. Württemberg	85.2	1	DE	B. Württemberg	101.0
2	DE	Baden Württemberg	87.8	2	IT	TAA	83.9	2	IT	TAA	101.0
3	DE	Bayern	85.7	3	DE	Bayern	83.0	3	DE	Bayern	100.9
4	DE	Hessen	82.6	4	BE	Vlaams Gewest	82.4	4	IT	ER	100.9
5	DK	Hovedstaden	80.8	5	IT	ER	82.2	5	IT	Veneto	100.9
6	DK	Nordjylland	80.4	6	IT	Veneto	82.2	6	ES	Navarra	100.9
7	IT	ER	79.9	7	ES	Navarra	81.9	7	BE	Vlaams Gewest	100.8
8	DK	Midtjylland	79.9	8	DE	Hessen	79.5	8	IT	Lombardia	100.7
9	IT	Veneto	79.5	9	IT	Lombardia	79.3	9	IT	FVG	100.7
10	DK	Syddanmark	77.6	10	IT	FVG	79.1	10	DE	Hessen	100.7
11	IT	FVG	76.8	11	IT	Umbria	78.4	11	IT	Umbria	100.7
12	DK	Sjælland	76.5	12	IT	VDA	78.0	12	IT	Marche	100.7
13	ES	Navarra	74.1	13	IT	Marche	78.0	13	IT	VDA	100.7
14	IT	Lombardia	73.3	14	IT	Toscana	77.1	14	IT	Toscana	100.6
15	DE	Hamburg	73.2	15	ES	País Vasco	76.9	15	ES	País Vasco	100.6
16	IT	TAA	73.0	16	DK	Midtjylland	76.7	16	DK	Midtjylland	100.6
17	IT	Umbria	72.8	17	DE	Hamburg	75.0	17	DE	Hamburg	100.5
18	DE	Schleswig Holst.	72.5	18	DK	Hovedstaden	74.6	18	DE	Rheinland-Pfalz	100.5
19	DE	Rheinland-Pfalz	72.0	19	DK	Nordjylland	74.6	19	IT	Liguria	100.5
20	DE	Thüringen	71.1	20	DE	Rheinland-Pfalz	74.5	20	DK	Nordjylland	100.4
21	DE	Niedersachsen	70.7	21	IT	Liguria	73.9	21	DK	Hovedstaden	100.4
22	IT	VDA	69.7	22	DK	Syddanmark	73.7	22	DK	Syddanmark	100.4
23	ES	País Vasco	69.4	23	DE	Schleswig Holst.	73.4	23	DE	Schleswig Holst.	100.4
24	DE	Saarland	67.0	24	IT	Piemonte	72.7	24	IT	Piemonte	100.4
25	DE	Nordr. Westfalen	66.4	25	DE	Niedersachsen	72.4	25	DE	Niedersachsen	100.4
26	IT	Toscana	66.4	26	DK	Sjælland	71.4	26	DK	Sjælland	100.3
27	IT	Marche	65.8	27	DE	Thüringen	70.9	27	ES	Cantabria	100.3

(continued)

Table 11.16 (continued)

Rank-country-region			Survey min-max	Rank-country-region			Data min-max	Rank-country-region			Data Z-score
28	DE	Sachsen	65.3	28	DE	Sachsen	70.3	28	DE	Thüringen	100.3
29	DE	Brandenburg	65.1	29	ES	Cantabria	70.2	29	ES	Madrid	100.2
30	IT	Liguria	64.1	30	IT	Lazio	69.7	30	DE	Sachsen	100.2
31	DE	Bremen	60.5	31	DE	Nordr. Westfalen	69.5	31	IT	Lazio	100.2
32	IT	Molise	60.1	32	IT	Abruzzo	69.3	32	IT	Abruzzo	100.2
33	IT	Piemonte	59.8	33	DE	Saarland	69.1	33	DE	Nordr. Westfalen	100.2
34	IT	Lazio	57.4	34	ES	Madrid	69.0	34	DE	Saarland	100.2
35	IT	Abruzzo	55.7	35	DE	Brandenburg	67.4	35	IT	Molise	100.1
36	DE	Meckl. Vorp.	54.6	36	IT	Molise	67.1	36	DE	Brandenburg	100.1
37	ES	Cantabria	52.6	37	ES	Cast. y Leon	65.4	37	ES	Cast. y Leon	100.1
38	DE	Berlin	51.3	38	DE	Bremen	65.0	38	DE	Bremen	100.0
39	BE	Wallonia	46.4	39	ES	Aragón	63.9	39	ES	Aragón	100.0
40	IT	Basilicata	43.4	40	DE	Berlin	63.8	40	DE	Berlin	99.9
41	ES	Madrid	36.1	41	ES	La Rioja	62.2	41	ES	La Rioja	99.9
42	ES	Cast. y Leon	32.3	42	DE	Meckl. Vorp.	60.2	42	ES	Galicia	99.8
43	ES	Aragón	31.1	43	ES	Galicia	59.6	43	DE	Meckl. Vorp.	99.7
44	ES	La Rioja	31.0	44	ES	Asturias	58.7	44	ES	Asturias	99.7
45	ES	Cataluña	30.3	45	IT	Basilicata	57.8	45	IT	Basilicata	99.7
46	IT	Calabria	29.8	46	BE	Wallonia	57.7	46	ES	Cataluña	99.6
47	ES	Galicia	29.6	47	ES	Cataluña	56.9	47	BE	Wallonia	99.6
48	ES	Asturias	28.9	48	IT	Puglia	55.5	48	IT	Puglia	99.6
49	IT	Puglia	26.3	49	IT	Sardegna	55.2	49	IT	Sardegna	99.5
50	ES	Castilla	25.8	50	ES	Balears	48.8	50	ES	Balears	99.2
51	IT	Sardegna	24.0	51	IT	Calabria	48.7	51	IT	Calabria	99.2
52	ES	Valenciana	23.5	52	ES	Valenciana	45.4	52	ES	Valenciana	99.1
53	ES	Illes Balears	23.5	53	ES	Castilla	43.1	53	ES	Castilla	99.0
54	ES	Murcia	23.0	54	ES	Murcia	41.7	54	ES	Murcia	98.9
55	ES	Canarias	22.8	55	BE	Bruxelles	41.2	55	BE	Bruxelles	98.8
56	ES	Extremadura	22.5	56	IT	Campania	36.5	56	IT	Campania	98.6
57	IT	Sicilia	21.8	57	IT	Sicilia	34.7	57	IT	Sicilia	98.5
58	ES	Andalucía	21.0	58	ES	Canarias	33.1	58	ES	Canarias	98.4
59	IT	Campania	20.3	59	ES	Extremadura	31.9	59	ES	Extremadura	98.4
60	ES	Ceuta	20.0	60	ES	Andalucía	31.8	60	ES	Andalucía	98.4
61	ES	Melilla	19.5	61	ES	Melilla	21.8	61	ES	Melilla	97.7
62	BE	Bruxelles	18.5	62	ES	Ceuta	19.5	62	ES	Ceuta	97.5

References

Anand, S., & Sen, A. (1994). *Human development index: Methodology and measurement. Human Development Report Office (HDRO)*. New York: United Nations Development Programme (UNDP).

Atkinson, A. B., Cantillon, B., Marlier, E., & Nolan, B. (2002). *Social indicators: The EU and social inclusion: The EU and social inclusion*. Oxford: Oxford University Press.

Atkinson, A. B., Marlier, E., & Nolan, B. (2004). Indicators and targets for social inclusion in the European Union. *JCMS: Journal of Common Market Studies, 42*(1), 47–75.

Bertin, G., Carrino, L., & Giove, S. (2016). The Italian regional well-being in a multi-expert non-additive perspective. *Social Indicators Research, 3*, 1–37.

Blackorby, C., & Donaldson, D. (1982). Ratio-scale and translation-scale full interpersonal comparability without domain restrictions: Admissible social-evaluation functions. *International Economic Review, 23*, 249–268.

Boarini, R., & D'Ercole, M. M. (2013). Going beyond GDP: An OECD perspective*. *Fiscal Studies, 34*(3), 289–314. doi:10.1111/j.1475-5890.2013.12007.x.

Bourguignon, F., & Chakravarty, S. R. (2003). The measurement of multidimensional poverty. *The Journal of Economic Inequality, 1*(1), 25–49.

Cherchye, L., Knox Lovell, C. A., Moesen, W., & Van Puyenbroeck, T. (2007). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review, 51*(3), 749–779. doi:http://dx.doi.org/10.1016/j.euroecorev.2006.03.011.

Chowdhury, S., & Squire, L. (2006). Setting weights for aggregate indices: An application to the commitment to development index and human development index. *Journal of Development Studies, 42*(5), 761–771.

Decancq, K., & Lugo, M. A. (2008). *Setting weights in multidimensional indices of well-being*. Oxford: Oxford Poverty & Human Development Initiative.

Decancq, K., & Lugo, M. A. (2009). Measuring inequality of well-being with a correlation-sensitive multidimensional Gini index. *University of Oxford Department of Economics Discussion Paper Series, 459*.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews, 32*(1), 7–34.

European Commission. (2009). *Portfolio of indicators for the monitoring of the European strategy for social protection and social inclusion Bruxelles*.

European Commission (2010). Combating poverty and social exclusion. A statistical portrait of the European Unio. European Union.

European Communities Commission. (1992). *Towards a Europe of solidarity. Intensifying the fight against social exclusion, fostering integration. Communication from the commission. COM(92) 542 final, 23 December 1992*. Bruxelles: Commission of the European Communities.

European Council. (2000). *Lisbon European Council 23–24 March 2000*. Presidency conclusions.

European Council. (2001). *Laeken European Council. 14–15 December 2001*. Presidency conclusions and annexes.

Giovannini, E., Nardo, M., Saisana, M., Saltelli, A., Tarantola, A., & Hoffman, A. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD.

Hoskins, B., & Mascherini, M. (2009). Measuring active citizenship through the development of a composite indicator. *Social Indicators Research, 90*(3), 459–488. doi:10.1007/s11205-008-9271-2.

Kasparian, J., & Rolland, A. (2012). OECD's 'better life index': Can any country be well ranked? *Journal of Applied Statistics, 39*(10), 2223–2230.

Kim, Y., Kee, Y., & Lee, S. (2015). An analysis of the relative importance of components in measuring community wellbeing: Perspectives of citizens, public officials, and experts. *Social Indicators Research, 121*(2), 345–369. doi:10.1007/s11205-014-0652-4.

Klugman, J., Rodríguez, F., & Choi, H.-J. (2011). The HDI 2010: New controversies, old critiques. *The Journal of Economic Inequality, 9*(2), 249–288.

Lefebvre, M., Coelli, T., & Pestieau, P. (2010). On the convergence of social protection performance in the European Union. *CESifo Economic Studies, 56*(2), 300–322.

Maggino, F., & Nuvolati, G. (2012). *Quality of life in Italy: Research and Reflections. 48*: Springer.

Martinetti, E. C., & von Jacobi, N. (2012). Light and shade of multidimensional indexes: How methodological choices impact on empirical results. *Quality of life in Italy: Research and Reflections. Social Indicators Research Series, 48*.

Mascherini, M., & Hoskins, B. (2008). Retrieving expert opinion on weights for the Active Citizenship Composite Indicator. *European Commission–Institute for the protection and security of the citizen–EUR JRC46303 EN*.

Mazziotta, M., & Pareto, A. (2015). On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social indicators research*, 1–21. doi:10.1007/s11205-015-0998-2.

Meyer, P., & Ponthière, G. (2011). Eliciting preferences on multiattribute societies with a Choquet integral. *Computational Economics, 37*(2), 133–168.

Murias, P., Novello, S., & Martinez, F. (2012). The regions of economic well-being in Italy and Spain. *Regional Studies, 46*(6), 793–816.

Nguefack-Tsague, G., Klasen, S., & Zucchini, W. (2011). On weighting the components of the human development index: A statistical justification. *Journal of human development and capabilities, 12*(2), 183–202.

OECD, & European Commission, J. R. C. E. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publishing.

Pestieau, P. (2009). Assessing the performance of the public sector. *Annals of public and cooperative economics, 80*(1), 133–161.

Pinar, M., Cruciani, C., Giove, S., & Sostero, M. (2014). Constructing the FEEM sustainability index: A Choquet integral application. *Ecological Indicators, 39*, 189–202.

Ravallion, M. (2011). On multidimensional indices of poverty. *The Journal of Economic Inequality, 9*(2), 235–248.

Ravallion, M. (2012a). Mashup indices of development. *World Bank Research Observer, 27*(1), 1–32.

Ravallion, M. (2012b). Troubling tradeoffs in the human development index. *Journal of Development Economics, 99*(2), 201–209.

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 168*(2), 307–323.

Sen, A., & Anand, S. (1997). *Concepts of human development and poverty: A multidimensional perspective*. *Poverty and human development: Human development papers*. NewYork: UNDP.

Silva, R., & Ferreira-Lopes, A. (2013). A regional development index for Portugal. *Social Indicators Research*, 1–31.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2010). *Report by the commission on the measurement of economic performance and social progress*.

Giovanni Bertin, Ludovico Carrino, Silvio Giove, The Italian Regional Well-Being in a Multi-expert Non-additive Perspective. Social Indicators Research

Chapter 12
Evaluation of Life Satisfaction in Italy: Proposal of a Synthetic Measure Based on Poset Theory

Giovanna Boccuzzo and Giulio Caperna

Introduction

The definition of tools for evaluation and comparison of different alternatives in a multicriteria framework has been hotly debated for 20 years and will continue to be an important topic. In this context, the computation of a synthetic measure is one of the most accepted alternatives because the output is very simple and easily understood by inexperienced audiences (Nardo et al. 2005). The research on synthetic indicators aims to measure complex and unobservable concepts, such as quality of life or sustainability, with a unique measure by starting from a group of variables that describes the phenomenon from several observable perspectives. The simplification of complex and unobservable concepts is the main reason this methodology has gained a vivid interest in applied sciences such as chemometrics, sociology, psychometrics, and economics.

This research proposes a synthetic indicator using a statistical methodology equipped with the necessary mathematical formalization to define synthetic measures out of a group of ordinal and dichotomous variables.

The scope is achieved without using scaling procedures for the elementary variables or other assumptions that transform qualitative variables into quantitative ones.

The fundamental definitions and concepts for a brand new approach come from the theory of Partially Ordered Sets (poset), which is a mathematical approach to define space according to the order relations between observations and not to their position in a Euclidean multidimensional space. Theory of poset contains several tools for the formalization and solution of the problem under analysis (Davey and

G. Boccuzzo • G. Caperna (✉)
Department of Statistical Science, University of Padova, via Cesare Battisti 241, 35121 Padova, Italy
e-mail: giovanna.boccuzzo@unipd.it; caperna@stat.unipd.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_12

Priestley 2002). One of the most common is the definition of the position of an element inside a partially ordered set and is called average rank (or medium height), and it is a simple and easy-to-read synthetic measure representing the information on order relations between the elements of the set. Many approaches for the approximation of average rank have been proposed in the last decade, like Brüggemann and Carlsen (2011) and De Loof et al. (2011).

The purpose of this text is the presentation of a slight modification of the tools, presented by Fattore (2015), which is an innovative approach that allows an intuitive application of poset theory in social sciences (Maggino and Fattore 2011). The innovation of this method is the connection between the poset formalization and the framework of multidimensional decision making for evaluation and identification. Because of its nature, we will often refer to it as the *evaluation* method.

The chapter is divided into four parts. In the first section, an introduction to poset theory and identification methodology is presented, the aim is to point out the most relevant features for this work. In the second part, a synthetic measure is derived starting from Fattore's methodology; in this part, the construction and characteristics of the proposed measure are described. In the third section, we propose observations on the utilization of the evaluation method, with some suggestions for readers who want to use this method for the first time. The last part of this work is dedicated to presenting the analysis of life satisfaction in Italy, with the innovative information given by this proposal.

Partially Ordered Sets – Posets

The theory of partially ordered sets is a mathematical formalization of discrete space. Poset theory has been used in the field of data analysis to compare the elements of small populations; it adapts to both quantitative and qualitative data. The best use of poset in the case of quantitative data produces information which cannot be interpreted in the classic Euclidean sense. As in the case for instance, of the effect of a variable on a response not following any function or the description of this function being impossible. In chemometrics, for example, the aim is the extraction of information from chemical systems by mean of the analysis of data; often in this scientific field it is not possible to assess the effect of chemicals respecting quantitative scales, and sometimes it is not even possible to compare two chemical mixtures.

In modern social statistics, the presence of qualitative data is massive, because it relies on the opinions and judgments of individuals. There is a limited availability of quantitative information, and the methodologies are required to handle this issue. This premise implies the usefulness of poset theory in the field of social statistics, which is the main interest of this work.

Poset theory allows using the order information contained in data, including if information on the distance between units is absent, or when it is reasonably better to avoid using it.

All the definitions about poset theory that are presented in this section can be found in the text: Introduction to Lattices and Order (Davey and Priestley 2002). A

Table 12.1 Incomparability example

	q_1	q_2
x	Low	High
y	Medium	Medium
z	Medium	Low

good introduction to many of the concepts related to partially ordered sets is presented in Chap. 8; we will often refer to the paragraphs of this chapter to avoid repetitions in the book.

What Is a Partially Ordered Set

The elements of a family can be ordered with respect to some criteria, for instance, the order: father $\geq$ mother $\geq$ older sister $\geq$ younger brother could be the order in a common family if we consider the age as a unique criterion. The order is the relation between the elements of the group (*set*) that respects some properties presented in Chap. 8 (see section "Formal definitions").

Whom do you like the most? Grandma or Grandpa? Sometimes it is possible that two elements are neither equal nor ordered; it often happens when more than one criteria are considered simultaneously. The relation ($\leq$) can be defined as a **partial order** if **incomparable** elements exist in the set:

$$\textbf{incomparability:} x \,||\, y \Leftrightarrow \quad \text{not} \quad x \geq y \quad \text{nor} \quad y \geq x, \quad x, y \in P$$

This case happens in the presence of multiple attributes that are conflicting (See section "Formal definitions"). An example can be observed in Table 12.1, where an individual answers two questions (called q_1 and q_2) about the quality of three objects (x , y , z). In this example, it is impossible to define which object is the best because $q_2(x) > q_2(y)$, but $q_1(y) > q_1(x)$. This conflict determines an incomparability and implies the absence of order between the elements x , y. The same situation happens between x and z but not between y and z where $y \geq z$ respect to every attribute.

In the usual notation, two incomparable elements a , b are represented by $a||b$; in the example of Table 12.1 the following relations can be assessed: $x||y$, $y \geq z$ and $x||z$. The mathematical formalization of posets allows one to describe some types of data that are commonly used in social and applied sciences. Every system of ordinal variables can be handled within this theoretical framework, which takes care of all order information and avoids the use of Euclidean space and other quantitative concepts.

Covering Relation

It is important to use the concept of *coverage*, to understand the construction of poset representation. One element covers another if there are no other elements between them, in mathematical language:

Given $x\,, y\,, z$ in the ordered (or partially ordered) set P,

$$x \text{ is } \textit{covered} \text{ by } y \quad (x \lhd y) \quad \text{if} \quad x < y \text{ and } \quad x \leq z < y \Rightarrow x = z.$$

If something is between two elements in a covering relation, it has to be one of these elements. Moreover, if P is finite,

$$x < y \Leftrightarrow \quad \text{a sequence such as } x = x_0 \lhd x_1 \lhd \cdots \lhd x_n = y \text{ always exists.}$$

This formalization assesses that the order relation determines and is determined by a list of covering relations. An object can cover and be covered by many others. As an example, if the objects $x\,, y\,, z$ are evaluated respect to two variables (q_1 and q_2):

	q_1	q_2	$\Rightarrow \quad x \triangleleft y, \quad y \parallel z, \quad x \triangleleft z$
x	Low	Low	
y	High	Medium	
z	Medium	High	

Hasse Diagram

In the final decades of the nineteenth century, mathematicians started to represent partially ordered data with a handwritten technique; they were used to represent every element with a *node* and every covering relation with an *edge*. Some more rules are useful to make the result easy to read in every case, but *directions*, *nodes*, and *edges* are sufficient to draw this graph. This representation is called *Hasse Diagram* (Fig. 12.1), named after the German mathematician *Helmut Hasse*. Hasse did not invent the representation but made extensive use of it (Birkhoff 1948), causing its more widespread usage.

This representation is an oriented graph that allows an observable description of the set. Due to the properties of covering relations, the diagram describes every relation of order between the nodes by drawing only the covering edges.

Down-Set and Up-Set

Two important families of sets are associated with every ordered (or partially ordered) set. They are central to the development of methodologies based on posets. If P is an ordered (or partially ordered) set, and $Q \subseteq P$:

- Q is a **Down-set** if $x \in Q\,, y \in P$, and $y \leq x$ implies $y \in Q$;
- Q is an **Up-set** if $x \in Q\,, y \in P$, and $y \geq x$ implies $y \in Q$.

Consequently, the down-set (up-set) of an element $x \in P$, is the set of all the elements of P that are lower (higher) than the element x itself.

Fig. 12.1 Example of Hasse diagram

	q1	q2
a	2	0
b	0	2
c	3	3
d	4	5
e	7	6
f	5	7

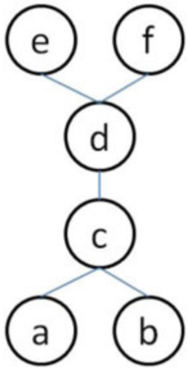
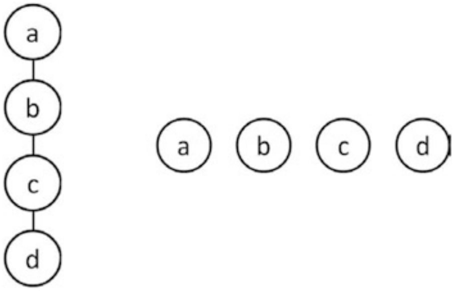


Fig. 12.2 Example of a chain and an antichain



Chains and Antichains

There are two extreme cases in the field of partially ordered sets:

Complete comparability occurs when every element of the set is comparable to all the others, and there are no incomparable pairs. Sets like this are usually called **chains** or **linear orders** because they form a complete order, and the shape of their Hasse diagram is clearly linear (Fig. 12.2).

Formally, a set P is a chain if

$$\forall x, y \in P, \quad either \quad x \leq y \quad or \quad y \leq x.$$

Complete incomparability is the extreme case in which every element is incomparable to all the others, and no comparison leads to an order. In such a situation, the Hasse diagram looks like a horizontal line of isolated nodes (Fig. 12.2) and is called an **antichain**.

A set P is defined as an antichain if

$$\forall x, y \in P, \quad x \leq y \Leftrightarrow x = y.$$

Linear Extensions

When we look for a complete order among the incomparable elements of a poset, there are many possible results, and these are defined *linear extensions* of the poset (see 8.2.2 for a wider introduction). In this chapter, we refer to the set of all the linear extensions of a poset P calling it $\Omega(P)$.

Every linear extension $\omega_i(P)$ can be interpreted as the original poset but enhanced by much information on the comparability of the elements. Naturally, a poset can have more than one linear extension, the dimension of the set $\Omega(P)$ depends on the dimension and structure of comparability in the set.

Information from Linear Extensions

The relevance of linear extensions is well established by a fundamental result from Schröeder (2012), where it is stated that two different finite posets have different sets of linear extensions and that every poset is the intersection of the set of its linear extensions. In other words, a poset coincides with the comparabilities that are common to all its linear extensions.

Given this result, any single linear extension can be considered the atomic level of order information of a poset, like a single observation in a population. Hence:

- Every linear extension describes one of the possible orders of objects.
- The set of all linear extensions of a poset $\Omega(P)$ identifies the partial order structure uniquely.
- The number of linear extensions $|\Omega(P)|$ suggests the complexity of the poset and how many pairs of objects are incomparable.

The Average Rank of a Profile

The rank of an object x on a single linear extension ω_i is called **height** $h_i(x)$ (or simply **rank**); knowing the height of objects among the entire set $\Omega(P)$ allows computing their **average rank** (**AR** or medium height). The AR of an object is the arithmetic mean of its ranks among the linear extensions; this measure summarizes the information about the object's position in the set.

$$\bar{h}(p) = \frac{1}{\mid \Omega(P) \mid} \sum_{\omega_l \in \Omega(P)} r_l(p) \qquad (12.1)$$

The acknowledgment of the position of a unit with respect to all the others is the main aim of this work. With the computation of the **average rank**, an order is obtained, which could be *Weak* or *Complete* if, respectively, there are some elements with the same value of AR ($\bar{h}(\cdot)$) or every element has a different value with respect to all the others. Nevertheless, the computation of the average rank makes every element comparable.

The Chap. 8 shows a simple example of average rank in par "Average rank".

Computational Issue

Observing the entire set of linear extensions is a difficult task. In their work, Brightwell and Winkler (1991) prove that the problem of determining the height of an element x of a given poset is #P-complete (pronounced *sharp p complete*). This assertion refers to the computational complexity of the problem. With a trivial expression, we can affirm that the number of linear extensions and, consequently, the procedure of information-gathering from them has a computational time that cannot be evaluated in a deterministic way. The best results for the approximation of the number of linear extensions started from the results of Dyer et al. (1991), but extremely satisfying results can also be found in the results of De Loof (2009).

Forecasting the time needed to compute the number of linear extensions in a deterministic way is not possible because the number of linear extensions is not directly dependent on the number of elements in the poset; it depends on the structure of the comparable and incomparable elements. In order to clarify the magnitude of variability, compare the following two posets:

- The first consists of 19 comparable elements and one element incomparable with all the others (*isolated* element)
- The second consists of ten elements that are all incomparable.

The number of linear extensions for the first poset is $|\Omega(\mathbf{I})| = 20$ because the isolated element can occupy every position in the chain with length 19. In the second case, every element can take every position, leading to an amount of linear extensions equal to

$$| \Omega(\mathbf{II}) | = 10! = 479001600.$$

In the case of *real* posets, made by more than ten elements, this number can be extremely high, leading to a clear truth:

$$\text{“The set of linear extensions is too big!”}.$$

To handle this information limitation, researchers follow two paths:

Approximation of the average rank: An algorithm is used to find an approximation of the average rank without observing any linear extension (Brüggemann and Carlsen 2011).

Sampling of linear extensions: Only a random subsample of the linear extensions is observed (Fattore 2015).

The insights and applications of the two approaches are very different. In the following, only the sampling approach is examined due to space constraints. For the approximation approach, we suggest Brüggemann et al. (2004), De Loof et al. (2011), and Brüggemann and Carlsen (2011).

A Method to Evaluate Posets

This subsection covers some of the basic concepts of the method proposed by Fattore (2015). It is based on the sampling of linear extensions, and we suggest reading the original work to understand the mathematical formalization completely. Note the different aims between this chapter of the book and the original work proposed in Fattore (2015) and Maggino and Fattore (2011). These cited works contain the conceptual and mathematical formalization for the measurement of social concepts in poset data; specifically, these are focused on a measure of *membership* of the profiles to a dualistic categorization; enriched with a measure of *severity* of this membership. On the other hand, the proposal of this chapter is the utilization of the severity tools given by these works to define synthetic indicators out of ordinal or dichotomous data.

Product Order of Variables

In the original formalization of this method, the poset contains all the possible values that are observable considering the starting variables' structures. In partial order theory, this particular poset is called *product order* because it is created by the interaction of the linear order determined by the single variables. For instance, two dichotomous variables define their linear orders made by two levels $\{0, 1\}$. The product of these two orders produces a poset made by the elements 00, 01, 10, and 11. Then, the elements of the poset need to be described by a sequence of values (e.g. 10), and this sequence is called **profile**, **p**, in the following.

Setting a Threshold

The poset derived from a product order is a mathematical structure without information about the meaning of the variables constituting it. It is possible to define a **threshold** τ which contains external information defined by the researcher to impose a meaning on this structure. The scope of the threshold is to cover the elements that are *deprived*, that is acting like a frontier between the low and the high levels of the poset. The threshold can only be defined externally. Because the framework is multidimensional, it is possible to define a multidimensional threshold, i.e., a list of profiles, which respects two requirements:

1. Every element of the threshold must be considered as completely deprived, or for the sake of generality, members of the lower subset;
2. The threshold must be made by incomparable profiles (i.e. to be an *antichain*, which was introduced in section "Chains and antichains").

Deeper information about the threshold and its meaning is presented in the next section.

Identification Step

The role of the *identification function* $idn(\cdot)$ is to determine the degree of membership to the deprived group for every profile. The *deprivation membership* score is contained in the interval [0, 1].

$$\begin{aligned} idn &: P \mapsto [0,1] \\ &: \mathbf{p} \mapsto idn(\mathbf{p}) \end{aligned}$$

The construction of the identification function is inspired by the principle of *reduction to linear extensions*, previously introduced in section "Information from linear extensions". Each linear extension l is interpreted as a binary classifier where a profile $\mathbf{p}$ is classified as either deprived or not. A small notation improvement is necessary to understand the procedure. In every linear extension l, there is a top element of the threshold τ_l that is better than all the other elements of τ. For every linear extension, a profile $\mathbf{p}$ is defined as deprived in that linear extension, and the value of the identification function is 1 if this profile is lower than or equal to τ_l; otherwise, the value is 0:

$$idn_l(\mathbf{p}) = \begin{cases} 1 & \textit{if} \quad r_l(\mathbf{p}) \leq r_l(\tau_l) \\ 0 & \quad \textit{otherwise} \end{cases}$$

The position of the profile $\mathbf{p}$ and the value of τ_l change in different linear extensions; the aggregation of the results among the observed linear extensions gives the value of the identification function:

$$idn(\mathbf{p}) = \frac{1}{\mid \Omega(P) \mid} \sum_{l \in \Omega(P)} idn_l(\mathbf{p}).$$

At the end of the procedure, every statistical unit will be associated with its identification value:

$idn(\mathbf{p}) = 1$ when the profile is under the threshold for every extension,
$idn(\mathbf{p}) = 0$ if the profile is always not deprived in every linear extension,
$idn(\mathbf{p}) \in (0, 1)$ if the profile is ambiguously defined in the middle.

Following the derived information, it is possible to define three subsets:

$W = \{s \in P : idn(s) = 1\}$, deprived;
$D = \{s \in P : idn(s) = 0\}$, non-deprived;
$A = \{s \in P : 0 < idn(s) < 1\}$, ambiguous.

For deeper explanation and properties, we again suggest referring to the original work Fattore (2015) or section "Evaluation over ordinal multi-indicator systems" of this book.

Severity

The measure of *severity* defines the intensity of deprivation of the deprived or ambiguous elements (subsets D and A) by assigning a numerical value to each profile of $D \cup A$:

$$\begin{aligned} svr &: D \cup A \mapsto \mathrm{R}^{+} \\ &: \mathbf{p} \mapsto svr(\mathbf{p}). \end{aligned}$$

In every linear extension, the measure of interest is the *distance* between a profile and the first element higher than τ_l called $\mathbf{q}_{\mathrm{l}}$. The distance is computed on the rank of the two objects:

$$svr_{\mathrm{l}}(\mathbf{p}) = \begin{cases} r_{\mathrm{l}}(\mathbf{q}_{\mathrm{l}}) - r_{\mathrm{l}}(\mathbf{p}) & \textit{if} \quad \mathbf{p} \lhd \tau_{\mathrm{l}} \\ 0 & \textit{Otherwise}. \end{cases}$$

The severity value is computed only on those linear extensions where the profile is considered deprived.

The *deprivation severity* of a profile is then obtained by aggregating all the results observed on the linear extensions:

$$svr(\mathbf{p}) = \frac{1}{\mid \Omega(P) \mid} \sum_{\mathrm{l} \in \Omega(P)} svr_{\mathrm{l}}(\mathbf{p}).$$

The Wealth function *wea* is complementary to *svr*; it measures the concept of intensity in the opposite direction, the direction of *non-deprivation*. The computation is the same as for *svr*, but it is oriented to the positive side and is evaluated on the subset $(W \cup A)$ (Fattore et al. 2011).

The *R* package devoted to the computation of these functions is called PARSEC and was developed by Fattore and Arcagni (2014).

A Synthetic Measure from Posets

In this chapter, a simple procedure to construct a synthetic indicator is proposed based on the proposals contained in Fattore (2015), Maggino and Fattore (2011), and Fattore and Arcagni (2014), presented in the previous section.

The definition of the synthetic indicator will be the topic of the next section "The height of a profile". In the following section, the central concept of the evaluation method will be described along with its implications. The threshold is the only source of exogenous information, and its role will be presented with examples. The last section contains the construction of the life satisfaction indicator and its use to study the relation of this concept to many socio-economical factors.

The Height of a Profile

The *evaluation* procedure includes the identification function and the severity functions. These functions were previously introduced in sections "Identification step" and "Severity". In the following, we propose to **combine** the information given by the method to obtain a unique measure.

The combination of the results transforms the meaning of the obtained measure: the functions of identification and severity are conceived to describe the membership of a profile to a group (e.g. the group of deprived individuals), and the intensity of this membership. The aim and meaning of the combined values are different because they try to represent the position of the profiles with respect to a complex concept mimed by the poset.

The measure we suggest consists of the combination of the severity indexes produced by Fattore and Arcagni (2014). It is possible to obtain a unique value given by the difference between the absolute *severity* and *wealth* measures:

$$H_\tau(p) = wea(p) - svr(p).$$

Given that

$$\begin{aligned} H_\tau(\mathbf{p}) = wea(\mathbf{p}) - svr(\mathbf{p}) &= \frac{1}{|\Omega(\Pi)|}\sum_{l\in\Omega} wea_l(\mathbf{p}) - \frac{1}{|\Omega(\Pi)|}\sum_{l\in\Omega} svr_l(\mathbf{p}) \\ &= \frac{1}{|\Omega(\Pi)|}\sum_{l\in\Omega}(wea_l(\mathbf{p}) - svr_l(\mathbf{p})), \end{aligned}$$

the index $H_\tau(\mathbf{p})$ coincides with the mean of the differences $\Delta_l(\mathbf{p}) = wea_l(\mathbf{p}) - svr_l(\mathbf{p})$ in the set of all linear extensions.

Meaning

The value $\Delta_l(\mathbf{p})$ represents a sort of height of the profile, *evaluated* on the defined threshold; in every linear extension, the profile is compared to the highest element of the threshold. It gets a positive value if it is higher than the top value of the threshold ($\mathbf{q}_l$) and gets a negative value otherwise.

The indicator $H_\tau(\cdot)$ can be defined as *evaluated* height because it contains the information given by the threshold τ. Two elements of the poset could be equal respect to the average rank and different respect to H_τ, or vice versa; the difference depends on the threshold and specifically on its *length* and *shape*.

In the following section, some thoughts about the meaning of the threshold and the criteria to define it are discussed.

Meaning and Use of the Threshold

The threshold τ proposed in Fattore's *evaluation* method is the cornerstone of the procedure; it inserts exogenous information to impress an evaluative meaning on the poset. This kind of information is essential in a multivariate evaluation framework, where an absolute *best* does not exist, and the relative most acceptable option is the aim (Munda 2008; Arrow and Raynaud 1986).

This section focuses on helping new users to determine the best way to choose a threshold, including whether it is better to look at elementary variables or to choose an expert defined set of profiles.

This section is divided into four parts. In the first, the effect of the *shape* of the threshold is investigated to understand how it can impress a system of values and priorities on the elementary variables. The second part shows the effect of the threshold on the discrimination skill of the identification function and the indicator H_τ. In the third part, the evaluation procedure and the indicator H_τ are compared to a more classic method based on the poset theory. The argument of interest is the added value of a method that contains a user-defined threshold. Finally, in the last part, we propose some tips for defining a threshold pulling together the deductions presented in this section.

The Shape of the Threshold

Although the threshold does not have a unique interpretation, in its original definition it has the purpose of covering every deprived element; all the elements that are lower than the threshold are evaluated as deprived respect to the concept evaluated by the poset. Without the meaning imposed by the threshold, two variables constituted by the same number of levels cannot be distinguished, neither conceptually nor mathematically. To perceive this limit, the case of a poset generated by two ordinal variables made by three levels is sufficient, as shown in Fig. 12.3.

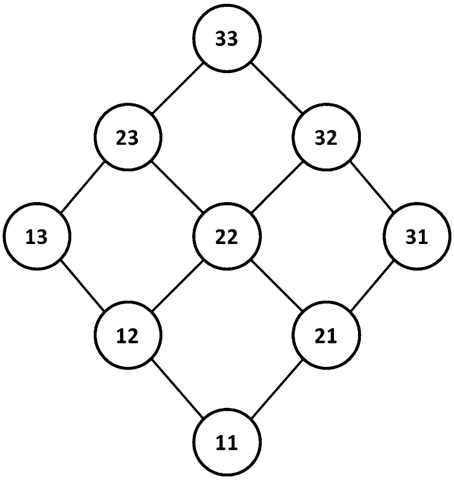
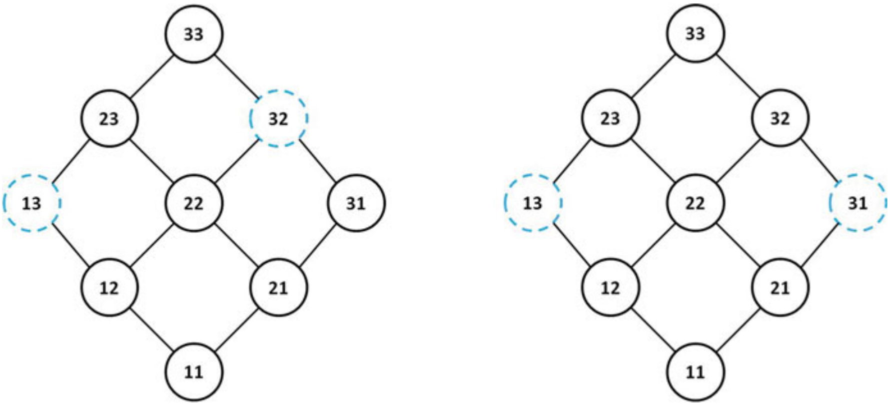


Fig. 12.3 Poset derived by two variables with three levels



Figs. 12.4 and 12.5 Asymmetric and symmetric threshold

Without an externally defined preference system, the constituting variables are completely interchangeable. For instance, these variables could represent the degree of appreciation for an ice cream flavour or the personal self-perceived safety in a residential zone; the two variables are identical if we consider only the structure of the poset. Instead, by adding a threshold to this structure, we impose a system of values and discriminate among the variables.

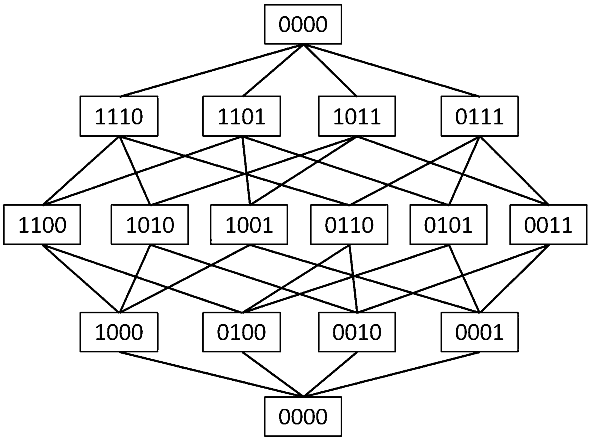
Figures 12.4 and 12.5 look very similar, and yet they impress a very different meaning on the structure. In simplest terms, in Fig. 12.4 the constituting variables are non-exchangeable. In this setting, profiles 32, 22, and 31 are always deprived. The same cannot be confirmed for 23 because it could be higher or lower than 32 depending on the linear extension. In this case, to be identified out of the *certainly deprived* group, a high value of the first variable is not enough; excluding the profile 33, every profile showing the maximum value on the first variables is deprived (lower than or equal to the threshold). On the other hand, with a high value of the second variable, it is possible to be defined as ambiguously deprived (profile 23). Therefore, there is an implicit assertion about the importance of the variables.

On the contrary, in the poset shown in Fig. 12.5, we can perceive a sort of symmetry among the variables; that is, the first variable is equated to the second. Thus, the poset is evaluated in a manner that is more similar to the computation of the average rank.

Due to this property of the threshold, it is possible to impress a system of *preferences*. These preferences are implicit and non-linear because they change intensity in different positions of the poset. So, it is not fair to interpret this preference effect in the same way as composite indicator weights.

Suffice it to imagine a poset made of several dichotomous variables, such as the one shown in Fig. 12.6, to underscore the relevance of such a property. This type of variable has only two levels, by definition, therefore, the levels are not very descriptive; nevertheless, two dichotomous variables could represent a deeply different meaning. For example, in deprivation studies where the ownership of

Fig. 12.6 Poset derived from four dichotomous variables



different goods is considered, owning a refrigerator and being the owner of a house are conceptually different pieces of information. In other words, these variables are different in their meaning and distribution but not in their level structure. In a European country such as Italy, to be deprived of the refrigerator is much more meaningful than not having the property of a house.

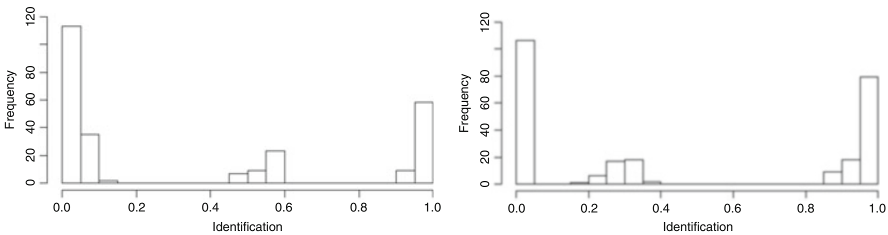
The Length of the Threshold

If the shape can influence the importance of the variables, the length of the threshold can change the level of *fuzziness* of the profiles that are non-comparable to the threshold itself. With the word *length*, we mean the dimension of the threshold, i.e., the cardinality of the set τ.

Indeed, if the threshold is defined by a single element p, only the elements in the downset of p will be certainly deprived, and the inverse result applies to the elements of the upset; all the other profiles will be included in the group of ambiguous elements. On the contrary, if the length of the threshold is big, the amount of uncertainty will decrease in number and intensity.

This concept is particularly intuitive in the case of *identification* function: in order to illustrate this effect, two evaluation functions have been computed on the same poset, made by four variables measured on four levels (256 nodes), using the following thresholds:

- **Single threshold** consisting of a single profile: 2232 that is quite central in the poset;
- **Extended threshold** consisting of profiles: 2232, 1233, 2133, 2223, 1242, 2142, 2241, 3132, 3222, 3231, 1332, 2322, 2331, actually all the profiles that are obtained from the *single threshold* and changing every elementary variable by one unit at a time.



Figs. 12.7 and 12.8 Distribution of the identification function with single and extended threshold

The results of the identification function show the effect of the threshold's length on the classification. In the *single* case (Fig. 12.7), more or less 40 profiles out of 256 (15.6%) are in an intermediate position around value 0.5, and only 67 have a value of deprivation close or equal to one. On the other hand, if the threshold is extended (Fig. 12.8), the number of certainly deprived elements increases to more than 100 units, because it now contains nodes that were ambiguous in the single-threshold case. Moreover, one out of three elements previously defined as non-deprived (Identification = 0), is now less certainly identified, and is distributed in the interval (0.2 – 0.4). There is more deprivation if we judge the whole picture, and the class of *absolute ambiguity*, identifiable with the values around 0.5, disappears.

The identification function is oriented to the evaluation of deprived profiles because of its definition, which takes into account the highest element of the threshold in every linear extension (τ_l).

According to this property, the enlargement of the threshold's dimension can only imply the equality or the increment of the deprivation level of the poset's elements. Using the same notation of the previous section,

$$if\ \tau_s \subset \tau_e \Rightarrow idn^{\tau_s}(p) \leq idn^{\tau_e}(p), \ \ \forall p \in P.$$

The same results apply to the severity function.

Comparison Between the Evaluation Approach and the Average Rank

Within the poset theory, a common tool is the computation of the average rank of an element in the poset. As described in section "The average rank of a profile", this method computes the medium height of observed units, making it possible to order them reciprocally and define their position on an unobservable dimension. This work focuses on the similarities and differences regarding the evaluation method previously presented.

Fattore's evaluation method includes two different measures: the *identification function* and the *severity function*. The first defines the degree of membership of the profiles to one of the opposite sides of the poset (the bottom or the top), the second measures the depth of a profile, describing how much it moves towards one of the two poles. The use of two different measures is motivated by the suggestion of Sen (1976), who proposed that a methodology for measuring multidimensional poverty should be made up of an identification method and an aggregated measure.

Moreover, the user-defined threshold impresses a meaning of centrality on all its elements. In this way, the poset gains some reference points, which do not exist in the definition of the average rank.

Therefore, the definition of the threshold gives a specific meaning to the results of the evaluation function. If a profile is always higher than reference group (resulting in an identification function equal to 0), it can be classified as not-deprived (using the terminology shared by Alkire and Foster (2011) and Fattore (2015)). Similarly, if the profile is always lower than the reference group, the deduction is the opposite, but it is an evaluated classification, not a score. For instance, if the poset is obtained from variables about the satisfaction of individuals, defining a good threshold will discriminate between satisfied and unsatisfied subjects.

This result is not possible in the case of average rank, where the definition of a limit value between two different poles (using quartiles or standard deviation) will result in something with a different meaning, due to the following:

- Using the *identification* function, if the value is equal to 1, there is no linear extension where the profile is greater than every element of τ. In other words, there is at least one element of the threshold that is better than the profile in every linear extension. This information is not accessible in the average rank approach;
- In the *average rank* framework, the classification of a profile as deprived implies that its *average value* is lower than a predetermined value, which means that the profile is sufficiently low in *average*. Due to the properties of the arithmetic mean, this result says nothing about the distribution of ranks among the linear extensions because there could be some linear extensions influencing the average rank more than others. In this sense, the average rank can be defined as less robust than the evaluation method.

The difference between these two approaches is conceptually important, especially in the definition of a classification procedure; in fact, these approaches imply two different definitions of classification. The simple fact of imposing a threshold impresses a meaning on the constituting variables. Instead, in the average rank method, the variables are handled as mathematical entities according to the observed levels and nothing else.

From the *severity* point of view, the additional value is determined by the shape and position of the threshold because it can influence the severity measure observed on the profiles. The point we want to stress is the centrality of the threshold; without that, the average rank and evaluation method are not so different. Indeed, if the complexity of the threshold increases, the difference between the two approaches increases too.

The last relevant difference lies in the profiles that are considered. The average rank method is based only on the observed profiles, and its complexity depends on the relation structure of the set. For this reason, it is used for small samples or performed with the help of approximation procedures (Brüggemann and Carlsen 2011; De Loof et al. 2011). On the other hand, the evaluation procedure is commonly based on all the possible profiles that can be observed starting from the elementary variables (section "Product order of variables"); hence, it does not depend on the numerical dimension of the population. Recent developments (Fattore 2015) considered the use of a smaller set of profiles in the evaluation method, avoiding the evaluation of those profiles that are not observed in the population.

Conclusions About the Threshold

According to our goal for this work, we propose three criteria as guidelines for the choice of the threshold:

- **Meaning of τ:** This can be dependent on the univariate distribution of the elementary variables or be defined externally by experts who evaluate the entire profiles;
- **Shape:** The structure impressed on the poset by the threshold has to be taken into account. An asymmetric threshold could be recommended in the case of variables with different importance;
- **Length:** The number of elements of the threshold influences the results significantly, especially the identification function. As a rule of thumb: if the information needed to define a large threshold is available, then it is better to use it, in order to reduce ambiguities.

Study of Life Satisfaction in Italy

The aim of synthetic indicators is the measurement of complex and non-observable concepts. In this work, the complex concept under study is life satisfaction as a proxy of well-being. It is a multidimensional concept that cannot be objectively defined because it depends on both life and socio-psychological conditions. It is fair to say that the same conditions (assuming the possibility of identical conditions among humans) are evaluated in different ways by different subjects. Such variability depends on culture, society, and psychology interactively, so we cannot measure effective well-being but instead measure its self-reported level from four points of view.

To analyze life satisfaction, we used the method proposed in section "The height of a profile". $H_\tau(p)$ represents the position (height) of the profile p with respect to the threshold τ. After the value of such an indicator is found for every observed unit, it is possible to use this information as a variable.

Data

The data come from a survey carried out by the Italian National Institute of Statistics (Istat). This survey is part of the multipurpose surveys, Aspetti della vita quotidiana, which means *Aspects of Daily Life*. This survey is extremely useful because of its longevity and complexity; it collects information on many life aspects such as work, health, safety perception, social inclusion, society, and much more. In 2012, more than forty thousand individuals were interviewed.

In this work, the focus is oriented towards the concept of life satisfaction and its determinants. Seven variables have been considered to produce a measure, each of them describing the satisfaction on a single aspect of life: Economic Situation, Health Status, Family Relations, Friendship Relations, Free Time, Working Conditions, and Environment. Three of these variables (i.e. Working Condition, Environment, and Friendship Relations) are not taken into account in this work for the following reasons:

- Working Condition is observed only among the individuals who were employed at the time, and who were homemakers. The present study is intended for all people.
- Environment is observed only in relation to recent times in the surveys. This study is meant to be part of a wider, time-crossing, application.
- Friendship Relations is too closely associated with family relations and free time, showing a small amount of original information.

Hence, at the end of the selection procedure, the elementary variables were: Economic Situation (*Economy*), Health Status (*Health*), Family Relations (*Family*), and Free Time (*Time*). All the satisfaction variables were measured on the four-level ordinal scale: *4. A Lot, 3 Enough, 2. A Little, 1. Not at All*. The poset produced by these variables is composed of $4^4 = 256$ nodes. Apart from specific reasons, the most limiting criterion in the selection of starting variables is computational. The number of linear extensions required from the *identification* method in a poset made by 256 elements, according to Karzanov and Khachiyan (1991), is around 6.1×10^{12}. This number increases to 7.8×10^{15}if we add one variable with four levels, more than 1000 times bigger, making the computation too long for any research purpose.

Construction of the Life Satisfaction Indicator

The procedure to determine the values of the life satisfaction indicator has been described in section "The height of a profile". In this application, the established

Table 12.2 Distribution of the satisfaction variables (%)

Variable	4. A lot	3. Enough	2. A Little	1. Not at all	Total
Economic	2.6	41.9	39.3	16.3	100
Health status	18.5	63.2	13.8	4.4	100
Fam. relations	37.3	55.8	5.6	1.4	100
Free time	16.2	51.5	25.7	6.6	100

threshold is made by a unique profile, $\tau = \{2232\}$; then it is imposed to be the first deprived profile and the less severe one (among the deprived). The profile 2232 means level 2, Little satisfied on Economy, Health, and Time, and level 3, Satisfied enough on Family.

This profile has been selected by taking into account the univariate distribution of the elementary variables (Table 12.2). In the threshold, only the Family variable has a value equal to Satisfied Enough. This threshold assesses that the border between satisfaction and dissatisfaction is located in this exact combination of levels. If a profile shows a Family value lower than 3, it implies that profile to be lower or incomparable to the threshold; the incomparability case is obtained only if at least one of the other variables is higher than 2.

This threshold takes into account the higher frequency of the highest levels of satisfaction observed on Family; the same care has not been repeated for Health due to the low frequency of the high level of satisfaction for this variable.

After the computation of the value of the indicator $H_\tau(\cdot)$ for every individual, the data have been scaled with the use of the *MIN-MAX* method. The scaling is important because the aim is the definition of an intuitive indicator, and the use of absolute values (which in this case have a range equal to $(-90; 166)$) is not a straightforward solution for the first description of the phenomenon. In the following, the scaled indicator is called $S(\cdot)$ to represent satisfaction. The life satisfaction is full if the indicator is 1; in the opposite case, it is 0.

The construction of the indicator defines a value for each one of the 256 nodes of the poset, as represented in Fig. 12.9. The highest concentration is in the middle values because the poset is larger in the central part. The graph is almost perfectly symmetric; this feature will be useful in the following because it implies that every asymmetry in the observed distribution is attributable to the distribution of satisfaction in the population.

In the following, the distribution of profiles presented in Fig. 12.9 is called *theoretical* because it does not depend on the amount of individuals observed on every profile.

Results

The observed distribution of the indicator S accounting for the frequency of the 256 nodes in the Italian population in 2012 is represented in Fig. 12.10, followed by

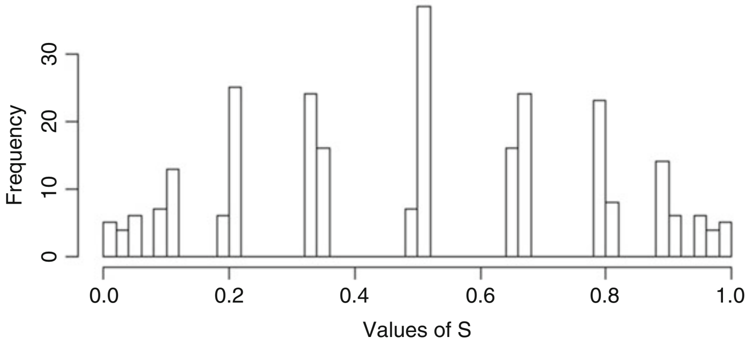


Fig. 12.9 Distribution of the indicator S(p) on the 256 nodes, without profiles' frequencies

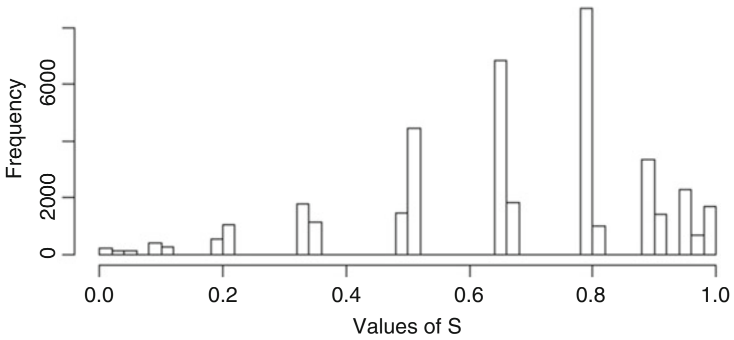


Fig. 12.10 Distribution of the Indicator S(p) taking into account the frequency of the 256 nodes in the Italian population, 2012

Table 12.3 Mean, standard deviation, median and Interquartile range of the S(p), in the theoretical case and for the observed population of Italy, 2012

Distribution	Mean	Std Dev.	Median	IQ range
Theoretical	0.500	0.284	0.502	0.590
Observed	0.671	0.230	0.668	0.306

the basic descriptive statistics of both the theoretical and the observed distribution of profiles (Table 12.3). The picture shows a strongly skewed distribution with a higher number of individuals assessing good levels of life satisfaction. The number of elements with a value lower than 0.4 is extremely low. The individuals who participated in the survey are quite satisfied; their answers show a population that is highly pleased with its level of satisfaction.

This trend is more observable in the representation of deciles (Fig. 12.11): the first 10% of the population has a life satisfaction level in the interval (0; 0.33). The difference compared to the distribution of theoretical profiles is large, which means

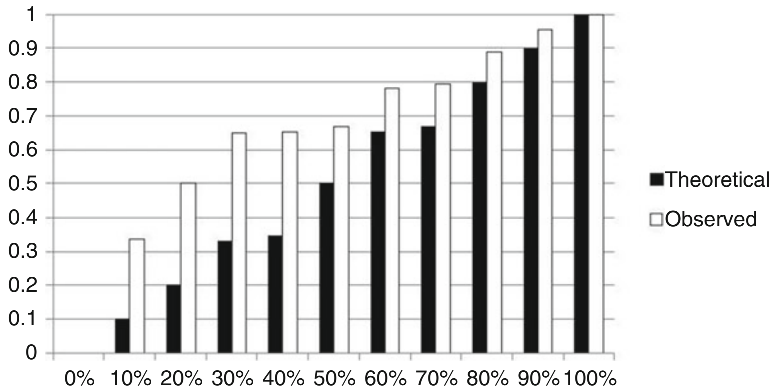


Fig. 12.11 Deciles of the indicator of life satisfaction, theoretical and with the frequencies

that a really small number of people define themselves as severely dissatisfied on every aspect. The difference between theoretical and observed frequencies is reabsorbed between the median and the ninth decile. The highest levels of the index (0.9; 1) are similarly distributed in theoretical and observed data. The number of individuals with a life satisfaction index lower than 0.5 is less than 20% of the population.

Just to have criteria for comparison, between these results and the basic identification approach, we can consider that the profile 2232, which is the only element of the threshold, has a value of $H_{\tau}(2232) = -1$, while the satisfaction index is $S(2232) = 0.35$.

Effect of the Elementary Variables on the Indicator

The relation between the life satisfaction indicator and the single variables constituting it is important to understand the behavior of the indicator; the regression tree method has been used to explore these internal relations. Given the construction procedure of the indicator, some of the observed effects are dependent on the distribution of the elementary variables.

The results of the regression tree are represented in Table 12.4 and Fig. 12.12. The procedure defines eight groups, with an increasing medium value of the indicator $S(\cdot)$; they also show different levels of the elementary variables.

The Family variable is not considered in the first three groups because it is not useful to discriminate among the low levels of the indicator; this elementary variable seems to be more influential at the highest values. Indeed, on the two top levels, the value of Family is 4 (A lot); the most satisfied group with Family lower than 4 has a mean value of the indicator equal to 0.785.

The role of Health is complementary to the Family's role: it shows its importance in the lower levels, where the medium value is 0.324 or less; in those groups, the value of Health is equal or lower than *A Little* (1 or 2). The Free Time variable is the first on the regression tree; it is strictly connected with the mean of life

Table 12.4 Groups of individuals identified by the regression tree

Group	Mean $S(\cdot)$	Time	Health	Economy	Family	
1	0.121	$1 + 2$	$1 + 2$	1	—	
2	0.324	$1 + 2$	$1 + 2$	$2 + 3 + 4$	—	Key:
3	0.486	$1 + 2$	$3 + 4$	$1 + 2$	—	1: Not at all
4	0.582	$3 + 4$	—	$1 + 2$	$1 + 2 + 3$	2: A little
5	0.714	$1 + 2$	$3 + 4$	$3 + 4$	—	3: Enough
6	0.785	$3 + 4$	—	$3 + 4$	$1 + 2 + 3$	4: A lot
7	0.812	$3 + 4$	—	$1 + 2$	4	
8	0.937	$3 + 4$	—	$3 + 4$	4	

satisfaction, it probably suggests how free time could be an overall criterion that is influenced by all the others. The key variable of the severe dissatisfaction is Economy; indeed, in the lowest group, every observation has value Not at All on it.

The seventh group is peculiar in that the elements of this group have a high satisfaction, but the level of Economy is 1 or 2. Therefore, if the family relations are completely satisfying and the free time is at a good value, the level of satisfaction could be very high, despite a low level from the economic point of view.

Life Satisfaction in Society

In the following, the indicator of life satisfaction H_τ (not scaled) is used as a response variable in a quantile regression procedure. In every regression model, the aim is the estimation of the effect of the explanatory variables on life satisfaction. In the quantile regression, the aim is the estimation of a quantile of the response variable conditionally to the effect of explanatory variables (Koenker and Basset 1978; Koenker 2005). This construction makes quantile regression highly recommended when the response variable is not normally distributed, and this is the reason why we are applying it to this data.

The estimated quantile could be the median or the quartiles but also every other quantile. Therefore, this method is suggested in particular when the effect of the explanatory variables is supposed to change along the quantiles.

The method of quantile regression has been applied by defining the indicator H_τ as the response variable and 14 socioeconomic variables as explanatory ones. In Table 12.5, the selected explanatory variables are shown together with the estimation of the parameters at the 50th quantile.

For example, the estimated parameter for men (compared to women) is 2.84.

The original range of the indicator is close to 250, precisely (−90; 166).

In this model, the estimated value of the median of our indicator of life satisfaction can be easily computed; keeping fixed the subject's qualitative variables to the reference levels (the values between brackets), the estimated median of life satisfaction for a median age individual in a family with dimension 2 is

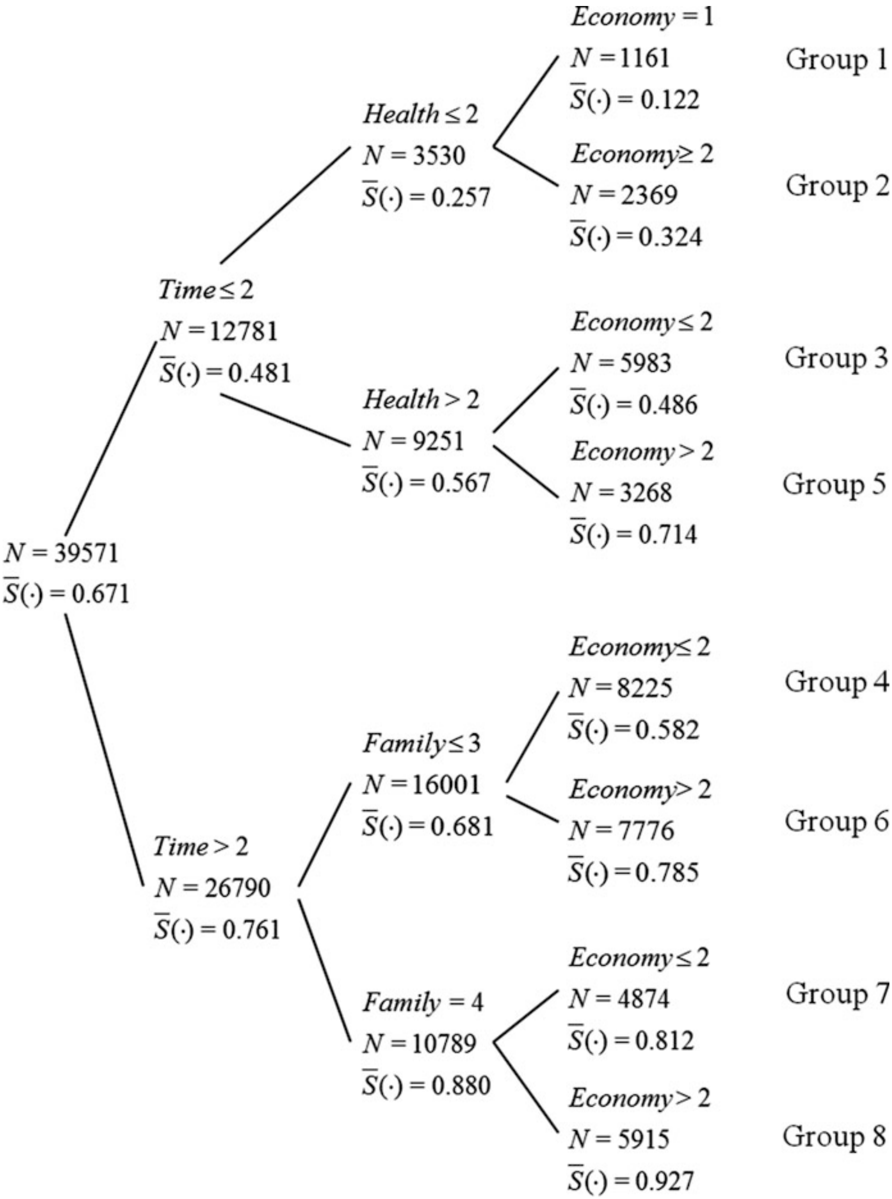


Fig. 12.12 Regression tree of the Italian population with respect to S(.), 2012

$$83.23 - 0.56 \cdot 49 - 1.57 \cdot 2 = 52.95,$$

where 49 and 2 are the values of the variables age and N. of family members. On the other hand, if the individual is a man, the estimated value has to be increased by the gender parameter 2.84, reaching the value of 55.79. The formalization of the model

Table 12.5 Quantile regressions' parameters are the median of the index

Variable (*Reference level*)	Level of variable	Estimate	Pr > |t|
Intercept		83.23	<.0001
Gender (*Woman*)	Man	2.84	<.0001
Marital status	Unmarried	7.96	<.0001
(*Divorced/Widower/Separ.*)	Married	8.22	<.0001
Education	University or higher	7.78	<.0001
(*Middle or lower*)	High school	4.76	<.0001
Work conditions	Occupied	−11.27	<.0001
(*Retired/Student*)	Unempl./housekeeper	−14.90	<.0001
Smoking (*Non-smoker*)	Smoker	−5.45	<.0001
Religious practice	Often	13.38	<.0001
(*Rarely*)	Sometimes	10.42	<.0001
Discuss politics	Often	4.42	<.0001
(*Rarely*)	Sometimes	3.67	<.0001
House contract	Rent	−6.79	<.0001
(*Free use without property*)	Property	2.75	0.059
Economic change (1 year)	Better	19.88	<.0001
(*Worse*)	Equal	19.72	<.0001
House type	Distinguished	16.12	<.0001
(*Rural/Public*)	Average city house	11.53	<.0001
Geo. Partition	North	22.57	<.0001
(*South*)	Center	13.57	<.0001
Age		−0.56	<.0001
N. Family members		−1.57	<.0001

allows a simple approach to interpretation: men have a higher median life satisfaction than women, given all the other explanatory variables in the model (at least at the median point).

The table of estimates at the median can be used as a reference to study the differences among the explanatory variables. The direction of results is not surprising, but their dimension is quite interesting. If we exclude age, the most important variables are Economical change and Geographical Partition. The economic change agrees with the concept of relative satisfaction, which assesses how the satisfaction is a relative measure; in this case, the effect comes from the comparison with the past. If the situation is worse than the previous year, the level of satisfaction decreases strongly. The comparison between equal and better level of this variable gives non-significant results under the 70-th quantile.

The geographical partition probably acts as a wide summary of relevant factors for the determination of life satisfaction. Indeed, the southern partition has the most negative effect on the quantiles of satisfaction. The comparison between the center and the north underlies a statistical difference, assessing the northern partition as the best environment for life satisfaction improvement.

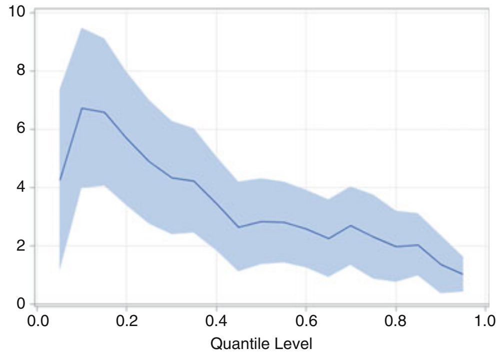


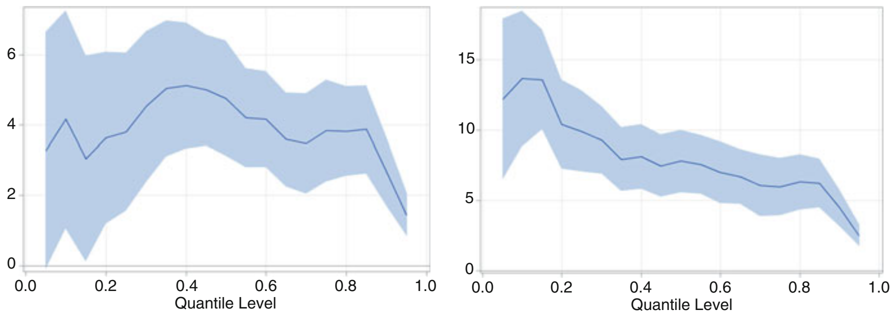
Fig. 12.13 Quantile regression of variable *Gender* – Parameter of Men from the 5-th to the 95-th quantile

It is useful to evaluate the quantile regression's estimates among many percentiles. In the following, those trends are represented: every graph represents the estimation at the percentiles from the 5-th to the 95-th with steps of 5 quantiles.

The values of the intercept move in the interval (−28; 160), almost linearly with respect to the increase of quantiles. In Fig. 12.13 the estimated parameters for the men group is plotted. There is strong evidence of a significantly higher level of satisfaction for the male population with respect to the female one. For instance, in the lower quantiles, the men' satisfaction has a value 4 and 7 points higher. Furthermore, the 95% confidence interval proves the male parameter to be significantly higher than zero in every quantile. Moreover, we can observe that the positive effect of being male is reduced in the proximity of highest life satisfaction to less than half of the observed maximum.

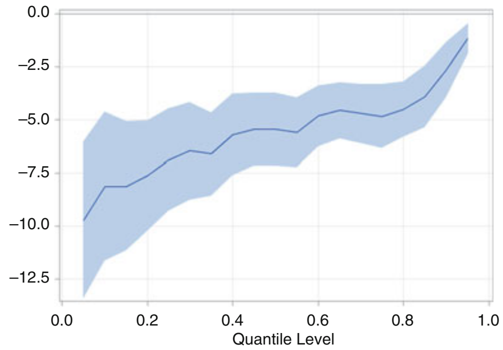
The effects of the Marital Status variable were to be expected: married and unmarried individuals show a higher level of life satisfaction with respect to the class of individuals whose relations were interrupted. An important note is in the composition of the group of unmarried people, which contains both singles and individuals that are too young for marriage. The status of unmarried has a different effect on the life satisfaction among the quantiles: the estimate is maximum around the 30-th percentile, where the value is 12.6, then, it constantly decreases to 0.25 where the estimation is not significantly different from zero from ninth decile. This type of fluctuation of the effect is visible only with the use of quantile regression. In this application, we get important information by this method: at the lower levels of life satisfaction, the Unmarried effect behaves differently from the Married one.

The effect of education is sharper: when the level of education increases, the median life satisfaction grows too. The parameters are evaluated on the level of middle school or lower. In Figs. 12.14 and 12.15 the effects are represented, showing that the effect of the university is stronger than high school; the effect of a high school education has the same direction as the graduation, but it has a smaller intensity. The university effect is stronger in the low levels of life satisfaction, which means that the less satisfied graduated individuals are significantly (around 14 points) more satisfied than less satisfied people with a middle or low education. The differences among the quantiles are also clear but strong as in some other



Figs. 12.14 and 12.15 Quantile regression of the variable *Education*. Parameters of High School and University from the 5-th to the 95-th quantile

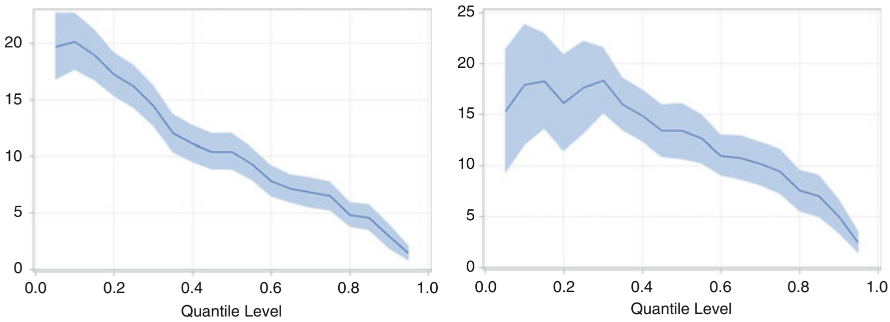
Fig. 12.16 Quantile regression of the variable *Smoking habit*. Parameter of Smoker from the 5-th to the 95-th quantile



variables (gender, marital status, and others); it probably means that education is useful as a tool for satisfaction at every level and not only for the low level of general satisfaction. Furthermore, the parameter of High school reaches its maximum around the 40-th quantile showing a different shape from the University's effect. Hence, education is certainly a good investment to gain good results and increase life satisfaction for everyone.

Studying the habits of the individuals, one result shows significantly: the individuals who belong to the group of smokers show a significantly lower level of life satisfaction respect to non-smokers and those who stopped smoking (Fig. 12.16). Furthermore, the group of Non-Smokers receives a satisfaction boost from the subset of people who stopped smoking recently and who experienced the physical and psychological effect.

These results are completely in line with literature. In their work, Grant et al. (2009) confirmed the association between life satisfaction and health-promoting behavior that is likely to be bidirectional. Regarding quitters, in contrast to continuing smokers, an improved subjective well-being has been reported that could be the motivation for the quit attempts by individuals Piper et al. (2012).



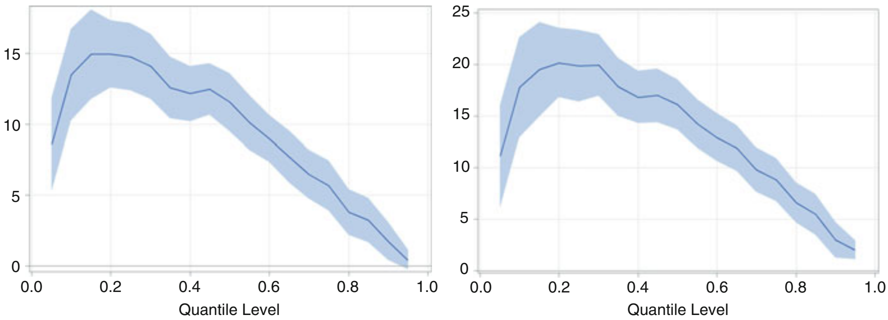
Figs. 12.17 and 12.18 Quantile regression of variable religious practice occasional and frequent religious practice from the 5-th to the 95-th quantile

The effect of religious participation is plotted in Figs. 12.17 and 12.18. The main goal of this type of information is to grasp the connection between religiosity and life satisfaction, but the limits of a single variable in the description of such a complex concept are straightforward. Hence, using this variable, together with the Political Participation variable, we want to present some perspectives on social inclusion and participation. People who attend religious services Often and Sometimes are both more satisfied with life than those who rarely practice. However, rigorous practicing individuals are not more satisfied than those interviewed who visit religious places once a week or less but more than once a month. To be exact, the confidence intervals of the groups are overlapping at many quantiles.

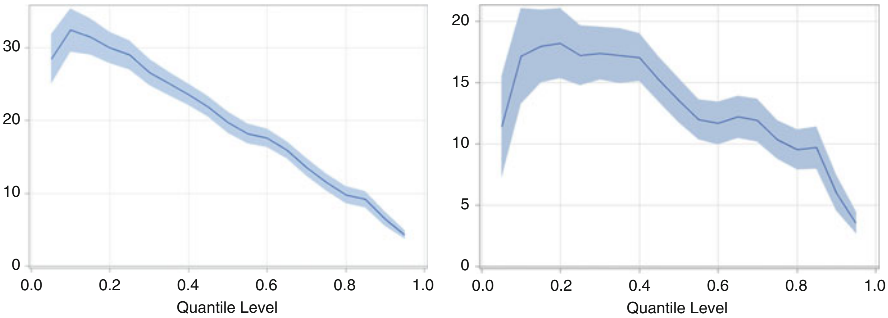
The effects of political participation have the same shape as a religious practice but are less intense (not shown). The Often level and the Sometimes level are almost overlapping. This variable is obtained by asking how often the individuals talk about politics. Hence, it becomes possible to understand part of the involvement of the individual in society changes.

This variable together with Religious practice are useful examples to appreciate the quantile regression results: both show the satisfactory effect of participation which is far more important at the lower levels and decreases with the improvement of life satisfaction.

There are three explanatory variables that help to model the economic situation of the subjects: the House type, the House contract, and the Economic changes. The effects of the first variable are represented using Public houses and Rural Houses as the reference level. The plotted effects of Distinguished and City houses are similar in shape. In the low level of housing, the subjects declare a lower life satisfaction; this difference is stronger between the second and the fourth deciles. We can compare average housing to distinguished housing only by regarding the magnitude of effects, while the shape remains similar for both the parameters among quantiles. The Average houses parameter has a maximum value of 15; this value is equal to 20 for Distinguished houses (Figs. 12.19 and 12.20). The value of the estimates for the group of average city houses is not significantly different from zero in the highest evaluated quantile.



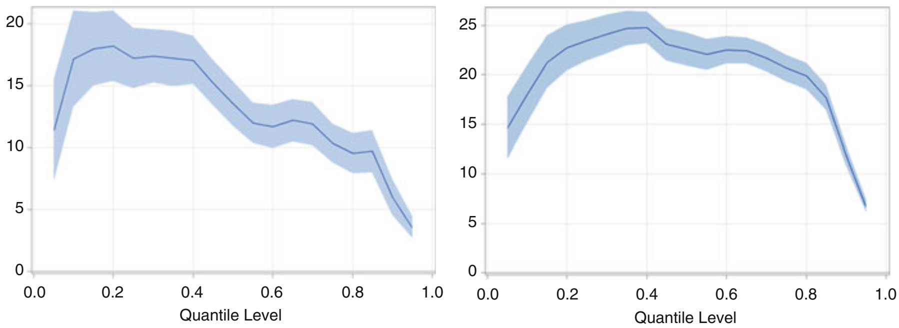
Figs. 12.19 and 12.20 Quantile regression of variable *House Type*. Parameter of an Average city house and Distinguished house from the 5-th to the 95-th quantile



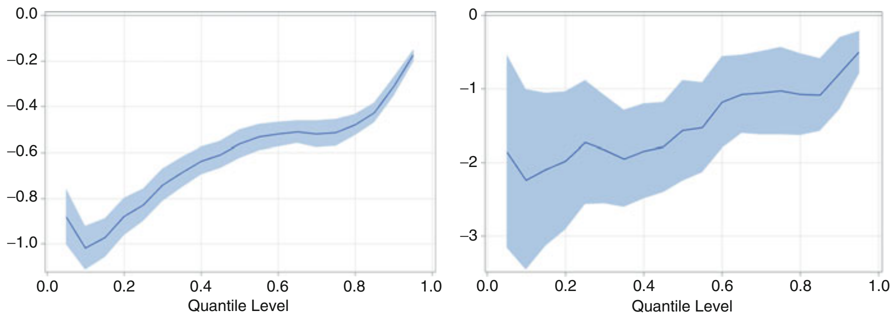
Figs. 12.21 and 12.22 Quantile regression of variable *Economic Situation Change*. Parameter of stable and improved economic situation from the 5-th to the 95-th quantile

The results given by the House contract are easier to interpret. The value Free use without property takes the reference role. This group is made by the individuals who own the right to use the house without paying (because it is free for them or because they sold the property in a delayed transaction). The satisfaction of house owners is higher, but this result is significant only in the first three deciles. The effect of being a tenant is negative, with a quite constant trend; in this case, the effect is significantly different from zero after quantile 0.10 where the estimate keeps showing very large confidence intervals.

In the Economic Change question, subjects were asked to say if they considered the economic situation of the family improved, stable, or deteriorated with respect to the previous year. The reference level is Worse, so it is possible to observe the estimated effect of a stable situation (Fig. 12.21), and the effect of an improved situation, which is presented in Fig. 12.22. The estimated trends for Better and Equal are almost overlapping in intensity and shape, which means that the real difference is observable between them and the group who perceives its situation as Worse. Indeed, the members of this negative class are less satisfied, by a large gap.



Figs. 12.22 and 12.23 Quantile regression of variable *Geographical Partitions*. Central and Northern geographical partitions from the 5-th to the 95-th quantile



Figs. 12.24 and 12.25 Quantile regression of variables *Age* and *Family Members*. Parameter of age and number of Family Members from the 5-th to the 95-th quantile

According to the estimation on Geographical partitions, the comparison between Southern and Northern is clear (Fig. 12.23): the inhabitants of the North are more satisfied along the entire spectrum of observation. Especially in the middle-low levels, the value of the Northern parameter is 10% of the entire range of the indicator H_τ, exactly 25 out of 256. Central Italy is still more satisfied than the South, but the effect is smaller respect to the North (Fig. 12.22). This result respects the knowledge about the socio-economical situation of these partitions.

Like the last issue, we want to take into account two demographic data. The Age and the number of Family members are quantitative variables. The estimated value is the gain/loss of life satisfaction quantile experienced by an individual when the considered variable increases by one unit.

The effect of Age is negative on Life Satisfaction (Fig. 12.24); age has a range of 90 units, then, in the 1-st decile, where the estimate of the parameter is -1.02, being the oldest of the population would mean losing 90 points with respect to the youngest. The number of family members has a negative effect too, according to Fig. 12.25; in the lowest deciles of life satisfaction, every additional member decreases the satisfaction by a value around 2.

Conclusion

In this chapter, the poset theory has been introduced and explained, with particular focus on the tools that can be used to construct a synthetic measure of complex concepts using a sub-sample of the set of linear extensions. A procedure for the construction of a synthetic indicator has been proposed and then used to obtain a measure of a complex and unobservable concept such as life satisfaction. The result is a unique indicator that can be used to represent the concept under study.

In this practice, the estimation of the effect of many socio-economical variables on life satisfaction has been enhanced, confirming many results in literature. We estimated the effects of the explanatory variables on the different quantiles of the index of life satisfaction: many variables show a stronger effect for low quantiles of life satisfaction; this obtained outcome justifies the use of a method such as quantile regression. In particular, the socioeconomic variables such as the economic change in the last year and the geographical partition show strong effects on life satisfaction.

References

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95*(7), 476–487.

Arrow, K. J., & Raynaud, H. (1986). *Social choice and multicriterion decision-making* (p. 1). Boston: MIT Press Books.

Birkhoff, G. (1948). *Lattice theory* (Vol. 25). New York: American Mathematical Society.

Brightwell, G., & Winkler, P. (1991). Counting linear extensions. *Order, 8*(3), 225–242.

Brüggemann, R., & Carlsen, L. (2011). An improved estimation of averaged ranks of partial orders. *MATCH Communications in Mathematical and in Computer Chemistry, 65*, 383–414.

Brüggemann, R., Sørensen, P. B., Lerche, D., & Carlsen, L. (2004). Estimation of averaged ranks by a local partial order model. *Journal of Chemical Information and Computer Sciences, 44*(2), 618–625.

Davey, B. A., & Priestley, H. A. (2002). *Introduction to lattices and order*. New York: Cambridge University Press.

De Loof, K. (2009). *Efficient computation of rank probabilities in posets*. Ph.D. thesis, University of Ghent.

De Loof, K., De Baets, B., & De Meyer, H. (2011). Approximation of average ranks in posets. *MATCH Communications in Mathematical and in Computer Chemistry, 66*, 219–229.

Dyer, M., Frieze, A., & Kannan, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM), 38*(1), 1–17.

Fattore, M. (2015). Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research, 128*(2), 835–858.

Fattore, M., & Arcagni, A. (2014). Parsec: An r package for poset-based evaluation of multidimensional poverty. In *Multi-indicator systems and modelling in partial order* (pp. 317–330). New York: Springer.

Fattore, M., Brüggemann, R., & Owsiński, J. (2011). Using poset theory to compare fuzzy multidimensional material deprivation across regions. In *New perspectives in statistical modeling and data analysis* (pp. 49–56). New York: Springer.

Grant, N., Wardle, J., & Steptoe, A. (2009). The relationship between life satisfaction and health behavior: A cross-cultural analysis of young adults. *International Journal of Behavioral Medicine, 16*(3), 259–268.

Karzanov, A., & Khachiyan, L. (1991). On the conductance of order Markov chains. *Order, 8*(1), 7–15.

Koenker, R. (2005). *Quantile regression*, *Number 38*. New York: Cambridge University Press – CUP.

Koenker, R., & Bassett Jr., G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society, 46*(1), 33–50.

Maggino, F., & Fattore, M. (2011). New tools for the construction of ranking and evaluation indicators in multidimensional systems of ordinal variables. In: *Proceedings of the "New Techniques and Technologies for Statistics"*, Brussels.

Munda, G. (2008). *Social multi-criteria evaluation for a sustainable economy*. Berlin: Springer.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on constructing composite indicators. *OECD statistic working paper*.

Piper, M. E., Kenford, S., Fiore, M. C., & Baker, T. B. (2012). Smoking cessation and quality of life: Changes in life satisfaction over 3 years following a quit attempt. *Annals of Behavioral Medicine, 43*(2), 262–270.

Schröder, B. (2012). *Ordered sets: An introduction*. Birkhäuser: Springer Science & Business Media.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica: Journal of the Econometric Society, 44*(2), 219–231.

Chapter 13
Joint Analysis of Structural Models and Performance: Merging Clustering and Composite Indicators in the Analysis of Europe 2020 Strategy

Tommaso Rondinella and Elena Grimaccia

Introduction

Representing synthetically the complexity of multidimensional phenomena is especially challenging when comparing different territorial areas according to multiple factors. In particular, in the comparison of territories (Countries in our specific case), the attention can be either on the performance shown by Countries, usually evaluated through rankings, or on the structural models which characterize different compositions of the various elements.

We present here a technique which combines cluster analysis with the use of a composite indicator, thus permitting to identify Countries both according to their structural characteristics and to their overall performance.

We chose to test the analysis upon the set of indicators of the Europe 2020 strategy, a reduced set of indicators which covers very diverse qualifications of economic growth. Set up by the European Commission Europe 2020, indicators benefit from societal and political consensus about welfare and development relevant dimensions. Moreover, they present all the methodological and accessibility characteristics that permit the analysis in terms of synthesis and classification.

In 2010, with the expiration of the Lisbon Strategy for Growth and Jobs, the governments of the European Union launched the Europe 2020 strategy (European Commission 2010), a set of guidelines of action in order to fix mid-term political economy targets which extended the notion of economic growth to some other key

T. Rondinella (✉)
Istat – National Institute of Statistics, Via dell'Agnolo 80, 50122 Firenze, Italy
e-mail: rondinella@istat.it

E. Grimaccia
Istat – National Institute of Statistics, Via Cesare Balbo 16, 00184 Rome, Italy
e-mail: elgrimac@istat.it

F. Maggino (ed.), *Complexity in Society: From Indicators Construction to their Synthesis*, Social Indicators Research Series 70, DOI 10.1007/978-3-319-60595-1_13

aspects which should characterize the European model: smartness, inclusion and sustainability. The broad policy objectives have been declined in quantitative and measurable targets, making governments "accountable" to the citizens and to the Commission. The targets are measured by eight headline indicators, concerning research and innovation, employment, education, poverty, climate change, renewable sources and energy efficiency. The strategy aims therefore at increasing European competitiveness maintaining a social market economy and improving resource efficiency.

We consider data for 10 years, before and after the economic crisis, in order to highlight the evolution of the similarities and dissimilarities of the 28 European countries in the considered period.

The combined use of cluster analysis and composite indicators allows us to represent the level of development combined with the different structures which characterise groups of countries.

Finally, the building of composite indicators for the three Europe 2020 domains of monitoring (smartness, inclusion and sustainability), as well as for the overall strategy, describe the growth pattern of European countries according to different categorisations.

Conceptual Framework: The Europe 2020 Strategy

In June 2010, the European Council adopted the "Europe 2020 Strategy", put forward by the European Commission (2010). It defines three priorities for growth in the European countries:

- smart growth: developing an economy based on knowledge and innovation;
- inclusive growth: fostering a high employment economy, delivering social and territorial cohesion;
- sustainable growth: promoting a greener and more resources efficient, economy.

The broad policy objectives (smartness, inclusion and sustainability) were declined in measurable, and thus well-defined, numerical targets, making governments "accountable" to the citizens and to the Commission.

Targets have been fixed for each country and indicator and the monitoring system is foreseen within the European Semester of economic policy coordination. This implies that Government must not only monitor the dynamic of the indicators but have also to make explicit the measures for achieving the goals (European Commission 2010).

"Smart growth" is measured by:

- the proportion of early leavers from education and training, and should be under 10%,
- the quote of persons (25–34 years old) with tertiary educational attainment should be at least 40%, and

- the percentage of gross domestic expenditure on research and development (R&D) on GDP should reach 3% in the EU as a whole.

"Inclusive growth" is measured by:

- the employment rate, with a target of 75% of the population aged 20–64 should be employed, and
- the number of people at risk of poverty or social exclusion which should decrease by 20 million from 2008 in total in EU.

"Sustainable growth" is measured by:

- the greenhouse gas emissions, to be reduced by 20% compared to 1990,
- the share of renewable energy sources in final energy consumption, to be increased to 20%, and
- the energy efficiency that should be improved by 20%, too.

Methodologies

Cluster Analysis

We have applied a cluster analysis to the 28 EU Member States, with the aim to identify the groups of Countries with similar performance with regard to the eight Europe 2020 Indicators. We used a hierarchical clustering, and we chose an Euclidean distance as a measure of dissimilarity, and Ward's criterion as a linkage method. Ward's method gave the best results and it is also the most common in literature).[1]

In general, more homogenous groups form at an earlier stage and are better characterized, while those with higher internal dissimilarity present less defined communalities. In Fig. 13.1 we present the dendrogram for year 2013.

We repeated the analysis for each year between 2004 and 2013 fixing a value of 55 for the dissimilarity measure, and have identified seven groups of Countries every year, (Fig. 13.2).

[1]Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. In order to decide which clusters should be combined (for agglomerative, "bottom up" approaches), or where a cluster should be split (for divisive, "top down" approaches), a measure of dissimilarity between sets of observations is required. In most methods of hierarchical clustering, this is achieved by use of an appropriate measure of distance, and a linkage criterion which specifies the dissimilarity of sets as a function of the pairwise distances of observations in the sets. In Ward's minimum-variance method, the distance between two clusters is the ANOVA sum of squares between the two clusters added up over all the variables. At each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation (Milligan 1980).

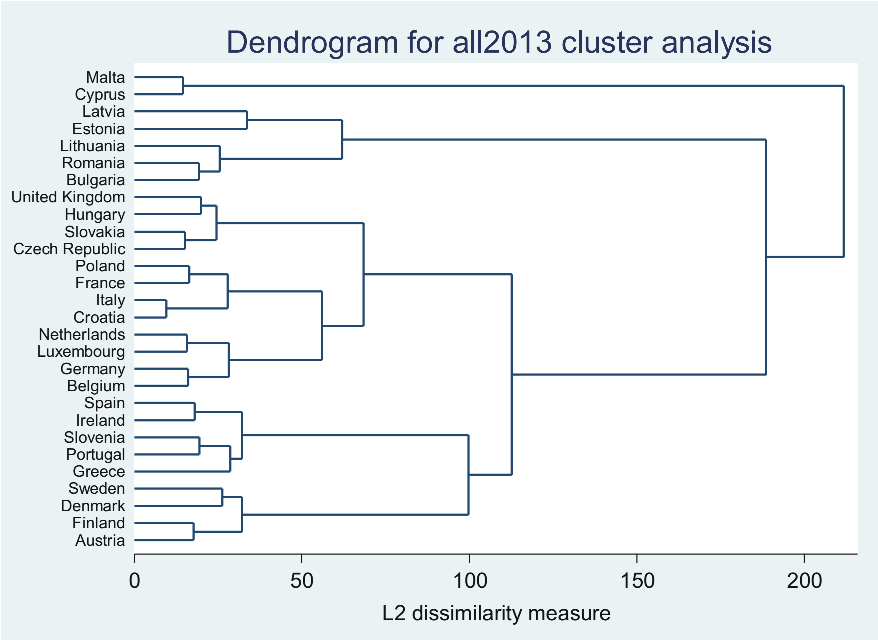


Fig. 13.1 Dendrogram for year 2013 cluster analysis on Europe 2020 indicators *(Ward method, Euclidean distance)*

The analysis of the eight indicators in each group permitted to identify their most relevant characteristics.[2]

The first group considered is the one of "Nordic Countries" (Sweden, Finland and Denmark), that present the highest level of education in the population and research expenditure (smartness), the highest level of employment and lowest of poverty (inclusion), and the best performance in terms of environmental sustainability. Austria and Slovenia present a peculiar pattern, being part of the first group most of the time, but with worse performance in tertiary education and GHG emission, and also less school drop-outs.

The group of Western Europe Countries (France, Benelux, UK and Germany) is characterised by a high level of research & development expenditure, good performance of the education system, a good level of employment and inclusion, but a moderate level of emission.

The Eastern Europe Group (substantially the "Visegrad Group") shows a good performance in smartness but lower level of inclusion and, mostly, it can be identified by a lower attention on energy saving.

[2]With regards to as the analysis of each single Country, we refer to the group where the country is most frequently included (mode).

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Mode
Finland	1	1	1	1	1	1	1	1	1	1	1
Sweden	1	1	1	1	1	1	1	1	1	1	1
Denmark	2	1	1	1	1	1	1	1	1	1	1
Austria	1	4	4	1	4	1	1	1	6	1	1
Slovenia	1	4	4	1	4	1	1	1	6	6	1
France	2	2	2	2	2	2	2	2	2	4	2
Belgium	2	2	2	2	2	2	2	2	2	2	2
Netherlands	2	2	2	2	2	2	2	2	2	2	2
Luxembourg	2	2	2	2	2	2	2	2	2	2	2
UK	2	2	3	2	2	2	3	3	2	3	2
Germany	2	2	3	3	3	2	3	3	2	2	2
Poland	3	3	3	3	3	2	3	2	2	4	3
Czech Rep	3	3	3	3	3	3	3	3	3	3	3
Hungary	3	3	3	3	3	3	3	3	3	3	3
Slovakia	3	3	3	3	3	3	3	3	3	3	3
Croatia	3	3	4	3	3	2	3	6	3	4	3
Italy	4	4	4	3	4	2	6	6	3	4	4
Greece	4	4	4	4	4	6	6	6	6	6	4
Ireland	4	4	4	4	4	6	2	2	6	6	6
Portugal	6	6	6	4	6	6	6	6	6	6	6
Spain	6	6	6	6	6	6	6	6	6	6	6
Cyprus	6	6	6	6	6	7	7	7	7	7	7
Malta	6	6	6	6	6	7	7	7	7	7	7
Estonia	5	5	5	5	5	5	5	5	5	5	5
Latvia	5	5	5	5	5	5	5	5	5	5	5
Lithuania	5	5	5	5	5	5	5	5	5	5.5	5
Bulgaria	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5
Romania	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.5

Legend:	1	2	3	4	6	7	5	5.5
	Nordic	Western Europe	Eastern Europe	Other	Mediterranean	MT+CY	Baltic	RO+BG

Fig. 13.2 – Groups of Cluster analysis on Europe 2020 indicators (*cut L2 = 55*), years 2004–2013

Mediterranean Countries present a more problematic situation, with lower percentage of the population with high level of education, low level of expenditure in R&D, quite low employment and problems of inclusion.

The group of Baltic Republics and that of Romania and Bulgaria present a low percentage of R&D expenditure, high level of poverty but a good performance on sustainability, with low emission and high percentage of the use of renewable sources.

Finally, Malta and Cyprus are characterized by a very weak performance on sustainability indicators.

Through the years, the analysis presents quite stable results in the identification of groups. In the "Nordic Countries" group, Finland, Sweden and Denmark are present for the whole period. We can say that the best performing countries present also a stability in the Europe 2020 indicators patterns trough the economic crisis. "Western Europe" cluster is built around the core of Benelux and France that exits in 2013. Germany swings between Eastern and Western clusters. The "Eastern Europe" group is formed by Czech Republic, Slovakia and Hungary for all the decade, with Poland before the crisis. Mediterranean Countries group present a different composition before and after the crisis, most of the time including Ireland and with Cyprus and Malta constituting a different group after 2008. The last two clusters, those of Baltic countries and Romania and Bulgaria, are substantially stable all over the decade.

Composite Indicators

In order to assess the performance of Member States according to Europe 2020s indicators we opted for the building of a composite indicator for each one of the three dimensions (smartness, inclusion and sustainability) as well as for the overall strategy.

The methodology we adopted is inspired by the work done by Mazziotta and Pareto within the BES initiative (Istat 2015) for building a composite index able to maintain comparability both in space and time.

Most aggregation techniques (OECD and JRC 2008) allow comparisons between units (countries) yet changing every year the standardization reference, thus the comparison over time shows only relative changes with respect to other units. The Mazziotta-Pareto method, a min-max based on an initial reference value, is able to monitor also time trends as a based index.

Differently to the MPI (Mazziotta and Pareto 2014) we opted for a more transparent measure with perfect substitutability among indicators and therefore we have not applied their correction based on the internal variability in one country's indicators.

The standardization is expressed as:

$$r_{ij} = \frac{\left(x_{ij} - Inf_x\right)}{\left(Sup_x - Inf_x\right)} 60 + 70 \qquad (13.1)$$

Where x_{ij} is the value of the variable x for the unit j at time i; Min_x and Max_x are the "goalposts" for the variable j. If the variable has negative polarity (in our case school leave, poverty, greenhouse gasses and energy consumption), then the complement of (1) with respect to 200 is computed.

$$\begin{cases} Inf_x = Ref_x - \Delta \\ Sup_x = Ref_x + \Delta \end{cases} \qquad \Delta = \left(Max_{A_x} - Min_{A_x}\right)/2$$

The goalposts are calculated as an interval $\pm\Delta$ around a reference value for the variable x, where Δ is the midrange of the dataset A_x comprising all units and times taken into account for the variable x.

In our case, the reference value considered is the one for European Union 28 in year 2004, while A_x comprises the values for the 28 countries in years 2004–2013.

Using this normalization, every variable will assume value equals to 100 for EU28 in 2004, and a value relative to such reference for each country and each year allowing both space and time comparison.

We built composite indexes of Smartness, Inclusion and Sustainability, as simple average of normalised values of the relevant indicators. A composite index for Europe 2020 is the simple average of the three. Should we have built it as the average of all the eight considered indicators, inclusiveness, represented only by two indicators, would have been underweighted.

The results of composite indicators, as known, are characterized by a number of pros and cons (OECD and JRC 2008). They provide for an effective quantitative synthesis of multiple variables allowing for comparison and ranking among units, yet for their interpretation it is always necessary to go back to initial data, especially when the index comprises very different phenomena. Therefore, a similar value can describe very different situations, typically for mid-range units.

Looking at the results from the application of the composite indicators over time, in the EU28 the smartness index shows a continuous increase (8% in the whole considered decade) thanks to the positive contribution of all the indicators. Similarly, also the sustainability index increases by 7%, but the pattern is determined by the rising use of renewable sources, and a substantial stability of GHG emissions and energy consumption.

Inclusion index, instead, rose only by 2% with a dynamic of fall and recovery through the crisis, and a number of countries experiencing a very weak or no recovery for both employment and poverty measures.

The overall Europe 2020 index rises by 6% during the decade. Looking at country results, we find Sweden, Finland and Denmark on the top and Malta at the bottom. Poland and Bulgaria show an increase by over 10%, while the index increased only by 3% in Croatia, Greece and Ireland.

As said, if we look at the value of the index we can find very similar values for countries which have different profiles. Figure 13.3 shows how clusters would reshuffle when ordering countries according to the composite index.

Joint Visualization

We propose a form of visualization of synthetic data which merges both cluster analysis and a composite measure, in order to show the level and the evolution of countries while still focusing on their structural similarities, as represented by the clustering (Fig. 13.4). Colouring composite indexes scores, it is easily possible to highlight the relative position of each country as well as it progression through the years. It also helps to show the different characterization of the clusters, and how the indexes' scores and the structural elements emerged from the clustering do not necessarily go together with sometimes very relevant differences within a cluster.

Performances within clusters tend to be quite homogenous, but relevant differences can be recognized in the smartness domain for the cases of Ireland or Latvia, in the Inclusion domain for the cases of the Netherlands or the Central European cluster, in the Sustainability domain for the Mediterranean cluster and Malta and Cyprus. Finally, in the overall Europe 2020 index, differences are obviously

Fig. 13.3 Europe 2020 index ranking, year 2013, and reference cluster

Cluster	Country	2013
1	Sweden	121.7
1	Finland	115.8
1	Denmark	115.0
2	Germany	111.0
5	Estonia	110.7
1	Austria	109.8
2	Netherlands	109.5
3	Czech Republic	109.4
2	France	108.9
2	United Kingdom	108.7
1	Slovenia	107.9
2	Belgium	107.2
5	Lithuania	107.1
2	Luxembourg	106.5
5	Latvia	106.2
6	Ireland	104.2
3	Slovakia	103.0
3	Hungary	101.7
6	Portugal	101.6
3	Poland	100.8
7	Cyprus	100.1
5.5	Romania	98.7
3	Croatia	98.0
5.5	Bulgaria	97.8
6	Spain	97.7
4	Italy	97.7
4	Greece	97.0
7	Malta	89.9

smoothened but still some relevant differences emerge for Czech Republic or Malta and Cyprus.

On the other way round, we can find very similar values for countries which have different profiles. For example Germany and Estonia, France and Czech Republic, UK and Slovenia, Poland and Cyprus or Spain and Bulgaria.

Conclusions

Europe 2020 indicators allow us to analyse the different characteristics and patterns followed by the European countries, pointing out differences and similarities. They worked well in describing the effects of the crisis and they seem robust and sensible to changes. Member States show different results with respect to the eight indicators allowing the identification of similarities characterizing different patterns of development, the most evident being “Nordic countries”, “Western Europe”,

Smartness

Cluster		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	% change	Cluster
1	Finland	120	121	122	123	123	124	124	124	123	121	0.3%	1
1	Sweden	117	118	121	120	122	123	123	123	124	123	5.4%	1
1	Denmark	115	115	115	112	114	117	116	117	118	119	4.0%	1
1	Austria	103	105	104	104	106	108	108	108	110	111	7.1%	1
1	Slovenia	103	103	105	106	107	108	111	115	118	117	13.6%	1
2	France	108	109	110	110	111	113	112	113	113	114	5.1%	2
2	Belgium	108	108	109	110	111	111	112	112	113	113	4.4%	2
2	Netherlan	105	106	107	107	109	109	111	112	113	112	7.0%	2
2	Luxembou	103	105	105	105	107	114	112	113	113	114	10.5%	2
2	United Kir	105	106	107	105	106	107	108	110	111	111	6.0%	2
2	Germany	106	105	105	106	108	110	109	110	112	112	5.7%	2
3	Poland	96	97	98	99	101	102	104	104	106	107	11.7%	3
3	Czech Rep	95	96	97	97	97	99	101	104	106	106	11.6%	3
3	Hungary	93	93	94	95	96	98	99	100	101	103	10.5%	3
3	Slovakia	91	92	92	92	93	94	97	98	99	100	9.5%	3
3	Croatia	97	96	95	96	98	98	100	100	99	101	4.0%	3
4	Italy	87	88	90	91	91	92	92	93	94	95	8.3%	4
4	Ireland	104	105	106	107	110	112	113	113	115	115	10.6%	4
6	Greece	93	94	93	94	94	94	95	96	98	101	9.1%	6
6	Portugal	77	78	80	82	86	89	91	95	96	97	25.9%	6
6	Spain	92	94	95	95	96	96	98	99	99	100	8.4%	6
7	Cyprus	97	98	102	104	103	104	103	105	106	106	10.2%	7
7	Malta	75	80	82	82	84	84	86	88	89	92	22.5%	7
5	Estonia	97	99	100	100	102	103	108	112	111	111	15.0%	5
5	Latvia	89	90	91	93	94	95	98	101	102	103	15.8%	5
5	Lithuania	99	104	105	103	105	105	107	109	110	112	13.0%	5
5.5	Bulgaria	89	90	91	93	94	94	95	96	96	97	8.4%	5.5
5.5	Romania	81	83	85	86	88	88	87	89	90	90	10.6%	5.5
	EU28	**100**	**101**	**101**	**102**	**103**	**104**	**105**	**106**	**107**	**108**	**8%**	

Inclusiveness

Cluster		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	% change	Cluster
1	Finland	111	112	113	113	115	112	112	112	113	113	1.8%	1
1	Sweden	117	119	118	121	121	118	119	119	120	120	2.5%	1
1	Denmark	117	117	119	118	119	116	114	113	113	113	-3.1%	1
1	Austria	109	109	110	112	110	111	112	112	112	112	3.3%	1
1	Slovenia	107	109	110	111	111	110	108	105	105	103	-3.5%	1
2	France	106	107	107	107	108	107	106	106	107	107	1.1%	2
2	Belgium	101	101	102	103	104	103	103	103	103	103	2.2%	2
2	Netherlan	114	114	116	118	119	119	117	117	118	116	2.4%	2
2	Luxembou	107	107	108	109	108	108	109	109	109	108	1.8%	2
2	United Kir	108	109	110	111	110	110	108	109	109	109	0.4%	2
2	Germany	107	107	108	109	111	111	112	114	114	114	6.5%	2
3	Poland	74	78	84	90	95	96	96	96	97	97	32.5%	3
3	Czech Rep	106	108	109	111	112	111	110	110	111	113	6.3%	3
3	Hungary	90	91	91	92	92	90	89	89	90	91	0.6%	3
3	Slovakia	89	93	98	103	105	103	100	101	101	101	13.7%	3
3	Croatia	88	89	89	93	94	93	91	88	86	87	-2.1%	3
4	Italy	93	94	95	95	96	95	94	92	91	90	-3.4%	4
4	Ireland	105	106	108	109	107	100	96	94	94	96	-9.0%	4
6	Greece	93	95	96	97	98	97	95	88	81	78	-16.1%	6
6	Portugal	105	105	106	106	106	105	104	103	99	97	-7.5%	6
6	Spain	98	101	103	104	102	97	95	94	91	90	-8.8%	6
7	Cyprus	109	108	110	111	112	110	109	107	102	99	-9.2%	7
7	Malta	92	93	94	94	95	94	95	96	97	98	6.5%	7
5	Estonia	103	105	112	113	113	105	102	105	107	108	4.5%	5
5	Latvia	87	89	96	103	104	92	89	90	95	97	11.3%	5
5	Lithuania	89	94	98	104	104	97	92	95	97	100	11.7%	5
5.5	Bulgaria	70	72	76	80	92	89	84	81	81	82	18.2%	5.5
5.5	Romania	84	84	85	85	86	85	88	87	88	88	5.6%	5.5
	EU28	**100**	**101**	**102**	**104**	**105**	**103**	**103**	**102**	**102**	**102**	**2%**	

Sustainability

Cluster		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	% change	Cluster
1	Finland	105	109	105	105	109	112	108	111	113	114	8.7%	1
1	Sweden	112	113	115	116	117	121	116	119	121	122	9.3%	1
1	Denmark	101	102	100	102	103	107	105	108	112	112	11.3%	1
1	Austria	102	100	102	103	103	107	104	106	107	106	4.8%	1
1	Slovenia	100	99	99	99	95	103	101	100	102	103	2.8%	1
2	France	99	100	100	102	102	104	103	106	106	105	6.2%	2
2	Belgium	96	98	99	101	99	104	100	103	106	105	9.3%	2
2	Netherlan	96	97	98	97	97	100	96	100	100	100	4.2%	2
2	Luxembou	92	91	92	94	93	97	94	94	96	97	6.1%	2
2	United Kir	98	98	99	100	100	105	103	107	106	106	8.5%	2
2	Germany	102	103	102	105	104	107	105	107	107	106	4.5%	2
3	Poland	100	100	97	97	97	99	94	97	97	98	-1.8%	3
3	Czech Rep	102	102	102	103	104	107	106	107	109	109	7.5%	3
3	Hungary	101	100	101	104	104	107	107	108	112	112	10.7%	3
3	Slovakia	103	102	103	104	103	108	105	108	110	108	4.2%	3
3	Croatia	101	100	98	97	97	100	101	103	106	107	5.8%	3
4	Italy	96	96	97	98	99	104	103	104	107	108	12.7%	4
4	Ireland	92	90	88	88	88	95	95	100	101	102	10.1%	4
6	Greece	95	94	93	92	94	97	101	103	107	112	17.0%	6
6	Portugal	97	96	98	99	101	103	104	106	109	111	14.5%	6
6	Spain	90	88	90	88	92	98	98	99	101	104	15.1%	6
7	Cyprus	89	88	86	85	83	86	87	88	91	95	7.4%	7
7	Malta	90	89	89	88	78	89	79	79	79	79	-11.7%	7
5	Estonia	112	110	110	106	108	115	112	114	113	113	1.1%	5
5	Latvia	117	116	113	110	112	117	113	117	117	118	1.0%	5
5	Lithuania	110	107	104	101	102	109	107	108	108	109	-0.3%	5
5.5	Bulgaria	107	105	104	104	106	113	113	110	112	114	7.0%	5.5
5.5	Romania	108	109	108	110	110	116	116	115	116	118	9.0%	5.5
	EU28	**100**	**100**	**100**	**101**	**102**	**106**	**104**	**106**	**107**	**107**	**7%**	

Europe2020

Cluster		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	% change	Cluster
1	Finland	112	114	113	114	115	116	114	116	116	116	3.4%	1
1	Sweden	115	117	118	119	120	121	119	121	121	122	5.7%	1
1	Denmark	111	112	111	111	112	113	112	113	114	115	3.7%	1
1	Austria	105	104	105	106	107	109	108	109	110	110	5.0%	1
1	Slovenia	104	104	105	105	104	107	107	107	108	108	4.2%	1
2	France	105	105	106	106	107	108	107	108	109	109	4.1%	2
2	Belgium	102	102	104	105	105	106	105	106	107	107	5.2%	2
2	Netherlan	105	106	107	107	108	110	108	110	110	109	4.5%	2
2	Luxembou	100	101	101	102	103	106	105	105	106	107	6.1%	2
2	United Kir	104	104	105	105	105	107	107	108	108	109	4.8%	2
2	Germany	105	105	105	107	107	109	109	110	111	111	5.6%	2
3	Poland	90	92	93	95	97	99	98	99	100	101	12.4%	3
3	Czech Rep	101	102	103	104	104	106	106	107	109	109	8.4%	3
3	Hungary	95	95	95	97	97	98	98	99	101	102	7.4%	3
3	Slovakia	95	96	98	100	100	102	101	102	103	103	8.9%	3
3	Croatia	95	95	94	96	96	97	97	97	97	98	2.7%	3
4	Italy	92	93	94	95	95	97	96	96	97	98	5.9%	4
4	Ireland	101	100	101	101	102	102	101	102	103	104	3.6%	4
6	Greece	94	94	94	94	95	96	97	96	96	97	3.4%	6
6	Portugal	93	93	95	96	98	99	99	101	101	102	9.4%	6
6	Spain	93	94	96	96	97	97	97	97	97	98	4.5%	6
7	Cyprus	98	98	99	100	99	100	100	100	100	100	2.2%	7
7	Malta	86	87	88	88	86	89	87	88	88	90	4.8%	7
5	Estonia	104	105	107	106	108	108	107	111	110	111	6.5%	5
5	Latvia	98	98	100	102	103	101	100	103	104	106	8.6%	5
5	Lithuania	99	102	102	103	104	104	102	104	105	107	7.7%	5
5.5	Bulgaria	89	89	90	92	97	99	97	96	96	98	10.4%	5.5
5.5	Romania	91	92	93	94	95	96	97	97	98	99	8.4%	5.5
	EU28	**100**	**101**	**101**	**102**	**103**	**104**	**104**	**105**	**105**	**106**	**6%**	

Fig. 13.4 Joint visualization of clusters and composite indexes scores *(EU28 = 100, scores and percentage values; colour: from dark red to dark green on the scores' table and on the % change column)*

“Eastern Europe” “Mediterranean countries”, “Malta and Cyprus”, “Baltic countries” and “Bulgaria and Romania”. Yet, results and trends for some indicators may deeply differ even within the groups. Such differences have sometimes strong influence when one tries to synthetize the complexity through a composite indicator. Composite indicators, as a matter of fact, hide the actual composition of original indicators, leading to similar results for territories which may experience very different results for the original indicators. On the other hand they provide a synthetic representation of the overall performance as well as of the trends countries have been following over time. The joint representation of clustering and composite indexes allows the association between overall performance and different models of development for a more comprehensive synthetic understanding of complex multidimensional phenomena.

References

European Commission. (2010). *Europe 2020 A strategy for smart, sustainable and inclusive growth*, COM(2010) 2020 final. Brussels.

Istat. (2015). *BES 2015. Il benessere equo e sostenibile in Italia*. Rome: Istat.

Mazziotta, M., & Pareto, A. (2014). Non-compensatory composite indices for measuring changes over time: A comparative study. *CESS 2014 Conference of European Statistics Stakeholders* – November 24–25.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45*(3), 325–342.

OECD and JRC. (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: OECD.