Researchers Value Passion and Knowledge, and Seek to Solve Problems at Hand and Those That Have Remained Unsolved for Many Years

Tatsuaki Okamoto
NTT Fellow, NTT Social Informatics
Laboratories

Abstract

One of the most fundamental problems in science is quantitatively defining the complexity of organized things. Over the past several decades, many quantitative definitions of complexity have been proposed, but a single definition has not been agreed upon. Tatsuaki Okamoto, an NTT Fellow at NTT Social Informatics Laboratories proposed a quantitative definition that simultaneously captures all three key features of complexity for the first time. We interviewed him about the progress of his research activities and the research environment in Japan and the United States.

Keywords: organized complexity, probability distribution, deterministic strings

The world's first quantitative definition that simultaneously captures all three key features of complexity

—It has been two years since our last interview. I heard that you have returned to Japan from Silicon Valley.

Yes, the previous interview took place in 2020. At that time, I was in Silicon Valley as the director of Cryptography and Information Security Laboratories (CIS Lab) of NTT Research, and I talked about the research themes in cryptography and blockchains pursued at the CIS Lab and what kind of research institute NTT Research is. After returning to Japan in July 2022, I have been researching cryptography theory at NTT Social Informatics Laboratories. While at NTT Research in the summer of 2021, I managed the CIS Lab while also providing research guidance for internship students on the topic of adaptor signatures, a cipher used in blockchain transactions.

The research I'll discuss today is related to cryptography in a broad sense, but it is more specifically

about the field of *complexity*. If we follow the microscopic world, such as the cells that make up organisms such as animals and plants, we will see they have extremely complicated structures. We also know that the universe has evolved from its simple form immediately after the Big Bang to a more complex form. When we look around us in this way, we see that we are in fact surrounded by an abundance of complex things. Complexity is the topic of scientific study through which the complexity of those things that are said to be complex in the world is defined and quantified from a unified viewpoint. The research field of complex systems has been around for more than 30 years, and many excellent research results have been reported; however, not much progress has been made in regard to the study of the underlying complexity of those systems. I have been thinking about this research field while researching cryptography and recently published a paper on a quantitative definition of complexity in the journal *Complexity* [1]. It is the world's first definition that simultaneously captures all three key features of complexity (described later).

The ideas and concepts behind cryptography have also been helpful for me to study complexity; for example, the concept of *knowledge complexity* introduced in connection with the theory of zero-knowledge proof used in cryptography provided one clue in deriving this definition.

—You have proposed a world's first in research that could be considered your life work. Would you tell us more about it?

First, let me explain the essence of the problem of complexity. Warren Weaver, an American scientist, classified scientific problems into three categories: problems of *simplicity*, problems of *disorganized complexity*, and problems of *organized complexity (OC)*. A typical problem of simplicity is one in which the motion of a few balls on a billiard table is accurately analyzed and predicted using dynamics. A typical problem of disorganized complexity is one in which millions of balls are rolling around on a giant billiard table and randomly colliding with one another and the cushions of the table, and their average motions are analyzed and predicted using statistical mechanics. A problem of OC is defined as one involving organized and complex things such as living organisms and ecosystems in which many cells are interrelated and form a single living organism as well as artificial objects. Problems of complexity are

thus categorized as "organized" or "disorganized."

In problems of complexity, the most fundamental and important notion—from a scientific perspective—is the quantitative definition of complexity. The quantitative definition of disorganized complexity (randomness) of physical systems was established to be entropy, which appears in thermodynamics and statistical mechanics. Moreover, the quantitative definition of the disorganized complexity (randomness) of information sources (a probability distribution) was established as Shannon entropy. These two entropies are essentially the same except for a constant.

Despite many attempts to give a quantitative definition of OC, there is no definition that is widely agreed upon. The difficulty with this definition lies in the fact that OC depends heavily on our senses—or can be perceived by intelligent creatures like humans—and is very different from the randomness of objects. Those previous attempts have covered either deterministic strings or probability distributions, and there was no definition that covered both simultaneously.

Capture all three key features of complexity (descriptive, computational, and distributional) simultaneously

—So, it is necessary to establish a definition that covers both deterministic strings and probability distributions at the same time.

I considered what should be the object of the definition of OC. The object of complexity is everything around us, which includes the stars and galaxies in the universe, living organisms, ecosystems, artificial things, and even human societies. The only way for us to recognize those things is observation via telescopes, microscopes, various instruments, and electronic devices. We directly recognize the existence of an object by actually holding it in our hands; however, this recognition is also a result of observation; that is, our brain processes the observed data obtained through our five senses (sensors) to interpret and recognize the existence of the object. In other words, we perceive everything on the basis of observed data.

In general, the source (information source) of observed data (i.e., a deterministic string) of physical phenomena is a probability distribution, and the observed data are values randomly selected in accordance with that distribution. It is therefore natural to assume that the object of OC is not the observed data (i.e., deterministic values) selected by chance from

the source but the source itself (i.e., probability distribution). Accordingly, I considered that the object of OC should be a probability distribution (i.e., the source). Since a deterministic string may be a special case of a probability distribution (in which only one value occurs with probability of 1), it is also an eligible object.

—What were the problems with the existing quantitative definitions of OC?

Let's take two cases as an example. First, we give a chimpanzee a computer keyboard and allow it to press keys freely and generate a character string of 1000 letters of the alphabet. Second, we take a character string of 1000 letters of the alphabet from one of Shakespeare's plays. Both cases are deterministic strings; however, comparing the two sources of information clearly reveals the difference in complexity of the two strings.

The information source of the chimpanzee's string is almost a uniform distribution (i.e., a simple distribution) of 1000 random letters, so its OC is at the lowest level. On the contrary, the information source of Shakespeare's string is the probability distribution of several candidate expressions in Shakespeare's mind. Regardless of whether Shakespeare chooses from several expressions that come to mind (i.e., a complex probability distribution of several candidate expressions) or he chooses one expression without hesitation (i.e., a probability distribution of one complex string occurring with probability of 1), both distributions are complex, so their level of OC is very high.

If we consider OC in Shakespeare's strings, we notice that complexity has several different features. Since the object (information source) of OC is a probability distribution, one feature of complexity is the complexity of the pattern of probability distribution (distributional feature). Another feature of complexity is the complexity of the amount of description (descriptive feature) such as the knowledge required to write a play (language, history, culture, etc.) and the play's story. Moreover, when Shakespeare decided what he wanted to write, he would have thought deeply to choose the appropriate literary expressions to express that decision. Choosing such linguistic expressions can be regarded as a graph problem, and choosing an appropriate expression has a computational complexity (computational feature) that is equivalent to solving a graph problem. Therefore, OC has three features: distributional, descriptive, and computational, and to define it, all of these features must be captured simultaneously. Of course, values of complexity should be computable, and the complexity of problems of disorganized complexity (as in the above-mentioned example of a chimpanzee) or problems of simplicity should be low.

The problem with the existing definitions is fourfold: first, they capture only one of these three features of complexity (distributional, descriptive, and computational); second, they target either deterministic strings or probability distributions and cannot seamlessly target both at the same time; third, the complexity of some definitions is incomputable; fourth, some definitions are not rigorously specified (**Table 1**).

Accordingly, I set out to and successfully established a new quantitative definition of OC that would solve all four of those problems.

You need to have enough energy and passion to keep pursuing your research—whether you are asleep or awake

—You have overcome a longstanding issue. Please tell us about your definition of OC.

The basic idea of my quantitative definition of OC is that OC is defined by using the smallest stochastic automaton-form of a circuit (*oc-circuit*) that simulates an object, i.e., information source (probability distribution). (Making the oc-circuit the "smallest" means my definition accords with the principle called Occam's razor.) While some computational models must be used to capture the descriptive and computational features of complexity, the existing definitions use a Turing machine (a computer). Instead, my definition uses a circuit (oc-circuit). The reason for using a circuit is that the three features of OC can be obtained simultaneously. Moreover, a definition using a circuit is computable, whereas a definition using a Turing machine is incomputable.

Therefore, my quantitative definition of OC satisfies all the requirements and overcomes all the problems of existing definitions: (i) the three features of complexity (distributional, descriptive, and computational) can be obtained simultaneously, (ii) probability distributions and deterministic strings can be seamlessly covered as objects, (iii) the definition is computable and (iv) rigorously specified, and (v) the complexity of problems of disorganized complexity and problems of simplicity is low (Table 1).

My definition can be applied to theoretical

Table 1. Comparison with existing definitions.

Definition	Object (probabilistic and deterministic)	Simply regular	Simply random	OC			Computable
				Descriptive	Computational	Distributional	
Kolmogorov complexity	×	✓	×	✓	×	×	×
Effective complexity	×	✓	✓	✓	×	×	×
Logical depth	×	✓	✓	×	✓	×	×
Effective measure complexity	×	✓	✓	×	×	✓	✓
Statistical complexity	×	✓	✓	×	×	✓	✓
My definition of OC	✓	✓	✓	✓	✓	✓	✓

Definition	Approach	Rigorously specified	Computational models	Computable	Features
Kolmogorov complexity	Occam's razor	✓	Turing machine	×	Descriptive
Effective complexity					Descriptive
Logical depth					Computational
Statistical complexity			Stochastic finite-state automata ($\in$-machine)	✓	Distributional
My definition of OC			Stochastic automaton form of circuit (oc-circuit)		All
Effective measure complexity	Information theoretic		None		Distributional
Natural complexity	Framework	×		—	—
Thermodynamics depth				—	—

—: Difficult to evaluate

foundations for machine learning and artificial intelligence, algorithms with OC, networks, analysis of amounts of computational and communication limits and averages, and semantic information theory. I believe that, for example, in contrast to Shannon's information theory, which treats entropy (disorganized complexity) as the amount of information for the problem of signal transmission (syntactics), a semantic information theory that deals with the problem of conveying meaning (semantics) can be constructed using my definition of OC as the amount of (semantic) information (**Fig. 1**).

—The impact you have academically and socially is immeasurable. Finally, please tell us what you value as a researcher and what you would like to say to Japanese researchers who are aiming to work abroad.

I believe that a researcher is someone who values passion and knowledge, and attempts to solve problems at hand and problems that have remained unsolved for many years. With the spread of the Internet, the world has become borderless. It is thus important for researchers to approach their research with a global perspective.

As I mentioned, I have worked with some of the best and brightest researchers at the CIS Lab, NTT Research in Silicon Valley. The CIS Lab is full of energy since the world's best and brightest researchers are stimulating and competing with each other. Some of the researchers are enrolled in doctoral programs and are quite enthusiastic. I believe it is important for young Japanese researchers and graduate students who aspire to become researchers to possess such enthusiasm for research. If you want to compete with powerful researchers around the world and accomplish research that will have an impact on the world, you will need to have enough energy and passion to keep pursuing your research—whether you are asleep or awake.

If we want to produce world-class researchers from Japan, we may have to reform our research system and structure that supports them. For example, in the U.S., PhD researchers usually start as postdoctoral fellows. A postdoctoral fellow is not positioned as an assistant researcher, but rather as an independent researcher with their own research theme. When their tenure is up, if their results are good, they will continue to research as a researcher in an upgraded
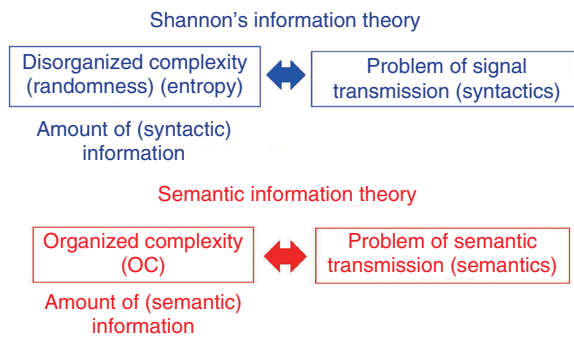
Fig. 1. Application to semantic information theory.

position (obtained through competition). (Although this is the case for theoretical research, it seems to differ in the case of experimental research.) As an independent researcher, you are more likely to develop yourself through competition with your peers than by learning from your professors. I therefore think it is necessary to create an environment in which outstanding researchers from around the world can gather so that young researchers can develop their career as researchers without depending on where they do their research.

Nevertheless, it is important for researchers and those who aspire to become researchers to choose where they can conduct research and where they can engage in competition—without blaming the environment if they make the wrong choice. One choice is to work at a research institute abroad. In that case, living abroad can be stressful owing to differences in language and culture. Such a situation could be a disadvantage, although time may solve that problem

to some extent as they get used to life in another country. I hope that when they determine to research abroad, they will bring enough enthusiasm that overcomes such stress.

Reference

[1] T. Okamoto, "A New Quantitative Definition of the Complexity of Organized Matters," Complexity, Vol. 2022, Article ID 1889348. https://doi.org/10.1155/2022/1889348

■ **Interviewee profile**
Tatsuaki Okamoto received a B.E., M.E., and Ph.D. from the University of Tokyo in 1976, 1978, and 1988. He has been working for NTT since 1978 and is an NTT Fellow. He is engaged in research on cryptography and information security at NTT Social Informatics Laboratories. He served as the director of Cryptography and Information Security (CIS) Laboratories at NTT Research in USA from 2019 to 2022. He also served as president of the Japan Society for Industrial and Applied Mathematics (JSIAM), director of International Association of Cryptology Research (IACR), and a program chair of many international conferences. He received the best and life-time achievement awards from the Institute of Electronics, Information and Communication Engineers (IEICE), the distinguished lecturer award from the IACR, the Purple Ribbon Award from the Japanese government, the RSA Conference Award, and the Asahi Prize.