Introduction

What is an artificial economy? It is a computational representation of an economic system, which allows us to simulate the interaction of artificial agents.

Artificial agents are the basic units that make up an artificial economy. These agents are computational objects containing information and rules for processing it. They can deploy very simple and silly behavior, or display sophisticated forms of artificial intelligence.

Artificial agents' characteristics and rules of behavior, their forms of interaction, and their spatial and temporal environments, are programmed and simulated computationally. From these simulations, economic structures and dynamics arise. In turn, these structures and dynamics can provoke changes in agents, affecting both their characteristics and their behavioral rules. These processes may converge, through disequilibrium situations, toward an equilibrium, or not. And they may lead to evolutionary and complex dynamics.

The methodology, assumptions and models of artificial economics (AE) contrast in at least four aspects with those of mainstream economics (ME) (a brief on the theoretical core of ME is presented in Annex B).

First, there is a contrast at the *methodological* and *instrumental* levels. ME is an essentially *mathematical* discipline, while AE is fundamentally *computational*. In ME, an economic system is usually represented by a system of equations, which in turn are obtained from the constrained optimization of functions. Therefore, the mathematical techniques of static and dynamic optimization, and static and dynamic equation systems, are among its main instruments. In AE, an economic system is programmed on a computer and its dynamic is simulated by the sequential application of its most basic rules of behavior. Thus, AE basic instruments are algorithms, software, and computer hardware.[1]

Second, there is a difference in how *individual agents* are conceptualized. ME builds its models assuming an agent endowed with *full rationality*, which acts as if it possesses a limitless capacity for gathering and processing information to maximize its well-being. This is the opposite of the artificial agent with which AE works, an agent with limited information and computational capacity and therefore with *bounded rationality*.[2]

Third, there is a contrast regarding *interactions between agents*. ME mostly assumes *global interactions*: each agent can interact simultaneously with all other agents in the economy. AE mostly works with *local interactions*: each agent only deals with other agents in its immediate environment.

Fourth, there is a difference in the *dynamics* of the economic systems being analyzed. ME tends to focus on economic *equilibrium* situations, those where the economy reaches a point at which it reproduces itself with no changes, or when it expands regularly. While AE tends to focus on the study of *disequilibrium* situations, often – although not necessarily always – evolutionary and complex.

While various aspects of the theoretical core of ME have long been questioned (e.g., criticisms coming from Marxian political economy or from Keynesian economics), in recent years there have been important developments leading to the challenge of one of its central pillars: the assumption of full rationality. This challenge comes from three new subdisciplines in economics: behavioral economics, neuroeconomics, and experimental economics. But none of these subdisciplines, so far, is a global alternative to the ME paradigm.

In this context, AE is of interest because, while for some researchers AE is complementary to ME, for others it is a globally alternative paradigm to the conventional one both in its basic concepts and its methodology. And because, like ME, it easily transcends the boundaries of economics to extend to other social disciplines such as demography, sociology, and political science. In other words, because AE may not only challenge ME

[1] As we will see in detail in Chapter 4, the computational nature of AE is not strictly due to its implementation on a computer, but to its algorithmic nature, which is not opposed to mathematics in general, but is linked to a special class of mathematics.

[2] It is argued that ME only assumes that an agent acts rationally if it behaves consistently: its preferences are complete and transitive, and no more than this. However, as we will see later, most typical ME models assume that agents display perfect forward-looking behavior in deterministic contexts, or rational expectations in stochastic environments.

within the economic realm, but also its "imperialist expansion" into other social sciences.[3]

There are other names that overlap with AE, depending on the methods and models used and the dynamics that emerge from simulations. According to the methods used, AE is seen as a part of the broader field of computational economics, which in turn includes computational methods to numerically solve typical ME models such as computable general equilibrium models, econometric models, and others. According to the models developed, mostly models based on agents or on cellular automata, AE is often referred to as agent-based computational economics, heterogeneous interacting agent modeling, or cellular automata based modeling. Finally, according to the type of dynamics generated, often of the evolutionary or complex type, a good deal of AE works is classified as part of the fields of evolutionary economics, and complexity economics.[4]

[3] "Economics imperialism" means the application of ME assumptions and methods to all or almost all areas of human behavior, including social and political actions; decisions on health, education and family; addictions; science; religion, etc. For a presentation of the concept, see Stigler (1984). For a critical discussion and appraisal, see Fine and Milonakys (2009).

[4] For information about the field of computational economics see the website of the Society for Computational Economics at http://comp-econ.org. The website www.artificial-economics.org contains information about annual AE conferences. For a broad view and for information on agent-based computational economics and on heterogeneous interacting agent modeling, see Tesfatsion and Judd (2006), Hommes and LeBaron (2018), the website created and maintained by Leigh Tesfatsion (a pioneer and great promoter of agent-based computational economics) at www2.econ.iastate.edu/tesfatsi/ace.htm, and the website of the Society for Economic Science with Heterogeneous Interacting Agents at https://sites.google.com/view/eshia-site. For introductory books to agent-based modeling, see Damaceanu (2013), Boero et al. (2015), Wilensky and Rand (2015), Hamill and Gilbert (2016), Gilbert (2019), Railsback and Grimm (2019), and Laver (2020). And for more advanced introductions, see Chen (2016) and Delli Gatti et al. (2018). For an introductory overview, and for arguments for modeling complex social dynamics with cellular automata, see Flache and Hegselmann (1998). For arguments for, and examples of, economic modeling with cellular automata, see Albin (1998). And for theory and applications (including economic ones) of cellular automata, see Him, Wu, and Li (2018). Some of the main academic journals that publish AE research are: *Journal of Artificial Societies and Social Simulation*, *Journal of Economic Interaction and Coordination*, *Computational Economics*, *Journal of Economic Dynamics and Control*, *Complexity*, and *Advances in Complex Systems*.

PART I

ARTIFICIAL ECONOMICS AND MAINSTREAM ECONOMICS

This first part introduces the concept of the artificial agent, and presents models of markets and games in which artificial agents interact. Each chapter contrasts concepts and models of AE against similar concepts and models of ME, emphasizing the differences between bounded rationality and full rationality at the agent level; between disequilibrium and equilibrium in markets and games; and between computational methods and mathematical methods.

Chapter 1, on the artificial agent, introduces the concept of the artificial agent, presents the main models of mental architectures that derive from cognitive science, and introduces some recent advances in neuroscience useful to build realistic models of artificial agents. Chapter 2, on artificial markets, presents a model of market interaction between artificial agents, and contrasts it against general static and dynamic equilibrium models typical of ME. Chapter 3, on artificial games, contrasts an artificial evolutionary game against a classic game typical of ME. Finally, Chapter 4 presents a discussion of the differences, as well as possible complementarities, between the computational methodology of AE, and the mathematical methodology of ME.

1

The Artificial Agent

The concept of *agent* has different meanings depending on the context in which it is used. In economics and in the social sciences, an agent is an individual with the capacity to act within his or her economic and social universe. More specifically, an agent is an individual who can initiate, perform, and control its actions, in order to achieve some goals.

In this chapter, we look at how the concept of economic agent developed historically within ME, and how the concept of artificial agent emerged while cognitive science became the successor of behaviorism. Then, considering that AE tries to build realistic models of artificial agents, we introduce the main models of mental architectures that derive from cognitive science, and some recent advances in neuroscience (specially within neuroeconomics, social neuroscience, and neurosociology) that relate directly to the economic and social behavior of individuals. Finally, we review some models and approaches that try to capture the cognitive, neurological, emotional, and social aspects of agents in an integrated way.

1.1 The Agent in Mainstream Economics

From the point of view of ME, an economic agent is conceived and formalized as a rational entity specialized in obtaining the greatest possible wellbeing, or utility, from each of its actions, given its preferences and the constraints within which it must operate. Specifically, an economic agent can be, for example, a consumer who, given prices of the goods she wants to consume, evaluates the quantities she could buy with her income, and chooses the best combination in order to obtain the greatest possible utility.

At the dawn of economic science, and especially from the work of utilitarian philosopher Jeremy Bentham at the end of the eighteenth century, the concept of utility was based on the seemingly obvious psychological assumption that the human being was always looking for the greatest happiness, well-being or, equivalently, the greatest amount of utility. This primitive conception regarded utility as a measurable magnitude in absolute terms and therefore comparable interpersonally, as if it were something that could be measured in the same way as the weight or height of individuals. A unit of measurement of utility, named "util," was even invented. Thus, it was assumed, for example, that from the consumption of an apple a person could obtain, say, ten utils, while, from a horseback ride, twenty utils, etc. So, one could add up the amount of utility achieved by an individual, for example, in a day, and compare it with that of another individual. The utility understood in this way is called cardinal utility.

However, as utility is a subjective magnitude, it became very difficult to sustain the claim of its interpersonal comparability. For example, two individuals can say that consuming an apple gives them great pleasure and, moreover, that on a scale of one to ten the two experience a level of pleasure or utility equal to ten. But from an objective point of view, we have no way of knowing whether a level ten effectively means the same thing to both individuals.

This inadequacy of the original concept of utility then led, from the works of John Hicks and R. G. D Allen in the 1930s, to their replacement by the concept of ordinal utility, which only requires for its construction the existence of a consistent order of individual preferences. For this conception of utility, it is only necessary for an individual to establish a preference ranking, specifying, for example, whether given two goods A and B, her preference for A is greater, equal, or less than for B, without attempting any quantitative measure, as was the case with the concept of cardinal utility.[1]

But there was still a problem, namely that such preferences, of a subjective nature, were inaccessible for direct observation. This led to Paul Samuelson's formulation of the theory of revealed preference at the end of the 1930s, which postulated that the observation of choices made by individuals reveals information about their order of preferences.[2] For example, if we observe that given two goods A and B, an individual with

[1] Introductions to the concepts of preferences and ordinal utility can be found nowadays in any introductory textbook to microeconomics. The first formulation of those concepts is in Hicks and Allen (1934).

[2] This theory is presented today in almost all microeconomics textbooks. The original formulation is in Samuelson (1938).

a enough income to acquire any of them, buys good A and not B, we can infer that she prefers A over B.

Samuelson's theory had a strong connection to behaviorism, a branch of psychology that was prevalent during the first half of the twentieth century. Behaviorism postulated either the nonexistence of mental phenomena or, if their existence was accepted, their inaccessibility to external observers. Therefore, it focused on the study of reactive behavior directly observable based on changes in the environment of individuals. The predominance of behaviorism – which, in the United States, lasted until the mid-twentieth century – was the intellectual and scientific environment within which Samuelson formulated the theory of revealed preference. This theory became, and remains to this day, the fundamental canon of ME to conceive and model the conduct of economic agents, without digging into their mental processes to perceive or decide. In this way, ME is as a self-contained discipline; that is, to articulate its explanations it does not need to refer to other branches of science, such as cognitive science, neuroscience, psychology, or sociology.[3]

1.2 The Agent in Artificial Economics

In the 1940s and 1950s, the main elements of an alternative paradigm to behaviorism began to gain momentum, with the pioneering works of Warren McCulloch and Walter Pitts in the field of artificial neural networks, and Alan Turing, John von Neumann, Herbert Simon, and Allen Newell in the fields of computer science and artificial intelligence. But it could be said that behaviorism begins to decline after the publication of Noam Chomsky's 1959 critical review of Burrhus F. Skinner's book *Verbal Behavior* (Skinner, 1957), then the greatest figure of behaviorism in the United States.

Contrary to behaviorism, which focused on stimulus and response functions bypassing the existence of internal representations in individuals, Chomsky demonstrated that to explain the development of language it was necessary to postulate the existence, in each individual, of a generative grammar model, that is, a set of rules that determines the grammatically correct combinations that can be made within a language. And that somehow those rules are already programmed in individuals

[3] For a defense of this position, see Gul and Pesendorfer (2008). For a critique, see Camerer (2008). For a discussion of mentalism versus behaviorism in economics, see Dietrich and List (2016).

from birth. Otherwise, it is impossible to explain the acquisition and development of language in children, a process whose complexity, speed, and empirical regularities greatly transcend the possibilities of conditional learning through the accumulation of experiences of stimulus and response.

This meant that in order to account for language in humans, the existence of not only internal mental representations but also of a structure of them must be recognized. In this way the rise of cognitive science began to consolidate, with an approach which postulates that by studying and modeling mental functions with methods of computer science it is possible to make verifiable inferences about the processes that take place in the mind; therefore, with a vision of mental phenomena strongly linked to the computational metaphor.[4]

In line with this paradigm, from the point of view of AE, an artificial agent is an entity capable of processing information by executing an algorithm, that is, by sequentially applying a set of rules.[5] This is no more and no less than the contemporary image of a computer. Each artificial agent has internal states and rules of conduct. Internal states are encoded into databases or data structures that contain information about agent characteristics (e.g., demographic, economic, social, cultural, etc.). While the rules of conduct are the programs (i.e., the algorithms), that encode its actions. The artificial agent takes "bits" of information from its environment, processes them to generate a representation of its world, and, given this representation, acts accordingly to achieve some goals.

In AE, an artificial agent is an artificial individual that interacts with others thus creating an artificial economy. The word *artificial* denotes here a purely algorithmic existence of agents and economies. Depending on the degree of sophistication of their behavior, different types of artificial agents can be identified. For example, Shu-Heng Cheng (2012) provides a historical review of the development of the artificial agent concept and proposes a taxonomy consisting of three categories: simple artificial agents, autonomous artificial agents, and human-like artificial agents. Simple artificial agents are those whose behavior is very elementary: for example, their behavior is always the same, or is completely random. Autonomous artificial agents are those whose behavior evolves autonomously and

[4] Historically, there have been several metaphors referring to the mind, such as a blank slate on which a person's experiences are recorded, a mechanical device with many gears, etc. The computational metaphor has gained prominence in recent decades and is the most used today.

[5] For a discussion of the use of the concept of agent in the social sciences, see Axtell (2000).

cannot be predicted from the initial conditions of the program that generates it, as is the case of agents built with genetic algorithms or artificial neural networks. Human-like artificial agents are those which autonomously deploy sophisticated forms of artificial intelligence, such as reinforcement learning and advanced models of decision-making under uncertainty, such as stochastic dynamic programming. Throughout the book we will see examples of these types of agents.

The various categories of artificial agents are of interest in themselves, since with them different models of artificial economies can be constructed – some that are based on very simple agents but that give rise to complex dynamics, and others that are based on sophisticated agents but yield relatively simple dynamics or, on the contrary, very complex ones. However, the category of artificial human-like agents is especially interesting because one of the main goals of AE is to have a concept of agent that contributes to, and at the same time is based on, the understanding of the behavior of real people, a goal that goes back to the origins of working with artificial agents and that was explicitly stated by pioneers of the stature of John von Neumann (1951 and 1958). Real people, in addition to possessing cognitive systems (minds) that can be interpreted as the software for information processing, have bodies of flesh and blood that can be interpreted correlatively as the hardware of such processes. And within that hardware, a fundamental part: the nervous system and, especially, the brain. When we see people in this way, we are assuming that the causal root of their behavior is in the virtual processes controlled by the software running on the brain hardware, while the body consists of input and output devices to interact with the world. Based on this characterization, to approach the goal of building realistic models of artificial agents (i.e., human-like models), AE draws mainly from two branches of modern science that relate directly to "the software and the hardware of behavior": cognitive science and neuroscience, which we will deal with in the next section. In this, AE contrasts with ME, which as we mentioned before, views itself as a self-contained behavioral science which does not need to resort to other disciplines.

1.3 Cognitive Science and Neuroscience

1.3.1 Cognitive Science

From the point of view of cognitive science, thought can be understood in terms of the existence of representational structures in the mind, and computational processes that operate on those structures. Thus, the mind is seen as an information processor.[6] There are several models that attempt to capture the way mental processes of representation and computing occur, but there are currently two main approaches to the functional architecture of the mind.[7]

The first one, predominant in classical cognitive science, is the rules-based architecture, whose founding fathers include Noam Chomsky, Allen Newell, and Herbert Simon and where Jerry Fodor, Marvin Minsky, and others have later stood out. This approach sees the mind as if it were a conventional computer that processes symbols sequentially using "if-then" rules and procedures that operate on them.

For example, consider the representation of thoughts about the economies of countries. When we think about such economies, we can represent them with concepts such as advanced or backward, innovative or traditional, productive or unproductive, with high or low investment, growing or stagnant, etc., and we can operate on such representations with rules such as:

a. If E is a productive economy, then E is an advanced economy
b. If E wants to be an advanced economy, then E must be innovative
c. If E is a backward economy, then E is a low-investment economy

From a database containing numerous such rules, we can make inferences that result in specific behaviors. A simple case would be:

a. E wants to be an advanced economy
b. If E wants to be an advanced economy, then E must be innovative
c. Therefore, E must be innovative

In short, a rule-based functional architecture has representational structures and operates on them (in a way, it reasons) by means of chains of rules. These operations can be complemented by efficient procedures to bring rules from memory without having to review all of them each time an inference is made; to resolve conflicts between rules; and to learn new rules,

6 For systematic introductions to cognitive science, see Thagard (2005), which is organized by modes of mental representation, Friedenberg and Silverman (2011), which is organized by subdisciplines, and Bermúdez (2014), which is organized around theories and problems. For comprehensive cognitive science coverage, see Frankish and Ramsey (2012) and Wilson and Keil (2001). For a comprehensive presentation of computational cognitive modeling, also known as computational psychology, see Sun (2008).

7 For a more extended and deeper presentation of mental architectures along the lines briefly introduced in this section, see Thagard (2012).