

The goal of this chapter is to present foundational concepts and some operational definitions in the field of Computational Social Science (CSS for short) by introducing the main assumptions, features, and research areas. A key feature of CSS is its interdisciplinary nature. Computational modeling enables researchers to leverage and integrate knowledge from many different disciplines, not just the social sciences. This chapter also provides an overview of the whole textbook by providing a “peek” into each chapter. The purpose is not to enter into many details at this stage, but to provide a preview of some of the main ideas examined in subsequent chapters.

One of the key challenges in the field of Computational Social Science is that several relatively subtle or complicated ideas need to be introduced simultaneously. Social complexity, complex adaptive systems, computational models, and similar terms are introduced in this chapter, and later elaborated upon in greater depth. What we need for now are some initial concepts so that we may get started in establishing foundations. There is no attempt in this chapter to provide an exhaustive treatment of each and every term that is introduced.

1.1 What Is Computational Social Science?

The origin of social science—in the *pre*-computational age—can be traced back to Greek scholars, such as Aristotle, who conducted the first systematic investigations into the nature of social systems, governance, and the similarities and differences among monarchies, democracies, and aristocracies. In fact, Aristotle is often considered the first social science practitioner of comparative social research. Modern social science, however, is usually dated to the 17th century, when prominent French social scientists such as Auguste Comte first envisioned a natural science of social systems, complete with statistical and mathematical foundations and methods to enhance traditional historical and earlier philosophical approaches. Since then, the social sciences have developed a vast body of knowledge for understanding human and social behavior in its many forms (Bernard 2012). This is how modern anthro-

pology, economics, political science, psychology, and sociology—the so-called Big Five (Bernard 2012; Horowitz 2006; Steuer 2003)—were born four centuries ago.

The new field of **Computational Social Science** can be defined as the interdisciplinary investigation of the social universe on many scales, ranging from individual actors to the largest groupings, through the medium of computation. This working definition is somewhat long and will be refined later as we examine many topics involved in the practice of CSS and the variety of computational approaches that are necessary for understanding social complexity. For example, the “many scales” of social groupings involve a great variety of organizational, temporal, and spatial dimensions, sometimes simultaneously. In addition, computation or computational approaches refer to numerous computer-based instruments, as well as substantive concepts and theories, ranging from information extraction algorithms to computer simulation models. Many more will be invented, given the expansive character of computational tools. In short, CSS involves a vast field of exciting scientific research at the intersection of all social science disciplines, applied computer science, and related disciplines. Later in this chapter we will examine some analogues in other fields of knowledge.

Another useful clarification to keep in mind is that CSS is not limited to Big Data, or to social network analysis, or to social simulation models.¹ That would be a misconception. Nor is CSS defined as any one of these relatively narrower areas. It comprises all of these, as well as other areas of scientific inquiry, as we will preview later in this chapter.

1.2 A Computational Paradigm of Society

Paradigms are significant in science because they define a perspective by orienting inquiry. A paradigm is not really meant to be a theory, at least not in the strict sense of the term. What a paradigm does is provide a particularly useful perspective, a comprehensive worldview (*Weltanschauung*). Computational social science is based on an **information-processing paradigm** of society. This means, most obviously, that information plays a vital role in understanding how social systems and processes operate. In particular, information-processing plays a fundamental role in explaining and understanding social complexity, which is a subtle and deep concept to grasp in CSS as well as in more traditional social science.

The information-processing paradigm of CSS has dual aspects: substantive and methodological. From the *substantive* point of view, this means that CSS uses information-processing as a key ingredient for explaining and understanding how society and human beings within it operate to produce emergent complex systems. As a consequence, this also means that social complexity cannot be understood

¹Big Data refers to large quantities of social raw data that have recently become available through media such as mobile phone calls, text messaging, and other “social media,” remote sensing, video, and audio. Chapter 3 examines CSS approaches relevant to Big Data.

without highlighting human and social processing of information as a fundamental phenomenon. From a *methodological* point of view, the information-processing paradigm points toward computing as a fundamental instrumental approach for modeling and understanding social complexity. This does not mean that other approaches, such as historical, statistical, or mathematical, become irrelevant. On the contrary, computational methods necessarily rely on these earlier approaches—and other methodologies, such as field methods, remote sensing, or visualization analytics—in order to add value in terms of improving our explanations and understanding of social complexity. In subsequent chapters we shall examine many examples pertaining to these ideas. For now, the best way to understand the information-processing paradigm of CSS is simply to view it as a powerful scientific perspective that enables new and deep insights into the nature of the social universe.

1.3 CSS as an Instrument-Enabled Science

CSS is by no means alone in being an *instrument-enabled scientific discipline*. Consider *astronomy*, a science that was largely speculative and slow in developing before the invention of the optical telescope in the early 1600s. What Galileo Galilei and his contemporaries discovered through the use of telescopes enabled astronomy to become a real science in the modern sense. In particular, the optical telescope enabled astronomers to see and seek to explain and understand vast areas of the universe that had been previously unknown: remote moons, planetary rings, sun spots, among the most spectacular discoveries. Centuries later, the radio telescope and infrared sensors each enabled subsequent revolutions in astronomy.

Or, consider *microbiology*, prior to the invention of the microscope in the late 1600s. Medical science was mostly a descriptive discipline filled with untested theories and mysterious diseases that remained unexplained by science. The microscope enabled biologists and other natural scientists, such as Anton von Leeuwenhoek and Louis Pasteur, to observe and explore minuscule universes that were entirely unknown. Later it was discovered that the majority of living species are actually microorganisms. Centuries later, another kind of microscope, the electron microscope, enabled biologists and other scientists to see even smaller scales of life and beyond, down to the molecular and atomic levels. Nano-science was also born as an instrument-enabled field, which also includes an engineering component, as does biology in the form of bioengineering.

Linguistics is a human science that experienced a similar phenomenon, through the application of mathematics. Prior to mathematical and computational linguistics the study of human languages was more like a humanistic discipline, where various interpretations and traditions contended side by side without each generation knowing much more than the previous, since the main tradition was to offer new perspectives on the same phenomena—not exploring and attempting to understand entirely new phenomena. Mathematical and computational linguistics propelled the discipline into the modern science that it is today.

Much the same can be said of *physics*. Greek and medieval scientists viewed the physical universe as consisting of substances with mysterious “essential” properties, such as a heavy object belonging at rest—a state caused by its essence. Physics became a modern, serious science through the application of mathematical instruments, especially the infinitesimal calculus of Newton and Leibniz, in addition to the empirical method. The empirical approach alone would have been insufficient, since theory was enabled by mathematical structures responsible for the main thrust of the hypothetic-deductive method.

What all of these and numerous other cases share in common in the long and well-documented history of science is quite simple: in every culture, science is always enabled and revolutionized by instruments, not just by new concepts, theories, or data. Instruments are the main tools that science uses to create new science. As computers have revolutionized all fields of science since the invention of digital computing machines in the 1950s, and many humanities disciplines in recent years (from the fine arts to history), so the social sciences have been transformed by computing. Moreover, such transformations are irreversible, as has been the case for other instruments in other fields. CSS is in great company; it is not alone in being an instrument-enabled science.

1.4 Examples of CSS Investigations: Pure Scientific Research vs. Applied Policy Analysis

Another stimulating characteristic of CSS is that it encompasses both pure science and policy analysis (applied science). It is not a purely theoretical science such as, for instance, mathematical economics, rational mechanics, or number theory.² This means that CSS seeks fundamental understanding of the social universe for its own sake, as well as for improving the world in which we live. In fact, as we discuss later in this chapter, CSS has a lot to do with improvement of the human condition, with building civilization. These are obviously large claims, but they are not different from those found in other scientific disciplines that attempt to better understand the world both for its own sake and to improve it. It is a misconception to think that pure/basic science and applied/engineering science are somehow opposed or incompatible pursuits. Again, the history of science is replete with synergies at the intersection of pure and applied knowledge. Examples of pure scientific research in CSS include:

1. Investigating the theoretical sensitivity of racial segregation patterns in societies of heterogeneous agents.
2. Modeling how leaderless collective action can emerge in a community of mobile agents with radially distributed, robot-like vision and autonomous decision-making.

²Number theory actually has very concrete application in cryptology, a highly applied field in national security and internet commerce.

3. Understanding how crowds may behave in a crisis when interacting with first responders and their respective support systems.
4. A project on the impact of natural extreme hazards of a generic variety to assess risk and the potential for causing catastrophes and plan for mitigation.
A parallel set of applied policy examples would read more or less as follows:
 1. A high-fidelity agent-based model of New York City neighborhoods to mitigate racial segregation without relying exclusively on laws.
 2. Modeling how the Arab spring may have originated, based on an empirically calibrated social network model of countries in the Middle East and North Africa.
 3. Understanding how the population of New Orleans responded when Hurricane Katrina hit the city and first responders and their respective support systems were activated.
 4. A geospatially referenced agent-based model of the Eastern coast of the United States to prepare for seasonal hurricanes and changing weather patterns caused by climate change.

The use of proper nouns is often (not always!) a give-away in applied policy analysis. However, there is more to applied CSS than the use of proper nouns. In particular, high-quality applied CSS must add value to other policy analysis approaches—it must provide insights or knowledge significantly and demonstrably beyond that which can be provided by other analytical tools. Another distinctive feature of applied CSS analysis is that it contributes to a better understanding of situations that are too complex to analyze by other methods, even when prediction or forecasting is not involved. For example, a good use of applied CSS might be the use of computer simulations to better understand and prepare for unintended consequences—or what are called *negative externalities*—of policies.

The pure-applied synergy in science is also present in CSS in another respect: this has to do with pure research that occasionally generates applications for improving policies, and, conversely, a so-called *wicked problem* in the policy arena inspiring fundamental research questions in pure research. Examples of the former kind of synergy (basic science improving policy) would include:

- Better understanding how crowds of panicky individuals “flow” in an emergency in order to improve building design and evacuation procedures.
- Comparing formal properties of organizational structures to improve the workplace.
- Inventing a new algorithm to improve security of communication in complex infrastructure systems and their management interface with humans.
- Deeper understanding of the formal properties of distributions to design better queuing systems, such as those used by air traffic controllers and similarly complex systems.

Conversely, examples of the latter kind (policy needs informing basic research) would include:

- Developing the social theory of communication in racially mixed communities out of the policy need to create a high-fidelity model of a refugee camp.
- Deepening our understanding of complex network structures based on the need to model transnational organized crime in trafficking of persons.

- Improving a theory of origins of civilization while attempting to improve anti-looting laws and regulations that govern world heritage archaeological sites.
- Working on formulating and testing a new theory of learning in individuals and collectives of agents while trying to revise public policy in health care and education.

The synergies highlighted by these examples are not contrived or invented for pedagogical purposes. They are real in the sense that they have either already occurred, or are likely to occur in the not-so-distant future. In other words, they are not purely notional examples. Moreover, such synergies are likely to grow as the field develops through more mature stages—as has happened in many other areas of science.

The powerful and fascinating synergy between science and policy notwithstanding, it is also fair to say—indeed, be emphasized—that basic scientific research and applied policy analysis are different activities along numerous dimensions, such that they generate different professions:

Expectations: Basic science is expected to produce new knowledge and understanding, whereas applied policy analysis is more results-oriented in a practical sense. People built bridges across rivers centuries (perhaps millennia) before the fundamental laws of mechanics were discovered.

Training: Scientists and practitioners train in different concepts, tools, and methodologies, even when they may share training in some common disciplines, such as in the use of simple statistics.

Incentives: Pure scientists and policy analysts have different incentives, such as academic rewards for the former and promotions to higher organizational roles for the latter.

Facilities: Pure science is best conducted in labs and research centers; think tanks are specialized venues for conducting policy analysis. Both kinds of venues can be academic, private, or governmental; what matters is the main mission and associated support infrastructure.

Publicity: Pure scientific research is most frequently highly publicized, especially when it touches on public issues, such as climate change, health, communication, the economy, or national security. Moreover, open sources are more typical of academic CSS research, except when researchers impose a temporary embargo in order to publish first. Applied policy research is often less public, especially when it concerns sensitive information pertinent to public issues, or when private consulting firms protect intellectual property by requiring and enforcing nondisclosure agreements.

Some features that are common to both pure and applied research in CSS include the need for terminological clarity (not the “Tower of Babel” decried by Giovanni Sartori), systematic concept formation, respect for evidence, rigorous thinking, and thorough documentation. Also, in both areas one can find excellent, mediocre, and outright awful work—“the good, the bad, and the ugly,” as in the proverbial phrase.

Throughout this textbook we will encounter cases of both pure CSS research as well as applied policy applications. Similarities and differences between the two are significant and instructive on the role of each and the synergy between the two orientations or activities.

1.5 Society as a Complex Adaptive System

Society is often said to be complex. What does that mean? In this section we examine this idea for the first time, developing deeper understanding in subsequent chapters.

1.5.1 What Is a CAS in CSS?

At the very beginning of this chapter we mentioned complex adaptive systems as being one of the key, fundamental ideas in the foundations of CSS. For now, we can define a **complex adaptive system** as one that changes its state, including its social structure and processes, in response to changing conditions. Later, especially in Chaps. 5–7, we will develop more rigorous definitions. A cybernetic system is an instance of a rudimentary CAS, whereas a system of government, an ecosystem, an international regulatory agency (such as World Bank or the International Monetary Fund), or a complex organization (such as NASA or the Intergovernmental Panel on Climate Change, IPCC), are more complete examples.³ An essential aspect of this initial definition is to note that a complex adaptive system operates through **phase transitions** (significantly different states and dynamics) in the operating regime of the system in order to maintain overall performance in the face of changing environmental conditions or evolving goals or changes in resources.

A family is a social organization that can be viewed as a complex adaptive system, one based on kinship relations that undergo numerous changes throughout the life cycle of individuals who are members of the family, when viewed as a human grouping. Everyone in the family ages, and some mature successfully into old age, experiencing many different situations, acquiring new knowledge, in the face of numerous opportunities and challenges. In spite of many changes, the overall system of kin-based relations in some families can endure for decades; in other cases that is not the case and the system breaks down. Adaptation in the history of a given family manifests itself in numerous ways: children grow up and must adapt to going to school; parents might change jobs or occupations, having to adapt to labor market conditions or to changing priorities; social mobility also requires adaptation, perhaps to new norms or new locations; making and losing friends also requires adaptation. Adaptation is common and frequent in many social systems because internal components and relations are willing and able, even required, to change in order for the open systems to endure, sometimes improving or prospering.

Adaptation in social systems is best seen as a *multi-stage process*, not as a single event. As such, several occurrences are required for adaptation to operate successfully. We may view this as consisting of several events, which later we will refine in

³The example of a cybernetic system as a CAS is not by chance. In fact, the Greek etymology of the term government, or κυβερνήτης (kybernētēs), means the rudder or steering mechanism in a ship. It's the same in Italian (*governo*), Spanish (*gobierno*), French (*government*), and in other languages.

more formal ways. First, the system, or the actors within the system, must be aware that there is a *need* to adapt—to undertake adaptive behavior. Second, there must be an *intent* to adapt, which is separate from the recognized need to adapt. Third, there must be *capacity* to adapt, since adaptation costs in terms of resources, be they tangible or intangible. Finally, adaptive behavior must be *implemented* in some form, which may involve executing plans or overcoming various kinds of difficulties and challenges. A key idea to understand regarding adaptation in social systems is that it is never automatic or deterministic, at least in the most interesting or nontrivial situations. Whether a person, a family, a group, an economy, an entire society, a whole nation, or even a global society adapts to change, such a process always consists of several stages.

A particularly noteworthy aspect of complex adaptive systems from a *computational perspective* is the key role played by information-processing:

1. Information is necessary for assessing the need for a complex system to require adaptation.
2. The activity of determining resources also requires information.
3. Information flows in the form of interpersonal and inter-group communication when adaptation is decided on, prepared for, implemented, or subsequently monitored for its effects on restoring a viable state for the system.

This is obviously a sparse and simple summary of the role of information in CAS, which serves to highlight the usefulness of the information-processing paradigm discussed earlier. Information-processing is pervasive and critical in complex adaptive systems; it is not a phenomenon of secondary importance. An interesting aspect of information in CAS is that it has many other interesting properties, as well as insightful connections to other essential ideas in CSS, such as complexity, computability, and sustainability, as we will examine later.

1.5.2 Tripartite Ontology of Natural, Human, and Artificial Systems

Another important distinction in CSS is among natural, human, and artificial systems—an ontological or categorical distinction that is different or does not exist at all, at least not to the same degree, in other fields of knowledge. The first computational social scientist to introduce this idea of a tripartite classification of entities was Herbert A. Simon, who used it as foundation for his theory of artifacts and social complexity through the process of adaptation. We will examine this soon, but the tripartite distinction is needed now. Complex adaptive systems of interest in CSS often combine all three categories of systems, so understanding the composition of each, as well as their similarities and differences, is important before entering more theoretical territory.

1. A **natural system** consists of biophysical entities and dynamics that exist in nature, mainly or completely independent of humans and their artifacts. Common examples are wilderness landscapes, animals other than humans, regional ecosys-

tems, and the biochemistry of life, including the biology of the human brain as a natural organ (*not* just mental phenomena).⁴

2. A **human system** is an individual person, complete with thoughts and body. Decision-makers, actors, agents, people, and similar terms denote human systems. The complexity-theoretic perspective highlights the human ability to create artifacts.
3. An **artificial system** is one conceived, designed, built, and maintained by humans. Artificial systems consist of engineered or social structures that act as adaptive buffers between humans and nature.

These initial conceptual definitions serve as building blocks that for now are sufficient for our initial purpose of establishing foundations. We shall return to these ideas to develop a better understanding of their properties and interrelationships.

1.5.3 Simon's Theory of Artifacts: Explaining Basic Social Complexity

Laws describe; theories explain. Having presented and discussed the first conceptual building blocks, now our main task is to move forward by providing an initial statement of Herbert A. Simon's theory of artifacts for providing an initial explanation of social complexity. Simon presented most of these ideas in his classic monograph, *The Sciences of the Artificial*, which first appeared in 1969, followed by a third and last edition in 1996.

From the previous ideas, it is important to note that artifacts exist because they have a function: they serve as adaptive buffers between humans and nature. This is the essence of **Simon's theory of artifacts and social complexity**. Humans encounter challenging and often complex environments, relative to their own simple abilities or capacities. In order to adapt to these circumstances, and not be overwhelmed by or succumb to them, humans pursue the strategy of building artifacts that enable their goals.

- Roads were first invented for moving armies and other military and political personnel from one location to another. They were also used for commercial and communications purposes. Without a proper road it is either very difficult or impossible to achieve such goals.
- Bureaucratic systems, and in some cases writing (e.g., Mesopotamia, China), were first created for maintaining records related to the governance and economy of a city. This enabled the first urban populations to attain the goals of becoming established and developed.

⁴The wording here is intentionally and necessarily cautious and precise. The paradigm being presented here separates humans from the rest of nature, based on the human ability to build artifacts, some of which are used to build other artifacts, especially intelligent, autonomous artifacts, using mental, cognitive, and information-processing abilities that are far more complex than those found in any other natural living organism. Ants might build colonies, corals build reefs, bees build hives, beavers build dams, but none of these or other examples of "animal-made artifacts" compares to human artifacts.

- The first large aqueducts, built by the Romans, required careful planning, engineering, and maintenance in order to provide water for large urban populations located at great distances from the sources (springs, rivers, lakes, or reservoirs).
- The International Space Station (ISS) is an engineering structure of unprecedented complexity, operating in the challenging environment of space, managed by a ground crew in coordination with the station's crew.

As already suggested by the previous examples, the artifacts that humans have been building for thousands of years, across all societies, can be **tangible** (engineered, i.e., physical) or **intangible** (organizational, i.e., social), as required by the goals being sought. Some adaptive strategies require tangible, engineered artifacts, such as dwellings, bridges, roads, and various kinds of physical infrastructure systems. At other times, an adaptive strategy may require planning for and creating an organization, such as a governing board or committee, that is to say, a social system of a given size and complexity to enable attainment of the goal being pursued.⁵

A fascinating aspect of this tightly coupled *synergy* between tangible and intangible, or engineered and organizational artificial systems, is that they often require each other—as in a *symbiotic relationship between humans and their artifacts*, where the latter enable human attainment of desired goals. This feature of social complexity is supported by historical and contemporary observation. To build a road or a bridge it is also necessary to create teams of workers supervised by managers, who depend on supply chains for the provision of building materials and other necessities: the tangible artifact (bridge) cannot be built without the intangible one (organization). Modern cities provide another excellent example of the same symbiotic relationship between engineered and social artifacts. The complex infrastructure that supports the life of humans in cities (as opposed to cave dwellers) requires numerous, specialized buildings and artificial systems—especially when cities are built in mostly inhospitable environments. This was also true of the earliest cities, which were supported by an organizational bureaucracy of managers, city workers, and other social components, working in tandem as a coupled socio-technological system to support urban life. For example, the capital of the USA, Washington, is built on a swamp, as is the Italian city of Venice. Both are enabled by physical and organizational infrastructure.

In sum, what does Simon's theory *explain*? It explains *why* artifacts exist, *why* humans build artifacts, and the fact that artifacts are *adaptive strategic responses* for solving the many challenges faced by humans in societies everywhere since the dawn of civilization.⁶

⁵This idea prompted Simon to suggest—in *The Sciences of the Artificial*—that social scientists, lawyers, and engineers should undergo university-level training of a similar kind, perhaps under a common College of the Artificial Sciences.

⁶Herbert A. Simon's work in the social sciences is widely known for its contributions to the study of organizations and bureaucracy. In computer science his work is equally well known for contributions to artificial intelligence and related areas. His theory of social complexity grew out of an interdisciplinary interest across these domains.

1.5.4 Civilization, Complexity, and Quality of Life: Role of Artificial Systems

Simon's theory of artifacts and adaptation goes a long way toward explaining the genesis and development of social complexity. It also explains important aspects of the same patterns that endure to this very day and will likely continue into the future. Humans everywhere pursue goals that are often sought in challenging environments, so in order to accomplish those goals they build artifacts—both engineered and social systems that are tangible and intangible, respectively.

However, thus far the story is incomplete, because sometimes humans seek goals that are *not* necessarily linked to challenging environments. For example, they may already live in a city that is quite viable, but they simply wish to live in a better way, such as enjoying better services and amenities, living longer or more comfortably, or enjoying culture and the fine arts. An additional, essential ingredient for developing a more complete theory of social complexity, one that explains a broader range of social complexity, is based on the empirical observation that *humans everywhere prefer to live a better life*. This is also a purpose of government: "The care of human life and happiness, and not their destruction, is the first and only legitimate object of good government" (Thomas Jefferson, American President, 1809).

A significant variation on the very same theme would be, for example, to wish that their descendants or friends enjoy a higher quality of life. The pursuit of a higher quality of life is a goal for many humans, which may occur independent of or in combination with taming a given environment. The strategic adaptive response is the same or isomorphic: artificial systems are conceived, planned, built, and maintained in the form of physical or social constructs. Complexity in all these forms increases in each case. Therefore, *both* challenging environments *and* human aspirations—and quite frequently the interaction of both—cause social complexity in a generative sense.

Sometimes complex systems come and go in a transient way; at other times they become permanent artifacts that can endure for very long periods of human history. Systems of government, infrastructure systems, monetary systems, and cultural norms provide examples of long-term artifacts that have increased in complexity over the millennia. Civilization is the result of this process, from the theoretical perspective of CSS. The dawn of civilization in all parts of the world where humans have created and developed social complexity is marked by the earliest engineered and organizational artifacts. Contemporary civilization in the 21st century is no different from the earliest civilizations, as seen from this universal theoretical perspective. Societies in the earliest days of Mesopotamia, China, South America, and Mesoamerica built the first irrigation canals, structures for communal worship, villages, towns and cities, the earliest infrastructure systems and systems of government and bureaucracies that supported them. All these artificial systems and many others that have since been invented persist to this day, and spacefaring civilization—if we manage to launch and mature it—will demonstrate comparable patterns in the evolution of social complexity.

Information-processing, goal-seeking behavior, adaptation, artifacts—engineered as well as organizational—and the resulting social complexity that they cause are the main ingredients of this interdisciplinary theory. Its purpose is to explain how and why natural, human, and artificial systems interact in the creation of history. The theory is *causal*, in a strict scientific sense, because it proposes an empirically demonstrable process that links together—not in a superficial correlational way devoid of causation—the elements thus far presented in this chapter and examined in greater detail across areas of CSS.

1.6 Main Areas of CSS: An Overview

Computational social science is an interdisciplinary field composed of areas of concentration in terms of clusters of concepts, principles, theories, and research methods. Each area is important for its own sake, because each represents fertile terrain for conducting scientific inquiry, as basic science as well as policy analysis. In addition, these areas can build on each other and be used synergistically, as when network models of social complexity are used in simulation studies, or through many other possible combinations of scientific interest.

The chapters of this book are dedicated to each of these areas, which we will now survey by way of introduction. The main purpose in this section is to provide an overview, not a detailed presentation of each area. By way of overview, it should be mentioned that these areas of CSS are also supported by statistical and mathematical approaches, and in some cases other methodologies as well, such as geospatial methods, visualization analytics, and other computational fields that are valuable for understanding social complexity.

1.6.1 Automated Social Information Extraction

CSS is an interdisciplinary field where data play numerous and significant roles, similar to those in other sciences. The area of automated information extraction refers to computational ideas and methodologies pertaining to the creation of scientifically useful social information based on raw data sources—all of which used to be done manually. Other names for this area of CSS might be computational content analysis, social data analytics, or socio-informatics, in a broad sense. For example, whereas in an earlier generation social scientists would gather data from sources such as census records, historical sources, radio broadcasts, or newspapers and other publications, today much of the work that takes place in order to generate social science research data is carried out by means of computational tools. As we will see, these tools consist of computational algorithms and related procedures for generating information on many kinds of social, behavioral, or economic patterns.

Social information extracted through automated computational procedures has dual use in CSS. For instance, sometimes it is used for its own sake, such as for analyzing the content of data sources in terms of affect, activity, or some other set

of dimensions of interest to the researcher. An example of this would be a study to extract information concerning the political orientation of leaders or other governmental actors based on computational content analysis of speeches, testimony before legislative committees, or other public records.

Besides being used for analyzing the direct content of documents and other sources, information extraction algorithms can also be used to model networks and other structures present in raw data, but impossible to detect through manual procedures performed by humans. An example of this would be a model of organized crime organizations and their illegal activities, based on computational content analysis and text mining of court cases and other evidentiary legal documents that describe individuals, dates, locations, events, and attributes associated with criminal individuals. Another example would be automated information extraction applied to modeling correlations across networks, based on Internet news websites.

An extension of automated information extraction could also be used for building computer simulation models that require high fidelity calibration of parameters, such as models of opinion dynamics, international trade, regional conflicts, or humanitarian crises scenarios. The extraction of geospatial social data through computational algorithms represents a significant step forward in the development of CSS.

These and other examples illustrate how automated information extraction is sometimes seen as a foundational methodology in CSS: it can be used for developing models and theories in all of the other main areas of CSS, besides its intrinsic value.

1.6.2 Social Networks

Social network analysis is another major area of CSS, given the prominence of networks of many types in the study of social complexity. This area has become very popular in recent years, especially through the development of social media and Internet websites such as Facebook, Twitter, and numerous others. However, the analysis of networks in just about every domain across the social sciences—certainly in all the Big Five disciplines—predates computing by many years, so we should be examining the area of social network analysis from its historical roots. Social network analysis is the only area of CSS that has a well-documented history (Freeman 2004).

The advent of digital computing and CSS has transformed the study of social complexity through network analysis and modeling, expanding the frontiers of research at an unprecedented rate while advancing our understanding along many fronts in this area. There are numerous reasons for the exciting progress that this area is experiencing. For one, based on decades of pioneering research on networks, by the time computers became part of their methodological toolkit, social scientists had already developed a powerful set of concepts, statistical tools, and mathematical models and procedures, including formal theories, which enabled them to exploit computational approaches. Another reason for the explosion of progress on theory

and research in this area of CSS is that computational tools, especially the most recent generation of computer hardware and software systems, now enable efficient processing of high-dimensionality data and large matrices necessary for understanding complex social networks.

Social network analysis has intrinsic value, and it also contributes to the other areas of CSS theory and research. We shall examine examples of these synergies, but before that it is necessary to gain familiarization with basic concepts, theories, and research methods in this area—*almost* as if it had no applications in other areas of CSS!

1.6.3 Social Complexity

In this introductory chapter we have already previewed some initial ideas for understanding social complexity, because this is such a defining, foundational theme for CSS. However, there is much more to understanding social complexity and its many exciting scientific and policy implications, besides the preliminary introduction that has been provided thus far. For example, research in the area also requires an understanding of **origins of social complexity** in regions where the earliest civilizations emerged, and their subsequent, long-range historical development. The study of origins of social complexity should be seen in much the same way as a science course in astronomy examines the *cosmology* of the physical universe, in terms of how the physical universe originated and how and why the earliest structures and systems emerged—the formation of stars, planets, moons, planetary systems, galaxies, and clusters of galaxies that span the cosmos. Traditionally—and perhaps not so surprisingly, given the standard (read: “turf-based”) territorial disciplinary divisions of academic labor—most, albeit not all, of the study on origins of social complexity has been conducted by a relatively small community of archaeologists, mostly working in isolation from other social scientists. However, this is changing and CSS is playing an increasingly significant role in our scientific understanding of the origins of social complexity and civilizations.

In addition to understanding the origins of social complexity—just as astronomers are familiar with cosmology *and* contemporary theories and research for understanding the current universe—in this area of CSS it is also essential to develop a better understanding of interdisciplinary **concepts and theories of social complexity**. For example, whereas concepts such as information-processing, adaptation, and socio-technical artifacts provide some explanation of the phenomenon, CSS theory draws upon a broad array of other social science concepts, such as decision-making, coalition theories, collective action, and others. The **Canonical Theory** of social complexity provides a formal and empirically valid framework for describing, explaining, and understanding social complexity origins and development. Moreover, CSS investigation of social complexity also includes key concepts from **complexity science**, including the theory of **non-equilibrium distributions**, **power laws**, **information science**, and related ideas in contemporary science. This is another highly interdisciplinary area of CSS, bringing together quantitative and

computational social scientists, as well as ideas and methods from other disciplines across the physical, geospatial, and environmental sciences.

1.6.4 Social Simulation Modeling

The CSS area of **social simulation modeling** can be characterized as foundational, multi- as well as inter-disciplinary, and diverse, meaning it is based on many different methodologies in modeling and simulation disciplines. The area is increasingly significant and mature for conducting both basic science and applied policy analysis. Like social network analysis, this area is sometimes confused with the totality of CSS, whereas it is only an area, not the whole field of CSS.

The simulation modeling tradition began in social science many decades ago, during the earliest days of digital computing. There are several different kinds of social simulation modeling frameworks, as we shall discuss. Regardless of the specific type, all social simulation models share a set of common characteristics. Every simulation model is always designed and built around a set of *research questions*, which may concern basic science or applied policy analysis, sometimes both. Research questions provide essential guidance for simulation models, just as in other models (for example, in formal mathematical models). Another characteristic shared by social simulation models is that they are developed through a set of *developmental stages*, not as a single methodological activity, especially in the case of complex modeling projects or those involving teams of investigators. Such stages include model verification and validation, among others. In addition, specific types of models often require additional stages in their development. It should be pointed out that each of the social simulation modeling traditions is sufficiently large to include specialized journals, conferences, and other institutional components in communities of practitioners that often number in the thousands of researchers.

The earliest kind of simulation models in CSS are the **system dynamics models**, which gained highly significant international notoriety through the global models of the Club of Rome in the 1960s and 1970s.⁷ These social simulations built on the pioneering work of Jay Forrester and his group at MIT. From a computational perspective, these are equation-based models that employ systems of difference equations or systems of differential equations, as the situation and data might require. This class of models has been very significant for many decades—indeed, for half a century—because so many social systems and processes are properly amenable to representation in terms of stocks and flows, or levels and rates, respectively. Arms races, stockpile inventories in business enterprises, the dynamics of economic development, and numerous other domains of pure and applied analysis have been modeled through system dynamics simulations. A significant feature of theory and

⁷The Club of Rome is an international non-governmental organization founded in 1968 and dedicated to scientific analysis of the future and sustainable development.

research in system dynamics simulation models has been the availability of excellent software support systems, such as Forrester's DYNAMO, followed by the Stella system, and presently Vensim.

Another major tradition in social simulation models is represented by **queuing models**. As their name indicates, these models are used for social systems and processes where lines or queues of entities (such as customers, patients, guests, or other actors) are "serviced" by various kinds of stations or processing units. Banks, markets, transportation stations of all kinds, and similar systems that provide a variety of services are some examples. From a formal and computational point of view, these models are based on queuing theory, and various kinds of probability distributions are used to represent the arrival of entities at service stations, how long the service might take, and other statistical and probabilistic features of these processes. Hence, queuing models also belong to the class of **equation-based models**.

By contrast, the following kinds of social simulation models move towards the **object-based orientation** of modeling and simulation, rather than the equation-based paradigm. Of course, this is not to say that object-based models are devoid of equations; it simply means that the building blocks of this other class of models are object-like, as classes or entities. Their variables and equations are said to be "encapsulated" within the objects themselves.

The simplest kinds of object-based social simulation models are **cellular automata**, which generally consist of a grid or landscape of sites adjacent to one another, as in a checkerboard. The actual shape of the sites or cells can take on many different forms, square, hexagonal, or triangular cells being the most commonly used. The earliest work in cellular automata was pioneered by John von Neumann, who also invented game theory. The basic idea of social simulations based on cellular automata is to study emergent patterns based on purely local interactions that take place between neighboring cells on a given landscape. One of the most important and well-known applications of this kind of model has been the study of racial segregation in cities and neighborhoods, showing how segregation can emerge even among relatively unprejudiced neighbors.

Another major class of social simulation models is represented by **agent-based models**, often abbreviated as ABMs.⁸ In this case the actors being simulated enjoy considerable autonomy, specifically decision-making autonomy, often including physical movement from one place to another, which is why they have had so much success in modeling social systems and processes having a geospatial dimension. Agent-based models can be **spatial** or **organizational**, or both combined, depending on what is being represented in the model. Spatial agent-based models can also use a variety of data for representing landscapes, such as GIS (Geographic Information Systems) or remote sensing data. Organization agent-based models are akin to dynamic social networks, where nodes represent agents and links represent various kinds of social relations that interact and evolve over time. These kinds of social simulation models have become increasingly significant for solving theoretical and research problems that require representation of heterogeneous actors and a

⁸The computer science terminology for these models is multi-agent systems, or MAS.

spectrum of interaction dynamics that are simply intractable through mathematical approaches that require closed-form solutions. They are also particularly appealing for investigation of emergent patterns indicative of complex adaptive systems. For example, a significant application of agent-based models is the study of complex crises and emergencies, given their ability to represent human communities in environments prone to natural, technological, or anthropogenic hazards. In another important application, as we shall see, agent-based models provide the first viable methodology for modeling entire societies, politics, and economies, as well as national, regional, and global scales of these social systems.

Finally, **evolutionary computation models** represent the class of social simulations based on notions and principles from Darwinian evolution, such as evolutionary algorithms. Although evolutionary computation models are still relatively new in CSS, they already have shown great promise. For example, they allow us to derive patterns of social dynamics that are not well understood, so long as the simulation model can be made to match empirical data. This use of evolutionary models in a "discovery mode" is characteristic of this particular kind of simulation.

Each of the preceding types of social simulation models can, at least in principle, include ideas and components from other areas of CSS, such as results from automated information extraction, social network analysis, complexity-theoretic ideas, and the like. Conversely, social simulation models can provide significant input and improvements pertinent to research in these other areas.

This brief survey of simulation models in CSS covers most of the areas that have been developed during recent decades. No doubt other social simulation methodologies will emerge in the future, either as outgrowths of current modeling approaches (as agent-based models originated from cellular automata models) or as novel inventions to analyze problems or investigate research questions that remain intractable by the current types of simulation models.

1.7 A Brief History of CSS

Each of the areas of CSS that we have introduced in this chapter has its own, more detailed, history, the main highlights of which are provided in each of the chapters to follow. The purpose in this section is to provide an overall, albeit brief, history of the entire field of CSS, beginning with its historical roots.

How, when, why, and who began the field of CSS as a systematic area of inquiry is similar in some respects to the history of other scientific fields. The historical origins of CSS are to be found in the Scientific Revolution that occurred in Europe during the late Renaissance and early Enlightenment periods. This was the epoch when the social sciences began to adapt universally held concepts and principles of positive scientific methodology (not just particular quantitative methods, such as statistics), specifically with regard to measurement of observations, systematic testing of hypotheses, and development of formal mathematical theories for explaining and understanding social phenomena. Human decision-making and voting behavior (i.e., the foundations of social choice theory) were among the earliest areas of

inquiry. Statistics, initially intended to be the scientific discipline to study the state and improve policy analysis, was also invented during this period. Statistical and mathematical methods were introduced throughout the 18th and the 19th century by famous luminaries such as Denise Poisson, Adolphe Quételet, William Petty, Daniel Bernoulli, Pierre de Fermat, Jean Marie de Condorcet, Corrado Gini, and Vilfredo Pareto, among many others. The most important result of this formative period in the history of the social sciences was the adoption of a scientific culture concerning the quest for knowledge and understanding, a tradition that endures to this day.

For our purposes it is useful to mark the beginnings of CSS, in a strict sense, with the invention of digital computing during the closing days of World War II and the early days of the Cold War. This major milestone in the world history of science and technology affected the social sciences in two transformative ways, each of which is interesting in its own right. First, the modern digital computer enabled the emergence of CSS by providing the key instrument that would fuel and expand its research horizons in a way that would have seemed unimaginable just a few years earlier. For the first time social scientists were able to analyze vast amounts of data, test many novel scientific hypotheses, and explore the dimensions and structures of social space—from the human mind to the global system, with numerous levels of analysis in between. An early example of this was the invention of factor analysis—a powerful inductive, dimensionality-reduction methodology that led to many discoveries across the social sciences—by early CSS pioneers such as Charles Spearman, Rudolf Rummel, and L. Thurnstone. Among these was the discovery of the dimensionality of human cognitive spaces, as well as the structure of spaces wherein international interactions occur. Yet another example was the invention of the General Inquirer, a computational content analysis system that allowed social researchers for the first time to explore and test hypotheses concerning the content of an unprecedented volume of qualitative text data. Within the span of a single generation the volume of knowledge across the social sciences increased by many orders of magnitude thanks to the advent of the modern digital computer.

The second truly major, transformative way in which the modern digital computer affected the social sciences was as an inspiring metaphor that shed new light on classical and modern areas of investigation. Social scientists had known for some time the significance of communication and information-processing for understanding human and social dynamics. For example, the study of media and text data, as well as radio broadcasts and propaganda, had begun in earnest many decades before the advent of the computer. However, the digital computer inspired new concepts, hypotheses, principles, models, and theories about the vast array of systems and processes in the social universe. For instance, political scientists who became familiar with ideas from cybernetics and general systems theory (new fields pioneered by scientists such as W. Ross Ashby, Norbert Wiener, Ludwig von Bertalanfy, and Anatol Rapoport, among others) began viewing the structure and functioning of politics and other forms of political systems by highlighting the role of information-processing, goal-seeking behavior, social computing, and emergent phenomena. An example of this was the novel cybernetic theory of government formulated by Karl W. Deutsch and others, who played a leading role during the Behavioral Revolution

of the 1960s. A polity, as we will see in subsequent chapters, can be described and understood as a complex adaptive system that carries out numerous, coordinated computations, such as voting and policymaking. Herbert A. Simon's theory of social complexity through adaptation and artifacts—published for the first time in the 1969 edition of *The Sciences of the Artificial*—was another result of the influence of digital computing machines. Harold Guetzkow developed innovative computer simulation approaches, as well as hybrid simulations (so-called man-machine simulations) that are still highly influential to this day. 1969 was also a seminal year in which Hayward Alker and Ron Brunner published the first paper on comparative simulation research.

All areas of CSS have experienced remarkable growth since the early days of the field. Progress in social theory and research, as well as remarkable advances in all areas of computing, particularly applied computational approaches and methodologies, have contributed to the current body of knowledge in CSS. Today CSS is also beginning to reap the benefits of interactions and synergies among its main areas, as they fertilize and stimulate each other each other in new and exciting ways. For example, the early history of social network analysis, or even automated information extraction, developed in relative isolation or autonomy—by endogenous development. Today, by contrast, these areas experience frequent overlays and mutually beneficial collaborations, as witnessed by the application of text-mining algorithms to populate social network models. Another example is the application of network models to improve the specification of social structures represented in agent-based models for the study of emergence in complex social systems. The history of CSS as an emergent field is still in its infancy. However, the field has already demonstrated significant capacity and promise for contributing to new understanding across all areas of social science theory and research.

1.8 Main Learning Objectives

This textbook has a set of main learning objectives intended to be pedagogically appropriate as an introduction to the field of CSS. As indicated in the preface, these objectives include learning basic concepts, models, theories, and methodologies used in CSS. These objectives are designed to serve two purposes: a basic exposure to the field of CSS, as well as building foundations for further study at more advanced levels.

The following scientific learning objectives are among the most important. Examples are provided as illustrations.

- Basic understanding of key CSS *concepts*, including all those highlighted in boldface and included in the Index, to a level where the reader can provide additional examples. Conceptual proficiency is fundamental, including concept formation in CSS.
- Familiarization with the *scope and content of each area* of CSS, grounded in elements of computing, including areas of automated information extraction, social networks, complexity-theoretic understanding of social systems and processes,

and various kinds of social simulations. Examples include complex adaptive systems, coupled systems, multi-scale processes, bifurcation, criticality, metastability, phase transitions, autonomous agents, verification, and validation.

- Understanding of *main theories* that are part of the CSS paradigm as causal explanatory frameworks that shed new light on the nature of human and social dynamics. Examples include Simon's Theory of Artifacts, the Canonical Theory of Social Complexity, the Theory of Social Networks, the Theory of Non-equilibrium Social Processes, and others.
- Ability to distinguish and analyze the different *levels of analysis* of social complexity using computational approaches, ranging from mental phenomena to decision-making, social groups and their interactions, to the global system.
- Ability to work with one or more of the *methodological tools* covered in one or more of the chapters. Examples include extracting entities from text data, computing social network indices, testing a power law hypothesis, and building a basic agent-based model in a programming language such as Python or a simulation toolkit such as Netlogo.
- Familiarization with the *main classes of entities, objects, and relations* that are most common in computational analyses of social complexity. Examples include various types of actors, associations, attributes, and methods.
- Proficiency in the *interdisciplinary integration of knowledge* in the context of social phenomena, including the synergistic nexus between social science and computational methodologies.
- Basic knowledge of the *history of each area* of CSS, including prominent pioneers, with an understanding of roots in early development of the social sciences and computer science, at least to the level detailed in the brief histories provided in each chapter.

This minimal set of learning objectives applies throughout chapters in this textbook, ideally independent of the content of each area of CSS. In addition, each chapter contains its own set of main learning objectives that are more specific to the scope and content of each area.

Motivated readers will benefit from further study of the supplementary reading materials provided at the end of each chapter under the heading of Recommended Readings. These are intended to provide more advanced foundations and knowledge that extends beyond the scope of this introductory textbook. The bibliography contains additional sources that interested readers will wish to look up, both early classic literature in CSS, as well as some of the most current and influential contributions.

Recommended Readings

- H.R. Alker Jr., R.D. Brunner, Simulating international conflict: a comparison of three approaches. *International Studies Quarterly* 13(1), 70–110 (1969)
- H.R. Bernard, The science in social science. *Proceedings of the National Academy of Science* 109(51), 20796–20799 (2012)

- C. Cioffi-Revilla, Computational social science. *Wiley Interdisciplinary Reviews (WIREs): Computational Statistics*, paper no. 2. Available online (2010)
- D. Collier, J. Gerring, *Concepts and Method in Social Science: The Tradition of Giovanni Sartori* (Routledge, New York, 2009)
- R. Conte, G.N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffaunt, J. Kertesz, D. Helbig, Manifesto of computational social science. *European Physical Journal Special Topics* 214, 325–346 (2012)
- F. Fernandez-Armesto, *Civilizations: Culture, Ambition, and the Transformation of Nature* (Simon & Schuster, New York, 2001)
- A.M. Greenberg, W.G. Kennedy, N.D. Bos (eds.), *Social Computing, Behavioral-Cultural Modeling and Prediction* (Springer, Berlin, 2012)
- J.H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975)
- I.L. Horowitz, Big Five and Little Five: measuring revolutions in social science. *Society* 43(3), 9–12 (2006)
- M. Kline, *Mathematics and the Search for Knowledge* (Oxford University Press, Oxford, 1985)
- J.H. Miller, E. Page Scott, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton University Press, Princeton, 2007)
- H.A. Simon, *The Sciences of the Artificial*, 3rd edn. (MIT Press, Cambridge, 1996)
- L. Spinney, History as science. *Nature* 488, 24–26 (2012)
- M. Steuer, *The Scientific Study of Society* (Kluwer Academic, Dordrecht, 2003)
- C. Williford, C. Henry, A. Friedlander (eds.), *One Culture: Computationally Intensive Research in the Humanities and Social Sciences—A Report on the Experiences of First Respondents to the Digging into Data Challenge* (Council on Library and Information Resources, Washington, 2012)