Modeling and Analysis of the Effects of Information Contexts induced by the Structure of the Information

２０１

吉武　真

筑波大学大学院
ビジネス科学研究科　企業科学専攻

Modeling and Analysis of the Effects of Information Contexts induced by the Structure of the Information

Doctoral Dissertation

Makoto YOSHITAKE

Graduate School of Business Sciences
University of Tsukuba

UNIVERSITY OF TSUKUBA

Abstract

Graduate School of Business Sciences

Doctor of Philosophy

Modeling and Analysis of the Effects of Information Contexts induced by the Structure of the Information

by Makoto YOSHITAKE

This thesis explores an understanding of the fundamentals of information and develops a new framework of its mathematical description, *i.e.*, model of context–dependent, absolutely relative information at the fundamental level, on the basis of physical phenomena. Today, the most dominant picture of information in general must be a so-called container/conduit–model, which explains information as a parcel carried from the sender to the receiver via a conduit. This model successfully explains many kinds of information phenomena. However, this model has not attained enough success on the matter of understanding of fundamental aspects of information such as relativeness, contextuality of information. In this thesis, the relativeness and contextuality of information are considered to be essential, inherent properties of information phenomena, and a mechanism that realizes such features materially/extensionally is explored as a foundational model of information phenomena.

The present study grounds information phenomena on physics by modeling the process of extensional information articulation as a generalization of the theory of the extensive measurement in metrology. The model operationally defines the information phenomena that this thesis studies, eliminating speculativeness as much as possible. The model is demonstrated in the case of, so to speak, the attribute extraction from the universe. The scene of the attribute extraction may be considered to be a representative of fundamental information phenomena. In this concrete model of an information phenomenon, it is shown that the attributes as information may be obtained extensionally, that is, by mere physical interactions between the system under consideration and the universe, in a canonical manner.

Information, or its meaning, must be *canonically* extracted from the information phenomena with *universality*, to be utilized by the system that receives the information. Information is something useful afterward for the system. To be identified and used after a time interval, information must be universal, and be canonically obtained. The universality of the attribute is understood as a mathematical limit of the sequence of structures of (proto–)attributes. Each of the structure is formed by inclusion relations of the set of (proto–)attributes. These structures may form a sequence in their inclusion order. The actual attributes are understood as certain elements of the limit structure of this sequence. Limit has universality in mathematical sense. Thus attributes are understood as having universality to be qualified as information. The canonicy is attained by the Galois connection between the structure formed by the sets of (proto-)objects and the structure formed by the sets of (proto-) attributes and by the project limit of such structures. The actual object and attributes for the information gathering system under consideration are determined simultaneously in the mathematical dual relation that is formed by the connection. This model naturally explains the arbitrariness of the articulation of the attributes and the contextuality of the attributes.

To describe the inherent contextuality and other fundamental features of information, the language of category theory is proposed in this study. Basics of category theory are introduced, and it is shown that its features, such as the relativity represented by arrows, the structural contextuality of the elements, the structural universality, and the existence of the subobject classifier in some categories, are endowed with the properties suitable to investigate the information phenomena. The validity of this proposal is demonstrated also in the case of the attribute extraction. By using the language of topos, a kind of category, for modeling the attribute extraction, a new, internal description is obtained and this description provides a new understanding of the attribute as information.

The essential relativity of information causes some difficulty in understanding information itself, that is, a mysterious concept of *information dismantled of its context*. This thesis provides an understanding of it by using the algebra known as torsor. Torsor is utilized in an algebraic model for representation in this study. Use of torsor enables us to make no direct mention of the represented thing. This property of the model is suitable at the fundamental level of information phenomena studied in this thesis because the represented thing is assumed to be vague before it is articulated. The thesis presents how to interpret a representation

phenomenon to the algebraic structure of the model. The correspondence relations of the elements, the two-sidedness of representation and the closure concept play significant roles in the construction of the toy model. Through the examples, it is shown that the axioms of torsor tend to hold and correspond to some consistency requirement for the representation.

As is shown in this thesis, the relativeness and contextuality of information are considered to be underlying every information phenomenon. Therefore the understanding obtained in this study, and the model developed in this study are expected to contribute to explaining a wide range of information phenomena. Besides the attribute extraction studied in this thesis, which is a typical information phenomenon, there might still be other information phenomena to be investigated. Information phenomenon that is originated in some structure other than the inclusion structure, e.g., network structure, might be a candidate, if it would exist. The language of category theory, which naturally involves the relativeness and the contextuality, is expected to be used more widely in the studies of information. The characteristics of information understood in this thesis, such as the essential contextuality and the bootstrapping capturing process at the information level might be a matter of interest of the researchers in other disciplines such as linguistics, communication theory, robotics, and many others. This thesis may make some contribution in these fields.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Yasu-fumi Saruwatari for the continuous and tolerant support of my lengthy Ph.D study and research. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank all the reviewers inside and outside of GSSM, who have contributed to improve this thesis.

I thank my fellow, successive labmates in Saruwatari Group for the various stimulating discussions.

Last but not the least, I would like to thank my parents for giving birth to me at the first place and supporting me spiritually throughout my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 *Anomalous* information phenomena?

The word "information" has been given different meanings by various writers in the general field of information theory[1]. There are long lasting discussions on the concept of information and the way to study information scientifically. We only name a few general references: [4–8, 12].

There must be no dispute that one of the most successful theory, or the most dominating model of information is by Claude Elwood Shannon [24, 25][2]. Although Shannon prudently restrained himself from applying his theory to the fields other than the engineering of communication, as is known well, "Information Theory" originated with Shannon seems to have dominated as a foundational theory for information related disciplines – computer science, physics, communication theory, probability theory, statistics, and others [26].

Thus, today, the most prevailing folk picture of information is akin to so called the *container/conduit–model*[3] (Figure.1.1), originated from Shannon's model. The container/conduit model reflects our tendency to regard information as something

---

[1]In this thesis, "information" is defined operationally in Chapter 3. As is mentioned in [7], for a science like information science it is of course important how fundamental terms are defined. In scientific discourse, theoretical concepts are not true or false elements, but more or less fruitful. To avoid worthless disputes over the speculative definitions, the (class of) "information" in this thesis is presented as what is obtained by the process presented in Chapter 3.

[2]See 2.2.2

[3]This word is used in ordinary, broad sense as described in the following sentences, not strictly corresponding to the term in linguistics [91].

like a parcel to be delivered from the sender to the receiver. Meaning, content, or significance of the information is thought to be packaged in the representation of the information, *i.e.*, signs, words, messages, *etc.*, and to be passed through a conduit to the receiver. This picture certainly touches the matter of heart, even though it cannot escape censure for its superficiality. Thus, as is already mentioned above, "Information Theory" that is based on the container/conduit–model has been developed to a high–degree and has made a great success to bring the so–called "Information Age".
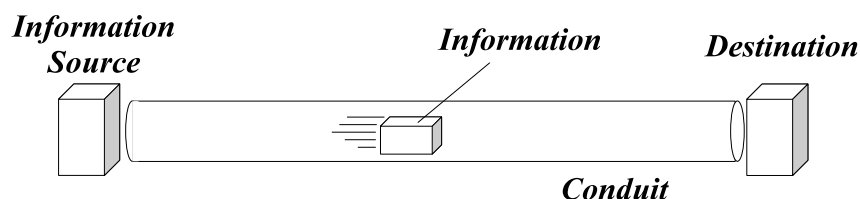
FIGURE 1.1: Container/Conduit model

However, if we keenly observe our daily experiences, especially around the scenes of information acquisition, we may find that the container/conduit–model is not enough. The following episodes are presented to help the readers to conjure up what is happening in their information lives. These episodes might be anomalous, or, pathological in the view of the container/conduit–model, but the readers may find that these episodes have something in common with ordinary, daily experiences of the readers themselves.

1.1.1   Episode 1: Arbitrariness of articulation of the universe

Suppose that there is a green apple. We, as experienced humans, can recognize that this fragment of the universe (*i.e.*, the apple) is green. We have information that the apple is green. But how about an infant or a brand-new *tabula rasa* autonomous robot just thrown out into the physical universe? When such a system tries to gather and utilize information about its surrounding universe, how can it obtain the information? They do not know what the *green* is, or, how *green* something is.

One may wonder if it is determined by the sensory system embedded in the system under consideration—our eyes, the retinae, the visual cortex, and other organs enable us to *see the green*. Yet this is not completely true. Yes, the visual system actually captures some optical stimuli from the surrounding universe. In general, the stimuli captured by the system that is gathering and utilizing information are actually restricted by the capability of its sensory subsystems. However, the way in which the information system *articulates*[4] the stimuli is another matter. In the case of the color system for example, it is known that the color system differs from race to race, from language to language [92][5]. Some may distinguish green from blue and others may not.

Thus even the information telling that "this is green" does not seem to come as a package delivered uniformly to all.

1.1.2 Episode 2: Contextuality of words

The meaning of a word is understood in the context of a language. The word pronounced as [bǽt] , *i.e.*, /bat/[6], is different from the word pronounced as [vǽt], *i.e.*, /vat/ in English, because there is a phonemic opposition, that is, a *distinction* between [b] and [v] in English. The word pronounced as [bǽt] might have the identical meaning as the word pronounced as [vǽt] in the context of the language that has no phonemic opposition between [b] and [v]. In other words, information that a word provides is not determined by the word alone. It is determined in a context which consists of other words in the language system to which the word is supposed to belong. The word /kani/ means a ⟨⟨rabbit⟩⟩ in Finnish, but it means a ⟨⟨crab⟩⟩ in Japanese[7]. There ought to be other *contexts* – cultural, social, communicative, conversational, and many others – that affect the information. Information differs in different contexts. It is a source of misunderstandings, puns, double meanings and other ordinary information phenomena.

---

[4]Throughout this thesis, the author always uses the word *articulate* in a similar sense to phonetics, that is, "*divided into segments.*"

[5]Their emphasis is on the cognitive commonalities not only within a color system but also among different color systems, though.

[6]In this section, single slash is intended to indicate a representation and double angles indicate a content. Thus, /xxx/ represents ⟨⟨xxx⟩⟩.

[7]It should be better to use other symbols such as phonetic symbols for /kani/ here to be precise, but the priority has given to the readability.

The container/conduit–model seems not to consider the context seriously. The words container and conduit imply that this information mechanism excludes surrounding circumstances. The container contains the definite information, and the conduit transfers it exactly as it is. There seems no room for context–dependency in the container/conduit–model.

1.1.3   Episode 3: One stimulus, multiple information

When a child finds a blot on the wall in an airport and the blot has a form like /P/, the child may regard the blot as ⟨⟨pee⟩⟩. But another child may regard the blot as ⟨⟨er⟩⟩. And the other child may regard the blot as ⟨⟨rho⟩⟩. One can think of a blot as a Latin letter, a Cyrillic letter, or a Greek letter. But what on earth are we talking about? The letter that one of the children found is *different* from the letters found by the other children. Can we say that we are talking about *the same blot*? It might just happen to be this simple case. But imagine a blot which has a more complex form such as a sequence of ideographs[8] and Arabic letters, and suppose that one does not have enough knowledge of the writing system of the language. When a child finds that a part of the blot is similar to a certain letter, can one articulate *the part* from the blot with confidence? Even the articulation of a letter is no easy task.

This might also be considered as a context–dependency, but it has a different aspect. In this case, information differs evidently between the receivers/informees. Whether the theory regards the receivers as a part of the context or not, a fundamental theory of information ought to explain this difference.

1.1.4   Episode 4: Information and physical process

The definition of information seems very difficult to deal with. Physical interactions might be considered to be a good point to departure. No one would deny that information has a physical base. But it is not easy to define information even on the physics base. There seems to be a certain spectrum of informational nature of physical interactions. A falling heavy body makes a dent on the surface of the floor. We, humans, tend to regard this as a mere physical consequence. The falling

---

[8]For instance, Chinese characters, or Kanji.

body or the *falling* may not be regarded as information for the floor. In contrast to this, let us consider a typical information phenomenon: in many countries, a change of a traffic signal to the red will make cars stopped. This situation may be intuitively informational. The change of the traffic signal may be regarded as information for the cars or the car-driver composite systems.

If information is always transferred as a package, this means that there are two distinctive types of physical interactions, one in which "the package of information is *emitted*," and the other in which "the package of information is not *emitted*." Even though the *emission* is not literally a physical process, this distinction seems strange to the author of this thesis. A fundamental study of information ought to explain the differences among such information related phenomena in the view of information, even if the phenomena are not apparently causally related, that is, not well-suited to be studied in physics.

1.2   Vicious Circularity of the Foundation of Information

Such information phenomena described in 1.1 seem ubiquitous in the depth of every information and moreover they actually are considered to form elementary, fundamental processes in every information phenomena. Information processes described in the above episodes are concerned with the origin of *distinction* for the system that is gathering and using information.

Making distinctions into the universe, namely, articulating the universe, by the system is understood to be the basics of information. A human may find an optical distinction between a certain closed area in the space and the background, and she/he may catch the information that the area is distinctive from the rest of the space. The area might correspond to be a green apple. An English speech recognition system may find a phonetic distinction between the word /bat/ and /vat/ and distinguish them. *Every* information is based on the distinction.

But the distinction itself is considered to be also based on the information of the differences within the distinction. A difference found in the world is a lack of uniformity of the universe. That is, where there is a lack of uniformity, there is distinction. To make a distinction, the system must find differences. In other

words, differences must appear as information for the system. Some of the readers may think that the differences are *inherently* found by the system as a physical existence. As a matter of fact, the visual system of the human has its own resolution, and the ocular ability of the human to capture the world is inevitably restricted by this resolution. However, the resolution that humans utilize as information is much coarser than the physical resolution of their visual systems. In contrast to the homology of the construction of physical visual systems among humans, the difference of the systems of visual information[9] is not negligible.

Information requires distinction and the distinction requires information. There seems to be a *vicious circularity* for the fundamental understanding of information. Thus our understanding of information should inevitably be suspected of a circular reasoning when we are naively assuming the distinction underlying information. If you consider it from the other point of view, such information phenomena like the above episodes of the scene of making distinction to capture information will be shed light on the foundation of information.

1.3 Difficulties of Fundamental Studies of Contextuality of Information

Foundational information sciences study information in general. As a scientific discipline, information sciences tend to consider information as some kind of *objective* existence that is isolated from the circumstances of each occasion of phenomena. They mainly focus on the universal characteristics of information which need not be bothered with the details of the circumstances of each occasion of phenomena, namely, *contexts*. Information is assumed not to be bearing its context in such information sciences.

Seemingly, such information phenomena described in 1.1 are very diverse, separately isolated and too much depend on the individual situation/context. Information in these phenomena seems lacking the universality that the foundational information sciences search for. Therefore such phenomena tend to be out of the scope of the target of foundational sciences.

---

[9]For instance, the color system.

Besides the difficulty of finding out the universality of context–dependent information phenomena, information as a contextual existence has another difficulty to investigate. "Information as a contextual existence" means that we should always take the individual, various context into consideration when we study information. If the *whole* information bears its context, each information phenomenon might inevitably be individual, separately isolated. The universality as the fruit of scientific endeavor must, if exists, reside in information whose context is ripped off.

However, it is difficult to understand, even imagine, information without its context. Many of us usually have the image of information with its context. Information is always accompanied by the background scene (in the case of visual information), surrounding silence or some noise (in the case of auditory information), other symbols (in the case of alphabet), other words in the same language (in the case of language), and so on. The information we imagine is inseparable from its context.

Information is a difference, as an anthropologist Gregory Bateson [37] defined: "information is … difference which makes a difference." It is a convincing definition because, as the author has already written above, information is based on distinction, *i.e.*, difference. Information without context might be *pure* difference. Difference is something in between. Thus information is inherently relative.

Even if we could attain ideation of "information without context," this relativistic nature of information would raise another difficulty of the fundamental studies of information. There are few ways even to describe the relativeness itself. It must be very hard to study information fundamentally without good tools that may describe, analyze and operate its relativeness naturally and properly.

1.4 Fundamental Information Studies on Context – Thus Far

Fundamental information studies seem to have not hitherto tackled the essential context–dependency of information squarely. There may be considered several reasons such as the followings.

Firstly, as we will see in Chapter 2, most of the preceding scientific studies assume distinctiveness of the information carrier[10] as a self-evident truth. The distinction is regarded as some kind of a gift from Heaven (via the physical construction of the system) in such studies. The vicious circularity which underlies beneath information phenomena and the essential context-dependence of information phenomena are put out of their scope of the studies in most cases.

Secondly, the multi-layered-ness, or, simply saying, the "complexity" of information phenomena easily hides the elementary context–dependent process. Information phenomena present a multi–layered structure. Optical stimuli form a shape of blot, the blot is interpreted as a letter of some alphabet, some such letters form a word, words becomes sentences, sentences mediate communication, and so on. Each of the studies of information has a specific layer of information phenomena that attracts the attention of the researchers in that field and that is investigated intensively. Because of its fundamentality, the context–dependence of information phenomena concerning with fundamental distinction makes an appearance even in the studies that pays no attention to it. In prior studies, the context–dependence is analyzed and understood at this specific layer of information phenomena. The layer has its own interest, and such interest at the "higher level" of the layers hides the elementary context–dependent process at the most fundamental level.

In fact, each related discipline at these higher levels addresses the context dependencies at its own level. For example, in the linguistics related disciplines, the more pragmatic the focused field of study is, the greater importance is attached to the contextuality. As is emphasized in Saussure's semiology [59, 68], a sign must be articulated from other signs to signify something. These other signs as "background" may be considered to form the context of the particular sign. Thus, even at the level of signs, the contextuality makes an appearance. In theories in pragmatics, information beyond "what is said" is explicitly modeled as constituents of the theory. Austin's perlocutionary act [61], Searle's Background [62, 65], Gricean conventional implicature [64], Sperber and Wilson's contextual implication [66], though from the different respective viewpoint, all of them tackle to information not included the utterance itself.

Although the phenomena addressed at these studies at the higher level of information may be considered as an appearance of the fundamental contextuality, from

---

[10]This is not a term that has a strict meaning. *Information carrier* here loosely means *the (physical/material) source of the information.*

the standpoint of this thesis, prior theories for these phenomena are naturally not enough general to be a fundamental information theory. They assume the distinction between the *main* part and the *context* part of a phenomenon. In other words, they analyze contextual phenomena in a postmortem fashion. But for the system itself, before the *context* part has been established *by information*, there seems no *main* part. The distinction required for information seems to require information. Vicious circularity is there. This should be resolved at the fundamental level of information.

Thirdly and lastly, without assuming the fundamental distinction, the basis of study, the starting point of investigation, the parts of the model for mechanism, of information seems hard to establish. Once you decide to pursue the way to resolve the matter, you will find that you must grope in the dark to search for the theory that forms the basis of the study, the methods that might be effective, and the tools that might be useful.

1.5 Beyond conduit models

In this thesis, we explore an understanding of the fundamentals of information. The aim of this study is to understand the essential contextuality of information in a non–speculative manner, resolving the difficulties mentioned in 1.3, by revealing a mechanism of information phenomena. Assuming the systems in general[11], we focus on the universality that the system utilizes as information, and on the canonicity that enables the system to capture information. Contexts are understood as those that are constructed structurally, and a new framework is developed for a mathematical description of information as a *context–dependent, absolutely relative* information phenomena to the system concerned.

To attain that aim, we concentrate on a particular aspect, or a particular stage, of information phenomena in this thesis. As a semiotician Umberto Eco [60] depicts (see Figure.1.2[12]), we can consider that there is a continuum at each end of an

---

[11]That is, the systems presupposed in this thesis include natural systems such as animals and humans, and also include artificial systems.

[12]Table has been rearranged and the illustration has been added by the author.

information phenomenon[13]. The continua are usually out of the scope of scientific information studies. However, as the readers will see, the present study focuses on this boundary[14] of the continuum and the *articulated* units, if its focus is attempted to be shown in this figure. This act of *articulating* is considered to be underlying every information phenomenon. The words like "fundamental" and "foundational" in this thesis are used in such a sense. Information is often anatomized into data plus meaning[15]. This picture of information gives us an impression of independence of the meaning from the data (and vice versa). Although this picture is indeed effective for postmortem analysis of information, it will be abandoned in the view of this study. From the aspect that the present study addresses, data and meaning are articulated from the continuum at each end in an intertwined manner, as will be shown in Chapter 3. Although this thesis is seemingly considering a limited part of information phenomena, the study from the aspect that is addressed in this thesis is anticipated providing better understanding of information in general, because of its fundamentality.

		Message			
Theory of Commnucation	**Channel**	**Sign-vehicle**		**Meaning**	**Source**
		Expression		**Content**	
Theory of Codes	**Stuff**	produced units (tokens)	syntactic system (types)	semantic system (types)	interpreted units (tokens)	**Experience**
		Code			
		Correlation between functives			
Formal Model	**Continuum**	posited units	system of empty positions	system of empty positions	posited units	**Continuum**
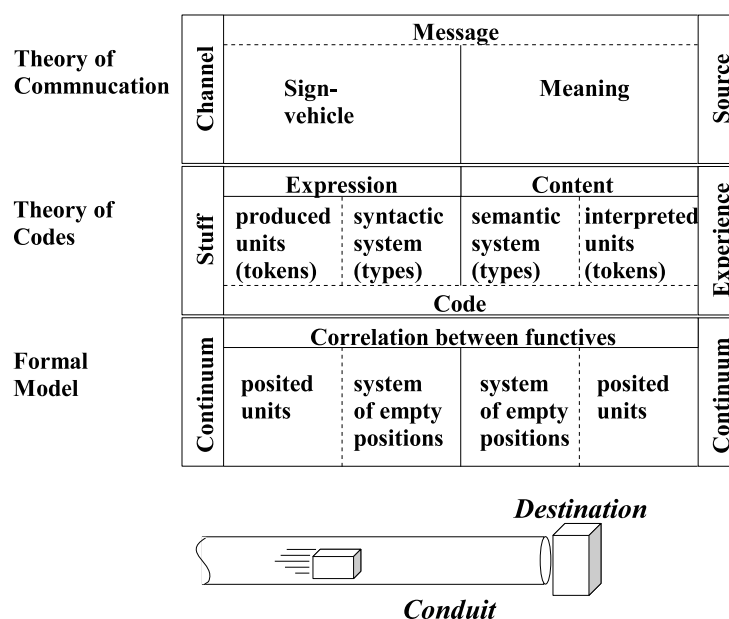
FIGURE 1.2: Model of sign–function and its underlying code [60]

---

[13]The usage of the word *information* varies even among the researchers. Eco [60], for instance, states his own view. However, the readers need not to be bothered with the difference between the semiotics and other information sciences here. They are boldly considered to be the same in this section.

[14]In the expression in [60], it is described as *the lower threshold of natural boundary of semiotics*.

[15]For instance, see the General Definition of Information in [14].

To eliminate speculativeness as much as possible, this study operationally defines the process of information phenomena at the "receiving" system. From the other point of view, *information in this thesis* is defined by physical operations[16] that will be presented in Chapter 3. Information must have a physical support [93]. More modestly, a class of information must have a physical support. Though there seems no definitive agreement on what information is, one may agree that information is induced by a physical phenomenon. That is, at a certain level of investigations, information may be considered to be physical phenomena related to a class of physical interactions between a certain part of the physical universe[17] and another part of the universe. In this thesis, we will call such an interaction as *information phenomenon*, and the latter part of the universe will be called an *Information Gathering and Using/Utilizing System*, abbreviated to IGUS[18] [94, 95]. We will use this abbreviation to specify such a subsystem of the universe. An IGUS is also merely a physical phenomenon in the universe. Ultimately, this physical world is just a bunch of physical interactions. Thus, in summary, information in the view of this thesis are considered to be (abstracted) physical interactions between an IGUS and the universe.

From this viewpoint, an information theory might be regarded as a kind of *phenomenological theory* of physical sciences[19] (Recall the episode in 1.1.4). *Phenomenological theory* here means a theory that expresses empirical observations of phenomena, in a way that is consistent with fundamental theory, but without paying detailed attention to their fundamental significance. The fundamental theory is physics here, and there are many phenomena among the word "information" that are not always directly explained effectively in the words of physics. Though this view may naturally arouse much discussion, in this thesis, the author will use the word *phenomenological* in this sense because this word seems to be convenient to transfer the grand view of this thesis concisely to the readers.

---

[16]Since the argument in Chapter 3 is primarily based on the macroscopic physics, information phenomena in this thesis are assumed to be macroscopic one. Although the possibility of a direct application of this study to the microscopic phenomena is not denied, quantum phenomena as information have not been sufficiently investigated in this study yet.

[17]Throughout this thesis, the word *universe* is consistently used to emphasize this uniformity—the status prior to being articulated. In contrast to the word *universe*, in most of the case, the word *world* used in this thesis implicates the existence of IGUS(es) capturing it.

[18]Note that those that regard a physical phenomenon as information for a certain IGUS are another IGUSes—human beings as IGUSes in many cases. This special kind of IGUS will be called a *human-IGUS* in Chapter 3. In general, such IGUSes are not restricted to humans.

[19]Note that it should not be confused with "a kind of phenomenological theory *in* physical sciences."

On this extensive, physics basis[20], the study in this thesis focuses on the emergence of both of information and the distinction that form the information, that is, what is represented as information and what represents the information. The emergence is explicated in a bootstrapping way through a connection between them. This thesis especially focuses on the scene of, so to speak, *attribute extraction*[21], as a representative of fundamental information phenomena. Attributes are information that is used to distinguish and classify objects. Since information and distinction are inseparable, the understanding of the information phenomena of the attribute extraction is significant.

The connection between what is represented as information and what represents the information is established as the connection between the structures of contexts. That is, *the context of information is captured structurally in this study*. There are two contexts that ought to be taken into consideration. One is in the IGUS side, and the other is in the universe side[22]. Although the former context may be considered to be in connection with the significance of the information, and the latter context may be considered to be in connection with the representation of the information, it will be found that these two contexts may not affect the information phenomenon independently. Information is relative to both of these contexts. In the case of the attributes, it is understood that the connection is established between the structure formed by the sets of (proto–)objects and the structure formed by the sets of (proto–)attributes. The actual object and attributes are determined in the mathematical dual relation that is formed by the connection. Each of the objects or the attributes is understood as relative existence embedded in their respective context.

Information, or its meaning/significance[23], must be *canonically* extracted from the information phenomena with *universality*, to be utilized by the IGUS. Information is something useful afterward for the IGUS. To be identified and used after an interval, information ought to be universal, and be canonically obtained (in contrast to the arbitrariness of articulations).

---

[20]The word *physics* here is not intended to mean a science that deals with matter and energy, and their interactions. By this word, the author means *the physical processes and phenomena of a particular system under consideration.*

[21]In other words, *attribute capturing* or *attribute attribution.*

[22]To be more exactly, the latter is also in the IGUS. It is an image of a part of the universe that reflects the circumstances in the universe.

[23]Do not mind the definition of the meaning of the word *meaning* or *significance*, here. Later, it will be defined (implicitly, though) in the language of the model developed in this thesis.

In this study, universality and canonicity are analyzed by, and realized as, a mathematical model. This thesis introduces the language of category theory and torsor, as tools for the investigation of information. The language of category theory has many advantageous features to this study. It may naturally describe the relativeness of information, provides the precise definition of the universality, and enables us to handle complex structures of information more easily. The torsor enables us to manipulate the torso-like characters[24] of information – the fact that torso has no head may be considered to correspond to the *information without its context*. With such a mathematical language, we can understand information more deeply with rigor.

In the view of this study, communication, *i.e.*, information transfer, is regarded as an information phenomenon, that is, an interaction between a system and its environment, not between the sender/informer and the receiver/informee. In this fundamental, elemental view, communication is a process in that an IGUS as a sender causes an event in the universe, and also in that the event causes something within another IGUS as a receiver. Indeed, there is no sender and receiver as in the container/conduit–model, instead, the IGUS simply "observes"[25] the surrounding universe to obtain information. Sometimes the observed event might be the one which is originated by another IGUS. And in such a case, *as a result*, the observed information *is regarded as* "communicated" one.

Thus this thesis will provide a new understanding of information by focusing on the mechanism of the emergence of information and by modeling the mechanism in the language of mathematics. This model will be considered to include the container/conduit–model as a special case, therefore the author expects that the understanding developed in this study may be regarded as deeper than before.

---

[24]This is just wordplay. What it means precisely will be revealed in Chapter 5
[25]This is just a personification to stimulate the imagination of the readers.

1.6   Summary of Contributions

In this thesis, the author:

- Defines information phenomena as extensive process (in Chapter 3);

- Demonstrates the validity of the above definition by the "attribute" information extraction, in a manner that clarifies the universality of the attribute as information and the canonicity of the extraction (in Chapter 3);

- Proposes the language of category theory as a description language of information phenomena (in Chapter 4);

- Demonstrates the validity of the language of category theory by providing an internal description of an origin of the "attribute" (in Chapter 4);

- Introduces torsor as a language of *information without context* (in Chapter 5).

The contents of the Chapter 3, 4, 5 are basically based on the papers [1], [2], [3], respectively.

Chapter 2

Review of the concept of information and its context

2.1 Introduction

As is known well, *information* is studied in many disciplines. For example, a sociologist Niklas Luhmann [114] defines social systems as systems of communication, naturally mediated by information, and whole society as the most encompassing social system, *i.e.*, the information system. Others say that the dynamic evolution of the physical world is (mere) a/the computation of information ([96], more recently [97], *etc.*). In the first place, the concepts of, for instance, *information*, its *meaning*, and its *context* have no generally agreed respective definitions.

In this chapter, we review the major discussions of information concepts in the studies that have information scientific flavor, focusing on ideas relevant to *context*. This review should clarify the possibility of explicating the *context* as a scientific research. Each of the description is detailed to some extent, to elucidate the point of the review and to offer the readers a glimpse into the discussion of information concept. Referred studies are limited to those that can be regarded a classic in each field, in order to avoid stepping into the latest, fragmented technical details that are too much for our purpose.

2.1.1   Concept of Information

The usage of the concept of information is very broad and complex. It ranges from bit [24] to society [114]. It is discussed both in the objectivistic view and in the subjectivistic view. It has been studied in natural and social sciences, in humanities, and in various interdisciplinary approaches. Merely a few general references for the concept of information are listed here again:[4–8].

In this chapter, we do not argue the issues such as the precise definition of the term *information*, information is whether an existing "thing"or interpreted and emerged "event", objectivity/subjectivity/intersubjectivity of information, and so on. This is because the concept of information is defined appropriately, if necessary, in each study that is referred in this chapter, and because, in most cases, the concept of information seems to be agreed with the common sense of the expected readers and such a rough understanding is expected not to introduce any difficulty. Those who are interested in such a topic should consult above mentioned references, especially [7].

In the following sections, though the definitive definitions are not required, we use several labels to indicate that each of them is used in a theory–neutral, limited meaning, in order to make explanation concise:

<**Information**> (Mental)[1] change of a certain kind which is occurred in a subject, *i.e.*, receiver of the information, with the observation of the external world;

<**(Information) Carrier**> External thing/event which is the source of the information;

<**(Information) Context**> Information is affected by many factors. The factors other than the information carrier are vaguely called *(information) context*. In a narrow sense, in contrast to the carrier without which information may not exist, the context is what may be changed without affecting the fact that the information itself is received by a subject, as long as the same carrier is observed. In a broad sense, context includes something like the background which enables to make a difference on the carrier, or a difference identifying the carrier itself.

---

[1]Note that this study does not assume that the subject always has mentality. The receiver of the information might be a "thing" in this study.

With wording of this chapter, the arbitrariness of the distinction between the carrier and the context is inevitable to some extent. In the researches that focus on the distinction, we must define the meaning of these words more precisely in each more restricted circumstance.

2.1.2 Viewpoints in this survey

To compare various theories from diverse fields in a unified way reflecting the author's view, we should bear some key questions in mind:

In each area surveyed,

1. What constituent of the theory is corresponding to the <Context>?
   — In each surveyed theory, the author identifies a constituent of the theory as the <Context>. Regardless of the role of the constituent in the theory, it is bravely interpreted.

2. Which difference in the Bateson's words is the constituent corresponding to?
   — As is mentioned in Section 1.3, Gregory Bateson [37] defines information as follows: "information is ... difference which makes a difference." The identified <Context> will correspond to one of (or, both of) the Bateson's differences. His first difference is related to identification, or articulation of the <Information Carrier>. The difference enables the <Carrier> to be identified and to try to fulfill its duties. His second difference is related to the <Information> received. By the difference, <Information> will be differently received, even if it is carried by the same <Carrier>.

3. Is it exogenous or endogenous?
   — Exogenous here means the <Context> is given from the outside of <Information Carrier>. In contrast to this, endogenous <Context> means that it is formed by the <Carrier> itself.

In addition to answering these questions, the role of the <Context> in respective area, and the possibility of the synthetic (rather than analytic) applications are discussed.

2.1.3   Survey Areas

Referring to the other surveys [7, 9, 10, 14], the author sorts out the following areas that seems to be fairly related to this thesis:

- Syntactic Approaches to Information

    - Thermodynamics
    - Mathematical theory of communication
    - Algorithmic information theory

- Cybernetics

- Semantic Approaches to Information

    - Bar-Hillel=Carnap "Semantic theory of Information"
    - Dretske "Theory of Semantic Information"
    - Barwise and Perry "Situation theory and Situation Semantics"
    - Barwise and Seligman "Information Flow"

For a general, more comprehensive survey, the readers should consult [7].

2.2   Syntactic Approaches to Information

In syntactic approaches, the concept of information often leads us to quantitative understanding of information at the sacrifice of the aspects which are popularly understood as *meaning*. In this section, following three major roots of such information measurement theories are surveyed:

1. Thermodynamics,

2. Shannon's Mathematical Theory of Communication,

3. Algorithmic Information Theory.

For more general introduction in the respective field, besides the mentioned in the following, please refer to the standard texts in each field such as [20, 22, 23] for thermodynamics, [26] for Shannon's theory.

2.2.1   Thermodynamic Entropy and Second Law

In nineteenth century, physicists were unveiling the heat phenomena, and became to understand that the origin of heat is the *randomness* of the movement of the elements included in the system.

Nicolas Léonard Sadi Carnot [16] studied the efficiency of heat engines on a mathematical footing, and argued the relation among heat, power and engine-efficiency in a cycle, which is called by his name today. He found a law: "No engine operating between two *heat reservoirs* can be more efficient than a Carnot Engine operating the same two reservoirs."

Rudolf Clausius [17] situated on the work of Carnot, introduced the concept of *entropy*, or energy lost to dissipation. He could state the basic ideas of the second law of thermodynamics using entropy: "In an isolated system, a process can occur only if it increases the total entropy of the system."

Ludwig Boltzmann [18] connected the entropy with probability. He introduced the formula for the entropy S:

$$S = k \log W \tag{2.1}$$

where $k \sim 1.38065 \times 10^{-23} JK^{-1}$ is Boltzmann's constant, and the logarithm is taken to the natural base e. W (Wahrscheinlichkeit = probability) is the frequency of occurrence of a macrostate, a macroscopic state of a system. More precisely, W is counted as the number of possible microstates corresponding to the macrostate — the number of (unobservable, but identifiable) microscopic "ways" the (observable) macroscopic thermodynamic state of a system can be realized by assigning different positions and momenta to the various molecules.

Kinetic understandings of heat phenomena won a sweeping victory, but also raised some annoying questions. Among them, Maxwell's demon is an 1867 thought experiment by the Scottish physicist James Clerk Maxwell, meant to raise questions about the possibility of violating the second law of thermodynamics. Leó Szilárd [19], and later, Léon Brillouin [21] suggested "solutions" emphasizing the importance of the role of measurement and memory of the observer.

2.2.1.1 <Context> in Thermodynamics

In physics, information is a measure of the order (or disorder) of physical phenomena, which is derived from statistical considerations. Namely, the definition of the information in physics is the Boltzmann's formula 2.1 or similar formulae, which calculates a quantity from particular physical situation. Thermodynamics enables us to discuss the physical phenomena that are in fact <Information Carriers> via analyzing thermal phenomena reflecting the randomness of affairs. Observation of physical states, distinguishability of micro/macro state, memory of observer, and so on, are understood as necessities to analyze information, and with these theoretical components, the entropy is found out as the concept corresponding to the information. <Information> in general sense is grasped as the transfer or correlation between the <Carrier> and the subject, and all the discussions and explanations are reduced to ones of the entropy.

In thermodynamics, <Information Context> may be said to correspond to the heat reservoirs of the system, physical surroundings of the system, the memory of the observer, and the ways to articulate the state of the system into some macro states. They may be regarded as ones that function as static exogenous/transcendental standard. That is, they are corresponding to the Bateson's first difference. For example, Carnot's heat reservoir works as a background to make the occurrences of thermal phenomena identifiable. This background as a static zero-level is necessarily required as presuppositions and rarely play an explicit role in the model. Thus we can say that the <Information Context> in thermodynamics is separated from the system under consideration, and that the system (*i.e.*, <Information Carrier> and <Information>) is discussed excluding the <Context> in thermodynamics.

2.2.2 Mathematical Theory of Communication

Claude Elwood Shannon and Warren Weaver describe a today's most well-known common communication model [25]. The model is consisted of the sender-channel-receiver and signals are transferred (with disturbances by noises) from the sender to the receiver (Figure.2.1).
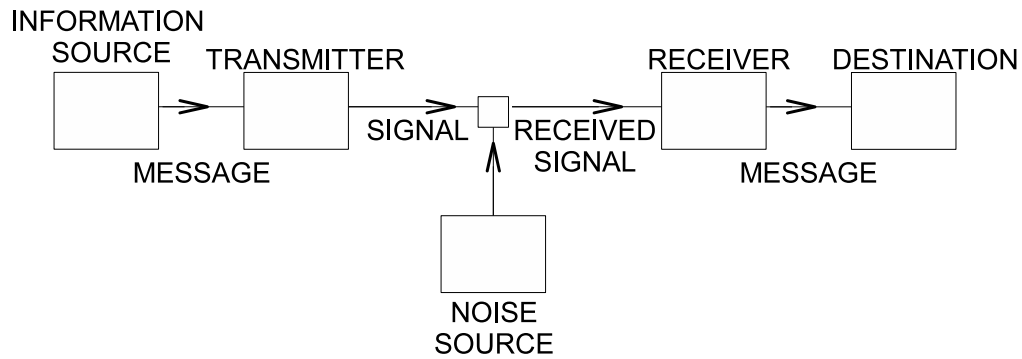
FIGURE 2.1: Shannon's schematic diagram of a general communication system

Shannon used the word *information* to mean the number of different messages possibly transferred through the channel as in the following quote:

> The word information, in this theory, is used in a special sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning. [25].

Information in Shannon's sense was a measure of orderliness (as opposed to randomness) in that it indicated the number of possible messages from which a particular message to be sent was chosen. One can say that Shannon initiated a new science of information by providing a precise definition of information, a way of measuring it, and theoretical results about the limits to its transmission [4].

Shannon's theory of information entropy, which measures information quantitatively, is as follows[2]. To develop his theory, Shannon examines the output of a discrete information source that generates a message as a Markov process. Each possible message that can be generated has associated with it a probability p_i of its occurrence and there are n such messages. Shannon searches for a function $H(p_1, p_2, \ldots, p_n)$ that will quantify our reduction in uncertainty on receiving the message subject to the following desiderata:

1. H should be continuous in p_i.

2. If all the p_i are equal, the H should be a monotonic increasing function of n.

---

[2]Following description in this paragraph is essentially an excerpt of [10].

3. Each event (symbol generation) should be capable of being linearly decomposed into two or more constituent events with their own proportional probabilities [24].

Shannon concludes that the only function satisfying the criteria is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \tag{2.2}$$

where K is a constant according the units chosen (*i.e.*, the base of logarithm used).

2.2.2.1 <Context> in Mathematical Theory of Communication

In Shannon's communication theory, <Information Context> is the set of all possible messages. In his theory, (the quantity of information)=(the <Carrier>'s capability of transferring different messages) and his information corresponds to Bateson's first difference. Bearing the communication, *i.e.*, engineering, in mind, one seems seldom care about the <Context> of the communication system, but it is embedded as the agreement of the sets of messages at Transmitter and at Receiver. It is trivial on the engineered communication system, but in general, it is clear that this is not the case, if you imagine a conversation between two persons who have different mother tongues.

That is, <Context> in the Shannon's theory is static, given, exogenous. Shannon's entropy, *i.e.*, quantity of information is related to <Information Context> because the probability p_i of message occurrence is measured against the whole messages, namely, the <Context>. The set of all messages works as a background for distinguishing a particular message. But as already mentioned, when all the messages are given, as typically in an engineered communication system, <Information Context> is determined and fixed, so the <Context> rarely becomes an explicit constituent of the theoretical model.

2.2.3 Algorithmic Information Theory

Algorithmic information theory was born out from the gaps between intuitive notions concerning regularity and answers provided by the standard probability theory. In order to illustrate these gaps, consider a coin-toss experiment in which a fair coin is flipped 23 times in a trial. Now imagine we obtained the following results from three trials [10]:

Trial 1: 1 0 1 1 0 0 0 1 1 1 1 0 0 1 1 0 0 0 1 1 0 1 0

Trial 2: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Trial 3: 0 0 1 1 0 1 1 1 0 0 1 0 1 1 1 0 1 1 1 1 0 0 0

Probability theory tells us the probability of obtaining the sequence shown in Trial 1 is 2^{-23}, and the value is same for Trial 2 and 3. However there is something unsettling about accepting that the first and second results are equally probable. The first result "looks" more random. The third trial may at first glance appear random, but there is also regularity in this sequence. It is a list of the integers 0 to 8 in binary format[3]. These gaps between the value of probability and the intuitive notions on regularity can not be resolved by the use of a probability method assuming equiprobable occurrences. The information embedded in the order in the sequences is mostly ignored there. Something that takes into account the generation process is required. Algorithmic information theory represents an attempt to meet that need.

The theory was proposed separately by Ray J. Solomonoff [27–29], Andrey Nikolaevich Kolmogorov [30], and Gregory John Chaitin [32].

2.2.3.1 Richard von Mises – "Interpretation of randomness"

Randomness lies at the heart of considerations of the algorithmic information theory. Richard von Mises' interpretation of randomness [33] is an important starting point for the development of the algorithmic information theory.

Von Mises' definition of randomness in terms of binary strings is:

> We say that it (an infinite sequence of zeros and ones) possesses the property of randomness if the relative frequency of the 1's (and therefore also that of 0's) tends to a certain limiting value which remains

---

[3]That is, the beginning of so-called Champernowne's Number in binary format.

unchanged by the omission of a certain number of the elements and the construction of a new sequence from those which are left. [34].

His definition has been clarified and improved with other researchers such as Alonzo Church and Abraham Wald. The amended version of von Mises' definition – *Mises-Wald-Church randomness* can be summarized as follows [10].

A binary string is *Mises-Wald-Church random* if and only if:

1. The relative frequency of all attributes of a collective approaches a definite value between 0 and 1 in the limit as the collection length approaches infinity;

2. Every subsequence chosen according to some admissible place selection function has the same limiting relative frequency as the entire collection;

3. A place selection function is admissible if and only if it is a member of the set of recursive functions.

2.2.3.2 Solomonoff – "Probability in inductive inference"

Solomonoff [27] examines the problem of inductive inference: "given a long series of symbols represented by T, what is the probability that it will be followed by a subsequence represented by an element a? That is, What is $p(a|T)$?" and reaches the concept of a binary description.

> Suppose that we have a general purpose digital computer M_1 with a very large memory. ... Any finite string of 0's and 1's is an acceptable input to M_1. The output of M_1 (when it has an output) will be a (usually different) string of symbols, usually in an alphabet other than the binary. If the input string S to machine M_1 gives output string T, we shall write

$$M_1(S) = T. \tag{2.3}$$

> Under these conditions, we will say that "S is a description of T with respect to machine M_1 ." If S is the shortest such description of T, and S contains N digits, then we will assign to the string, T, the a priori probability, 2^{-N}. [27].

Under Solomonoff's model, any observations or descriptions, namely, any information could be represented as binary strings with a priori probability.

2.2.3.3 Kolmogorov –"Kolmogorov complexity"

Kolmogorov directly considered the issue of the quantification of information. He attempts to define the information which is carried by object x, and tells us about object y [30].

By means of program p which generates string y from string x, *relative complexity* of y given x can be defined as follows:

$$K_\varphi(y|x) = \begin{cases} \min_{\varphi(p,x)=y} l(p) \\ \infty, \text{ if there is no } p \text{ such that } \varphi(p,x) = y \end{cases} \tag{2.4}$$

where φ is a partially recursive function which associates object x and program p with object y, and $l(p)$ is the length of the program p.

And also, one can specify the intrinsic information of an individual object:

$$K_\varphi(y) = \begin{cases} \min_{\varphi(p)=y} l(p) \\ \infty, \text{ if there is no } p \text{ such that } \varphi(p) = y. \end{cases} \tag{2.5}$$

That is, the information content of a string is the minimum length of a description – an instruction or program, p, to be executed by a universal computer, φ.

Though there is no provision made for determining the shortest possible p, Kolmogorov's formula provides the unambiguous specification of the generation of a sequence given another sequence as input.

2.2.3.4 Chaitin – "Algorithmic incompressibility"

Chaitin determined that a *random* string is one of that can not be algorithmically compressed. He considers a Turing machine that outputs a binary string [31]. In his theory, there are two functions which play the fundamental roles.

L, the first function, is defined on the set of all finite binary sequences S as follows: An N-state, 3-tape-symbol bounded-tansfer Turing machine can be programmed to calculate S if and only if $N \geq L(S)$.

The second function $L(C_n)$ is defined as

$$L(C_n) = \max_{S \text{ of length } n} L(S) \tag{2.6}$$

where the maximum is taken (as indicated) over all binary sequences S of length n. [31].

The symbol C_n conceptually denotes the most complex binary sequence of length n.

Chaitin argues that the patternless or random finite binary sequences of length n are those sequences S for which $L(S)$ is approximately equal to $L(C_n)$. He does not provide a rigorous proof of this, however, he attempts to "make it plausible by proving various results concerning what may be termed statistical properties of such finite binary sequences" [31]. These properties include familiar tests for randomness such as Simple Normality and Von Mises Place Selection.

Chaitin's work is essentially complementary to the work done by Solomonoff and Kolmogorov. Chaitin proposes the following model,

Formally, a binary computing machine is a partial recursive function M of the finite binary sequences which is finite binary sequence valued. The argument of M is the program, and the partial recursive function gives the output (if any) resulting from that program. $L_M(S)$ and $L_M(C_n)$ (if the computing machine is understood, the subscript will be omitted) are defined as follows:

$$L_M(S) = \begin{cases} \min_{M(P)=S} (\text{Length of } P), \\ \infty, \text{ if there is no such } P, \end{cases} \tag{2.7}$$

$$L_M(C_n) = \max_{S \text{ of length } n} L_M(S). \tag{2.8}$$

In this general setting the program for the definition of a random or patternless binary sequence assumes the following form: The patternless or random finite binary sequences of length n are those sequences

S for which $L(S)$ is approximately equal to $L(C_n)$. The patternless or random infinite binary sequences S are those whose truncations S_n are all patternless or random finite sequences. That is, it is necessary that for large values of n, $L(S_n) > L(C_n) - f(n)$ where f approaches infinity slowly. [31].

$L(S)$ is in essence Kolmogorov's $K_\varphi(y)$.

2.2.3.5 <Context> in Algorithmic Information Theory

In the algorithmic information theory, information or quantity of information is based on the algorithms generating each string, and is naturally endogenous. It is a number representing the degree of randomness or order of each string generated from a given set of symbols. The algorithmic information theory defines randomness in general endogenously, and measures the orderliness from that level. This relation between the randomness and order is in reverse to the case of the thermodynamics or Shannon's entropy. The <Context> in the algorithmic information theory is constructed by the interrelation of symbols in a string identifying each other. This is much different from the case of Shannon's communication theory in which <Context> is exogenously constructed. However, the algorithmic information theory describes the aspect of information which is comparable with Shannon's, and is also corresponding to Bateson's first difference.

2.3 Cybernetics

In his Cybernetics [35], Norbert Wiener founded cybernetics and pointed out the importance of the feedback in autonomous controlling. Although its developments encompass various fields such as engineering, natural sciences, and social sciences, they have common use of high level concepts such as *order*, *organization*, *complexity*, *hierarchy*, *structure*, *control*, and *information*, investigating how these are manifested in systems of different types, as written in [43]. Fundamental to all of these relational concepts is that of difference or distinction. That is, cyberneticians are not interested in a phenomenon in itself, but only in the difference between its presence and absence, and how that relates to other differences corresponding to

other phenomena. Using these concepts, they commonly aim at understandings of the purposiveness, or goal-directed behavior, of a system, which are independent of details of the system, and also try to control based on these understandings.

Since early 1970's, a part of the cybernetics which is especially interested in organizations and social systems has been developed as the *cybernetics of cybernetics* or *second-order cybernetics*[4]. It investigates the construction of models of cybernetic systems itself. Second-order cybernetics is aware that the investigators or observers are part of the system and emphasizes on the importance of self-referentiality, self-organizing, the subject-object problem, *etc.*
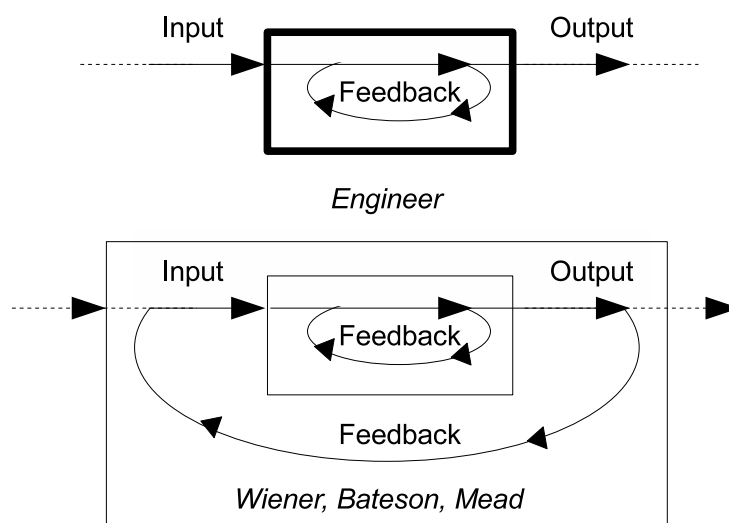
FIGURE 2.2: Second-order Cybernetics [38]

Succeeding the shift from first-order cybernetics' attention to homeostasis as an autonomous self-regulation in mechanical and informatic systems, to second-order concepts of self-organization and self-production in embodied and metabiotic systems, there emerges the lines of thought so-called *neocybernetics* [48]. Neocybernetics observes that cognitive systems are operationally bounded, semi–autonomous entities coupled with their environments and other systems [48]. Some authors in the field of information studies have been adopting the thought of neocybernetics, especially the idea of autopoiesis [39]. Nishigaki [46, 47], for instance, advocates the Fundamental Informatics, in the line of the neocybernetics. However, as has been pointed out repeatedly even by the insiders of the autopoiesis school, it is still in its gradual development, premature and not well understood [42, 45, 49, 50]. The concepts in it have been criticized and have been interpreted in diverse and

---

[4]See [44], for instance.

even contradictory ways. Moreover, it still merely has narrow connection with the existing empirical sciences. Though these critiques are for the autopoiesis theory and neocybernetics in general, the situation is similar for the studies in the field of infomation studies such as [46, 47][5]. Thus, though the dynamic, monadic[6] picture of information in neocybernetics attracts the interest of the author of this thesis[7], conforming to our principle of this chapter, the review of the neocybernetics in information sciences is left for the future.

We can say that the most fundamental contribution of cybernetics is its explanation of purposiveness, or goal-directed behavior, an essential characteristic of mind and life, in terms of control and information [43].

2.3.1 \<Context\> in Cybernetics

\<Information Context\> that is the most characteristic in cybernetics may be the one that is constructed by feedbacks. Feedbacks are generated according to the output, and may be interpreted as a self-created \<Context\> which affects the "meaning" of the input of the system (*i.e.*, an endogenous \<Context\> corresponding to Bateson's second difference). In second-order cybernetics, an observer is a part of the \<Context\> of the observed system, and there is a self-referential relation between the observer and the system. Second-order cybernetics points out that the observer, primarily as a receiver of information, inevitably affects the "meaning" of information. There is a mutual construction of \<Context\> between the observed system and the observer.

---

[5]See [49], for an example of criticism of Nishigaki's theory as an autopoiesis theory.

[6]In the Leibnizian sense.

[7]This is the reason why neocybernetics is mentioned in this section without a review in the view of the present study.

2.4 Semantic Approaches to Information

There are linguistic or logical semantic theories in which <Context> plays an important role, in addition to the linguistic pragmatics which are briefly mentioned in Chapter 1.

2.4.1 Bar-Hillel=Carnap – "Semantic theory of Information"

Yehosua Bar-Hillel and Rudolf Carnap developed a semantic theory of information [51, 52] in which they distinguish between information and amount of information within a linguistic framework.

In Shannon's communication theory, the amount of information is measured as statistical rarity, *i.e.*, the frequency of their occurrence in a transmission. In contrast to his theory, in their semantic theory of information, Bar-Hillel and Carnap measure the amount of information by a logical or inductive probability function over the contents of the sentence under consideration. Information content of a statement, which is a concept introduced by them, is the class of those possible states of the universe, which are excluded by the statement. They also introduced concepts such as:

State-descriptions It is a conjunction containing for every atomic sentence that can be formed in a language, either its affirmation or its negation but not both.
(It gives a complete description of a possible state of the universe of individuals with respect to all the properties and relations expressed by the predicates of the system.)

L-true (logical true) When the truthfulness of a statement in a given semantical system is known by reference to either the logical form or to the descriptive terms in the statement, without referring to extra-liguistic *facts*, the sentence is called L-true.

State-descriptions, L-true, and other various "L-" concepts are introduced to discriminate the (logical) meanings which are determined only within a semantical

linguistic system, without referring to the outer world from the "factual" meanings. Using these concepts, Bar-Hillel and Carnap provided various explicata for the presystematic concept (or concepts) of amount of semantic information.

To roughly grasp their idea, their derivation of the amount of information is briefly followed here [58]:

Let $m(i)$ be the logical probability of the statement i (Logical probability is the estimate of statistical probability, and it is the measure of the degree of confirmation). Then the quantity $1 - m(i)$ is the measure of the content of i, which may be called the *content measure of i*, denoted $cont(i)$. Thus,

$$cont(i) = 1 - m(i).$$

However, this measure does not have additivity properties, because *cont* is not additive under inductive independence. Thus insisting on additivity on condition of inductive independence, the authors propose another set of measures for the amount of information, which they call *information measures* for the idea of the amount of information in the statement i, denoted $inf(i)$, and which they define as:

$$inf(i) = \log \frac{1}{1 - cont(i)}$$
$$= -\log m(i)$$

This is analogous to the amount of information in Shannon's mathematical theory of communication but with inductive probability instead of statistical probability.

2.4.2   Dretske – "Theory of Semantic Information"

Fred I. Dretske developed a theory of semantic information [53] based on the distinction between information and meaning.

> It is common to think of information ..., as something that depends on
> the interpretive efforts – and, hence, prior existence – of intelligent life.
> According to this view, something only *becomes* information when it is

assigned a significance, interpreted as a sign, by some cognitive agent.

...

This is one way of thinking about information. It rests on a confusion, the confusion of *information* with *meaning*. Once this distinction is clearly understood, one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive process. [53].

The flow of such information should be based on the following *Xerox principle*.

Xerox principle If A carries the information that B, and B carries the information that C, then A carries the information that C [53].

In combination with Shannon's communication theory, Dretske states three conditions that a definition of information must satisfy, namely:

(A) The signal carries as much information about s as would be generated by s's being F.

(B) s is F

(C) The quantity of information the signal carries about s is (or includes) that quantity generated by s's being F (and not, say, by s's being G)

Moreover, he defines the information contained in a signal that simultaneously satisfies these three conditions:

Information content A signal r carries the information that s is F = The conditional probability of s's being F, given r (and k), is 1 (but, given k alone, less than 1).

The parenthetical k is meant to stand for what the receiver already knows (if anything) about the possibilities that exist at the source.

So, for example, if one already knows that s is either red or blue, a signal that eliminates s's being blue as a possibility (reduces this probability to 0) carries the information that s is red (since it increases this probability to 1). For someone who does not know that s is either red or blue (given what they know, it could be green), the *same signal* might *not* carry the information that s is red. [53].

Notice that, in this definition, the letter s is understood to be an *indexical* or *demonstrative* element referring to some item at the source. The definition defines what philosophers might call the signal's *de re* (versus *de dicto*) informational content.

2.4.3   Barwise and Perry – "Situation theory and Situation Semantics"

Kenneth Jon Barwise and John Perry originated situation theory and situation semantics [54]. At the heart of situation semantics lies, as the name suggests, the notion of *situation*. According to Perry [56], in its earliest forms, the central ideas of the situation semantics were:

- *Partiality* – Situations are contrasted with worlds; a world determines the answer to every issue, the truth-value of every proposition. A situation corresponds to the limited parts of reality we in fact perceive, reason about, and live in. What goes on in these situations will determine answers to some issues, but not all.

- *Realism* – Basic properties and relations are taken to be real objects, uniformities across situations and objects, not bits of language, ideas, sets of n-tuples or functions. In *Situations and Attitudes* [54], *courses of events* are partial functions from sequences of locations, relations and objects to truth-values. Complex properties and various types of objects were full-fledged objects, entering into courses of events.

- *The Relational Theory of Meaning* – The meaning of an expression ϕ is conceived as a relation between a discourse situation, a connective situation,

and a described situation, written

$$_{d,c}\left[\left[\phi\right]\right]_e.$$

The meaning of "I am sitting next to David," for example, would obtain between courses of events d (discourse situation); c (speaker connections), and e (described situations) if there are individuals a and b such that i) in d, a is the speaker of the sentence; ii) in c, a's use of "David" is used to refer to b; iii) in e, a is sitting next to b.

Driven by these ideas, [54] embodied the basics of realistic model-theoretic semantics of natural language.

> The basic idea of situation semantics is that in thought and action we use complexes of objects and properties to directly and indirectly classify parts and aspects of reality, or situations. This sort of realistic classification is more basic than linguistic classification, and underlies it. [56].

Consider a simple dialogue:

```
'What happened in the woods this afternoon?'
'Jackie broke her leg.'
```

> The question concerns a certain situation, a bit of reality: the events in the woods this afternoon. The answer directly classifies the situation in terms of an object (the dog Jackie) and a property (acquiring a broken leg). We classify situations by what goes on in them; which properties objects have, and the relations they stand in to one another in virtue of the events that comprise the situation. [56].

Consider the issue of whether Jackie broke her leg at a certain time t. The possibilities or *state of affairs* can be represented as:

$$\sigma : \langle\langle breaks - leg, t, Jackie, 1\rangle\rangle$$

or

$$\sigma' : \langle\langle breaks - leg, t, Jackie, 0\rangle\rangle \,,$$

corresponding to whether she did or did not. The factuality of whether or not Jackie broke her leg will naturally be determined by what goes on in the whole world. But also much smaller *situations* will determine the state of affairs. Let s be the situation in the woods this afternoon. Then,

$$s \models \sigma$$

i.e., *s supports* σ. In situation theory, various objects are built from the basic interplay of situations and states of affairs, permitting complex and abstract ways of classifying situations.

The points of situation theory are summarized as follows[8]:

- It captures the meaning of representations (utterance, diagrams, *etc.*) created by humans as a continuum over the information existing in the nature.

- It captures the situation individually in which a certain representation means a particular information. (context dependency of meaning)

- It captures every representation as a respective situation which has various properties.

2.4.4   Barwise and Seligman – "Information Flow"

Information Flow is the mathematical model proposed by Kenneth Jon Barwise and Jerry Seligman [55] as a formalization capturing the information flowing in any sorts of distributed systems ranging from physically distributed control systems to abstract ones. This theory first presents a model which shows the gap between the information sender and receiver. The basic idea of this model is that information exchange within a distributed system results from *regularities* that enable the information possessed by one component of the distributed system to be understood by another component. Such a notion is formalized in *Channel Theory*.

---

[8]Loose translations of [57]

In Channel Theory, a channel connecting components that participating in the information exchange by means of mappings among types and tokens. The interacting components might have their own vocabulary, *i.e.*, they might have different types and tokens and thus different classification between types and tokens; they might have different constraints regulating the behavior of the information that they are conveying. Capturing all the above notions, mathematical formulations are defined as follows:

In Channel Theory, each component of a distributed system is represented by a classification $A = (tok(A), typ(A), \models_A)$, consisting of a set of tokens $tok(A)$, a set of types $typ(A)$, and a classification relation $\models_A \subseteq tok(A) \times typ(A)$ that classifies tokens to types.

The flow of information between components in a distributed system is modeled by the way the various classifications that represent the vocabulary and context of each component are connected with each other through *infomorphisms*. An infomorphism $f = (f^\wedge, f^\vee) : A \rightleftarrows B$ from classification A to classification B is a contravariant pair of functions $f^\wedge : typ(A) \to typ(B)$ and $f^\vee : tok(B) \to tok(A)$ satisfying the following fundamental property, for each type $\alpha \in typ(A)$ and token $b \in tok(B)$:

$$
\begin{array}{ccc}
\alpha & \xmapsto{\;\;f^\wedge\;\;} & f(\alpha) \\
{\scriptstyle\models_A} \Big\updownarrow & & \Big\updownarrow {\scriptstyle\models_B} \\
f(b) & \xleftarrow[\;\;f^\vee\;\;]{} & b
\end{array}
\tag{2.9}
$$

The distributed system $\mathscr{A}$ consists then of an indexed family $cla(\mathscr{A}) = \{A_i\}_{i \in I}$ of classifications together with a set $inf(\mathscr{A})$ of infomorphisms all having both domain and codomain in $cla(\mathscr{A})$.

A basic construct of Channel Theory is that of a *channel*—two classifications A and B connected through a *core* classification C via two infomorphisms f and g:

$$
\begin{array}{ccccc}
 & & typ(C) & & \\
 & f^{\wedge} \nearrow & \vdots & \nwarrow f^{\wedge} & \\
typ(A) & & \models_C & & typ(B) \\
\vdots & & \vdots & & \vdots \\
\models_A & & tok(C) & & \models_B \\
 & f^{\vee} \swarrow & & \nwarrow f^{\vee} & \\
tok(A) & & & & tok(B)
\end{array}
\qquad (2.10)
$$

This basic construct captures the information flow between components A and B.

2.4.5   <Context> in Semantic Theory of Information

As presented in this section, all of the theories of semantics of information attack to the *situated* meaning, that is, meaning in <Context>, in the framework of logic.

Semantic theory of Information by Bar-Hillel and Carnap presents a way to explicate logically/scientifically the meaning of information. In their theory, <Context> for a particular sentence is divided into the *logical* <Context> which is determined by the interrelation of sentences in the semantical system, and the *factual* <Context>. The logical <Context> in this theory is, on the contrary to the Shannon's case, endogenous, but it is static and fixed, as is in the Shannon's case.

Dretske attributes information to the regularities among states of affairs. By doing this, information may be separated from the *meaning*, and discussed as objective commodities. According to Dretske, information is always relative to "a receiver's background knowledge"(*i.e.*, k). That is, information recognized by a receiver depends on the receiver's background knowledge, *i.e.*, <Context>. Such a <Context> corresponds to Bateson's second difference, is exogenous and can yield some multiple interpretations under the constraint that the state of affair is fixed.

Situation semantics by Barwise and Perry explicitly addresses the situations, namely, <Context>. In situation semantics, the meaning of an expression is conceived as a relation between a discourse situation, *i.e.*, <Context>. Their theory and a further developed theory – Channel Theory – by Barwise and Seligman are based on the idea of regularities between types of situations. Information is not a property of facts but it is constraint or <Context> dependent. In contrast to Dretske's concept of information which is within a cognitivistic framework, the theory of situation semantics defines information within a realistic and not just cognitivistic framework. Information contents are not dependent on the knowledge of the receiver, Dretske's k, but on types of situations. However it can be said that all these theories address the exogenous, dynamic situations, *i.e.*, <Contexts> and the contents of information are interpreted according to the <Context>.

2.5 Summary of the context in various information theories

Figure.2.3 presents the summary of how <Information Context> is addressed in information theories reviewed in this chapter.

Legends:
x   corresponds to
++ has advantages
+   addressed
-   has weakness

	Bateson's first difference	Bateson's second difference	Exogenous / transcendent	Endogenous / immanent
This thesis	x	x		++
Thermodynamics	x		+	-
Math. Theory of communication	x		+	
Algorithmic information theory	x			++
Cybernetics		x		+
Bar-Hillel=Carnap		x		+
Dretske		x	+	
Situation Theory Barwise & Perry		x	+	
Information flow Barwise & Seligman		x	+	

FIGURE 2.3: Summary of the context in information theories

All these prior works may be considered to be those which address some of the fundamental aspects of information, albeit in an insufficient manner. In the first three theories, the context constitutes the base of the quantitative measure in each theory, such as the probability space. It is for the Bateson's first difference, *i.e.*, difference which makes difference. In the thermodynamics and the mathematical theory of communication, the context is formed exogeneously with those that are assumed to be distinguishable without any other information beforehand. In the Algorithmic Information Theory, the context is constructed endogenously by the interrelation of symbols in a string identifying each other. However, the distinction of the symbols is assumed also in the Algorithmic Information Theory. Thus these theories may be considered to be those that do not undertake the vicious circularity seriously as a foundational theories of information. The context in the cybernetics is constructed endogeneously by feedbacks. Although this is interesting from our perspective, disappointingly, the cybernetics mainly concerns only with the Bateson's second difference. The distinction of the inputs, which corresponds to the Bateson's first difference, is considered to be assumed. The semantic approaches adopt logic as their base. Logic naturally presumes the distinction of its statements. In other words, the state of affairs that the logical statement describes is assumed to be distinguishable. Since the state of affairs corresponds to the Bateson's first difference, they may be considered to be concerned only with the Bateson's second difference.

As the readers will see, in this thesis, both of the Bateson's differences are incorporated into the model of information phenomena as essential constituents. Each of the differences is born in respective endogenous context that is built within the IGUS hysteretically[9]. This may bring a fresh view for the fundamental role of contexts in information phenomena.

---

[9]In other words, each of the contexts is built in dependence not only on current information related stimulus for the IGUS but also on past stimuli that have been received by the IGUS.

Chapter 3

Extensional Information Articulation from the Universe

3.1 Information as physical phenomena

Many researches point out the essential relativeness of information[1]. It is natural, especially in the physical view adopted in this study, because an IGUS[2] (part of the universe) inevitably interacts with surrounding systems (other parts of the universe) in its own way. But then, at the fundamental level, how can an IGUS articulate a specific uniformity (e.g., the color of green) in a part of the universe? The meaning of the word *articulate* here is *"to divide into fragments meaningfully arranged."* In principle, nothing other than the IGUS itself can provide information on how to articulate the universe.

Recall the episode in 1.1.1. The episode indicates that even the color of green might not be obvious information. We consider such an issue in this chapter. How does an IGUS articulate and categorize its surrounding environment, and how does it assign an identifier for the properties of the part, *i.e.*, a collection of attributes, to the articulated part of the environment? How can such information be obtained, or generated by a physical system? After all, this universe is a bunch of physical interactions. A system that is going to utilize information from the universe must extract the information through these physical interactions. In other words, if

---

[1]For instance, see the Locality Principle and other ontological principles in [15].
[2]Information Gathering and Using/Utilizing System. See 1.5.

we borrow terminology from logic, we might be able to say that the system must extract the information fully *extensionally*.

To our knowledge, few studies have been conducted to investigate the fundamentals of information in such a view.

One might consider that, in physics, information is tightly connected to physical phenomena that are obviously extensional. The distinction between the physical phenomena and the informational phenomena is subtle. Even in the study of physics, the materiality or the *thinginess* of nature has been pushed away in recent decades. Many authors have already declared that reality basically is an informational phenomenon (for example, [11, 13, 97–101]).

However, as is shown in the episode in 1.1.4, a fundamental study of information has to deal with non–causal phenomena on which ordinary physics does not study. Physics mostly focus on the information phenomena as we, human–IGUSes understand. Physics scarcely ever describes information *for a material* in a way such as for human–IGUSes. In other words, it scarcely describes information for a material immanently. Thus, in the view of this chapter, such an extensional approach to the fundamentals of information should cover broader scope than the information studies in physics.

In information theories, the approach developed by Alexander Chechkin (see [15]) might be regarded as one of such studies that have the nearest viewpoint.

> His approach is based on two main notions: an object and information about object. In his theory, properties of objects are represented (and denoted) by subsets of those objects that has this property, *i.e.*, each property is a subset of all objects and each subset is a property of objects. It is the classical set-theoretical approach to the concept of property. [15].

To represent properties of objects extensionally, it seems natural to use the sets of objects, as Chechkin. Needless to say, in many studies in various disciplines, a property of objects, or equivalently in this chapter, an attribute of objects, is treated as a set of objects. The attribute green is (represented by) all the *green-ish* objects. When someone asks "what is green?," one may answer by indicating something greenish and saying "this is green." Something green is one

of the greenish objects one has ever met, and the attribute "green" can be defined by all the greenish objects one has ever met. Borrowing a term in linguistics, *synchronically*, this may be sufficient. That is, at a fixed point of time, all the greenish objects that one has ever sensed can represent the attribute green for him/her.

However, if one defines the attribute green by a certain (fixed) set of *green-ish* objects, there will be no room for a green thing newly accepted. Something newly accepted cannot reside in the set of *green-ish* objects, which is already fixed as "the definition of the attribute green." *Diachronically*, an attribute is not merely that which is equivalent to a set of objects. There is a case in which modeling an attribute as a set of object is not appropriate.

This study addresses such a case. We will pursue the understanding of information as a universal structure that is fully extensionally obtained from finite physical interactions.

Note that, such an understanding may also demystify a kind of the bootstrapping process of information extraction from the universe. From the extensional viewpoint of this study, as already mentioned, information is naturally considered to be essentially relative to the (information receiving) system, the same as physical interactions are essentially relative. Information relative to the system implies the existence of the *context*, that is, another referred information within the system. But, then, how can a system articulate the universe without any contextual information (*i.e.*, so to speak, *criteria*), at the most fundamental level of, or, at the very beginning of, information process? In other words, how can a system have information of its own without any presupposed another information? Though this study might be very focused and not general enough, it may be considered as an attempt to understand information from such an absolutely relative viewpoint.

Extracting/creating information from the universe via physical interventions has been discussed in the measurement theory in metrology (see [102], *etc.*). The theory is called the *extensive measurement theory*, which will be described in detail in Section 3.2. The purpose of the study in this chapter is to develop a general model of extensional information articulation process as a generalization of the extensive measurement theory.

In this chapter, the author

- presents a model of the process of extensional information articulation as a generalized extensive measurement;

- provides a detailed example of a step of the extensional informational articulation process—a model of attribute creation.

The structure of this chapter is as follows. In Section 3.2, the extensive measurement in metrology is explained in detail. Referencing the steps in the extensive measurement, a model for information articulation process as a generalized measurement is developed in Section 3.3. Section 3.4 exemplifies the concepts of the developed process by a model for attribute creation. Section 3.5 concludes this chapter.

3.2   Extensive Measurement

A prototype suitable for describing how a system extracts information from the physical universe exists in the metrological measurement theory.

Suppose that we try to measure an attribute of a physical thing, for example, the weight. First, we prepare an apparatus for measuring the attribute (a measuring instrument): an equal arm beam balance. We place a thing a on the left pan and another thing b on the right pan, and we observe which pan drops (we observe if it is balanced). Thus we can introduce an order relation between the attributes of two things a and b. When balanced, two things are virtually the "same" in the aspect of the attribute, the weight. By these equivalences, we can use *copies* to represent each specific weight, instead of using particular instances (e.g., using sands instead of apples). We can also define the *concatenated* or *composed* attributes of two or more things by placing them in the same pan. To evaluate the attribute, we select a thing as the *unit*, and if another thing is found to be equivalent with n-copies of the unit thing, the attribute value of the latter thing can be determined as n (given in the unit).

Such measurement processes have been generalized and abstracted in the extensive measurement theory.

The extensive measurement (see [102], *etc.*) was formulated by Herman von Helmholtz [103], Otto Ludwig Hölder [104], and others. The extensive measurement relates the physical measurement situations/operations to a mathematical model. The mathematical model has three primitives [102]: a nonempty set A, a binary relation $\succsim$ on A, and a binary operation $\circ : A \times A \to A$. The interpretations of these primitives are: A is a set of objects or entities that exhibit the attributes in question; $a \succsim b$ holds if and only if a exhibits, in some prescribed qualitative fashion, at least as much of the attribute as b; and $a \circ b$ is an object in A that is obtained by concatenating (or composing) a and b in some prescribed, ordered fashion.

In this article, our main concern is the way in which the extensive measurement theory builds such a mathematical model from the operations in the real physical world.

The general steps in the extensive measurement are summarized as follows [105]:

1. Suppose that there is a collection I of things or events in the real world (in which all things/events are identifiable by the measurer), and let the elements (*i.e.*, things or events) in the collection I be countable. (Countability and identifiability enable to make *n-copy* of an element, later.)

2. A comparator m is introduced as a measuring instrument. It can compare an element in I with another one. That is, the comparator m shows that whether the compared elements are indistinguishable or distinguishable from the comparator's viewpoint (that is, in the aspect of the attribute under consideration). Introduce an equivalence relation $\sim_m^I$. Elements a and b in I have a relation $a \sim_m^I b$ if and only if the comparator m shows that they are indistinguishable, and under this equivalence relation $\sim_m^I$, all the elements in I can be partitioned into equivalence classes $X = \{\alpha, \beta, \gamma, \ldots\}$. Note that the equivalence class consists of real physical tokens (e.g., apples). There can be a collection of comparators M for a measurer and the relation $\sim_m^I$ is relative to each comparator $m \in M$.

3. Assume that one can define a physical relation $\succsim_m^I$ that has a total order, and a physical operation for *concatenation* (or *composition*) $\circ_m^I$ that is associative and strictly increasing. "Physical" here means that the relation and the operation are presented as physical phenomena in the real world. From

any pair of the equivalence classes α and β, choose any elements a and b, respectively. Equivalence classes α and β have a relation $\alpha \succsim_m^I \beta$ if and only if the comparator shows a distinctive state: "when a is on the right pan and b is on the left pan, the right pan sinks (or they are balanced)," for example. Choosing the elements a, b and c respectively from the any equivalence classes α, β, and γ, the concatenation is defined as: the concatenation of equivalent classes α and β, $\alpha \circ_m^I \beta$, is a physical state in which a and b are on the same pan; and $\alpha \circ_m^I \beta = \gamma$ if and only if c is on the other pan and they are balanced. By an abuse of notation, we also write $\alpha \sim_m^I \beta$ if and only if $\alpha \succsim_m^I \beta$ and $\alpha \precsim_m^I \beta$ holds.

4. Choose an equivalence class α from X. For any positive integer n, define $n\alpha$ inductively as follows: $1\alpha = \alpha, (n+1)\alpha = (n\alpha) \circ_m^I \alpha$. This $n\alpha$ is called *n-copy* of α. A copy of α can be found operationally. As for the weight, for instance, put $x_1 \in \alpha$ on a pan and find $x_2 \in \alpha$ which is balanced with x_1. Now what is being considered is "the attribute, weight," hence x_2 can be anything, say, a pile of sand, for example. In this way, infinitely many copies of the weight of x_1, $x_3, x_4, \ldots$, can be made in theory. Then one can define 2α as a unity of x_1 and x_2, 3α as a unity of x_1, x_2, and x_3, and so on.

5. *The Archimedes' condition* is required between the elements in X. The Archimedes' condition is: for any pair of equivalent classes α and β, there exists a positive number n such that $n\alpha \succ_m^I \beta$. Here $\alpha \succ_m^I \beta$ means $\alpha \succsim_m^I \beta$ and not $\alpha \sim_m^I \beta$. Though this condition has issues logically (because it cannot be described in first order predicate logics) and practically (because it implies infinite operations), it has great values to construct the theory.

6. Corresponding to the above physical situations, a mathematical model, *i.e.*, a formal system is set up. On the mathematical set A, introduce a totally ordered algebraic relation $\succsim_m$ that is derived from the corresponding relation in I, and a connective and strictly increasing operation $\circ_m$ that is also derived from the corresponding operation in I. (Thus, they have *operational* meaning in the real physical world.)

7. Introduce a *measure*, $\varphi : A \to \mathbb{R}$ as follows: (I) Select an element u from X as a unit, that is, $\varphi(u) = 1$. (II) Construct φ satisfying following two conditions: (II-i) $\varphi(nx) = n\varphi(x)$, (II-ii) If and only if $x \succsim_m^I y$, $\varphi(x) \geq \varphi(y)$ holds. The precision of φ can be arbitrarily increased by the countability

and the Archimedes' condition on the elements of I. Suppose that x is not equivalent to the unit u. If $x \succ u$, then the Archimedes' condition assures there exists a positive integer k such that $(k+1)u \succsim x \succsim ku$. By this and (II-i),(II-ii) above, $\varphi((k+1)u) = k + 1 > \varphi(x) \geq k = \varphi(ku)$ holds. Thus, $\varphi(x)$ is obtained within the *error* 1. Here an error is defined as the upper bound of the difference between the true value and the measured value. Then, taking 2-copy of the each side of the equation above, $(2k+2)u \succsim 2x \succsim 2ku$. With the Archimedes' condition, this means $(2k+2)u \succsim 2x \succsim (2k+1)u$ or $(2k+1)u \succ 2x \succsim 2ku$ holds. Applying the function φ and simplifying the equations with the conditions (II-i),(II-ii) yields $k + 1 \geq \varphi(x) \geq k + \frac{1}{2}$ or $k + \frac{1}{2} > \varphi(x) \geq k$. Thus $\varphi(x)$ is obtained within the error $\frac{1}{2}$. Thus, by considering the virtual n-copy on the mathematical model A, the function φ can be obtained with arbitrary precision.

Thus, we can obtain a measure of attribute associated with the measuring instrument, and the measure can have arbitrary precision, even if it is for a measurement in the potential future.

3.3 Extensional Information Articulation as Generalized Measurement

In the view of this study, the extensive measurement in metrology is considered to be a sort of the information articulation of the physical universe.

In this section, we reconfirm the points of the extensive measurement (3.3.1), clarify the implicit assumptions in the extensive measurement from the viewpoint of information articulation (3.3.2), consider the requirements imposed on the IGUSes (3.3.3), and finally, develop a model of the process of extensional information articulation (3.3.4).

3.3.1 Review of the extensive measurement

The points of the extensive measurement described in Section 3.2 are summarized as follows:

a. Introduction of a physical apparatus, *i.e.*, a measuring instrument (e.g., equal arm beam for the weight), representing the physical intervention(s): The measuring instrument is introduced by the third person (the measurer) and observed by her/him. This naturally means that the measurer can identify and manipulate the instrument.

b. Identification of target objects by a third person (the measurer): In a measurement, the target objects are identified by the measurer beforehand.

c. Identification of required physical operations and their structures: Relations such as *equivalence relation* and operations such as *concatenation* are introduced physically, in other words, they are introduced with an operational semantics.

d. Introduction of *n-copy*: Beyond actual physical tokens, multiple copies/alternatives for (potential) repetitive operations are modeled to be obtained physically.

e. Establishment of maps between physical operations and mathematical operations and relations.

f. Establishment of the measure, that is, the map from the mathematical structure representing the physical situation to the real number.

g. *The Archimedes' condition*: Universality of the measurement even for the potential measurements in the future is theoretically secured by the Archimedes' condition (and the concatenation operation). The Archimedes' condition theoretically assures the arbitrary precision of the measurement. Moreover, with the concatenate operation, it theoretically secures the universality of the measurement, in the sense that it enables to measure beyond a particular physical situations specified (by the actual physical tokens) as a measurement.

These points need to be re-interpreted and reconsidered for the extensional information articulation. This will be done in the next section.

3.3.2   Implicit Assumptions in Extensive Measurement

As Baird [106] pointed out, a measuring instrument plays two roles in information phenomena. (1) To begin with, a measuring instrument has to do something physically. It has to incorporate physical interactions related to the quantity to

be measured. An equal arm beam balance, for example, is under the influence of gravity. In addition, the measuring instrument moves in a particular way reliably in the physical interactions. That is, it moves in an approximately same manner in the approximately same situations. The action of the instrument has been separated from human agency and built into the reliable behavior of an artifact. This is simply a physical matter. (2) A measuring instrument also has to provide representations. The results of measuring something must be located in an ordered space of possible measurement outcomes. A representation of this ordered space has to be built into a measurement instrument. This can be as simple as the scale to show the balance on an equal arm beam balance.

The latter role presupposes human agency, e.g., the existence of a measurer. The representations are for the measurer. If one considers the extensive measurement process as simply an information phenomenon, there are several implicit assumptions originated in this presupposition.

Reflecting on the points listed in Section 3.3.1, major implicit assumptions related to the subject of the present study are:

a. Identities of objects in I are secured by the measurer. That is, the measurer is assumed to be able to identify the objects to be measured. In addition, the identities of objects are almost stable. That is, in most cases, the identity of each object under consideration is kept during the measurement (at least).

b. The comparator m (a part of the universe) is identified by the measurer. That is, an attribute corresponding to a comparator can be identified by the measurer. Moreover, the comparator can be manipulated by the measurer. Manipulation here means, (1) putting the object(s) into the situation to be measured by the comparator, (2) capturing the information on the object(s) the comparator displays. That is, the physical interactions between the comparator and the measurer are distinguished (and utilized) by the measurer.

c. The measurer has abilities to construct the suitable representations (sets for collection of objects, equivalence relations, real numbers for measured values, and so on).

To generalize the extensive measurement process to the information articulation process, these assumptions are reconsidered and turned into the requirements for the IGUS in the next subsection.

3.3.3   Requirements for the IGUS

The implicit assumptions of the extensive measurement in Section 3.3.2 do not hold in general information articulation processes.

In an information articulation process, an IGUS plays the role of the measuring instrument and the measurer simultaneously. There is no outsider who can observe and manipulate. There are merely physical interactions. Hence we need to reconsider the assumptions to generalize the extensive measurement.

First of all, before articulating the physical universe by information, there is no identity for everything in principle. We should not assume any privileged physical interactions. It is impossible, at least initially, for an IGUS to identify a fragment of the universe (FoU for short, from now on), because the IGUS has no clue, *i.e.*, information to distinguish a part of the universe from the others[3]. Note that identification here means that to conceive several physical interactions as united even if there is a spacio-temporal separation of these interactions. An IGUS is assumed to extract information to identify an FoU all by itself.

In addition to this, what an IGUS can rely on as a basis of differentiation is the IGUS itself only. There is no specific physical apparatus such as measuring instruments to be used by an IGUS to articulate the universe. If there is such an apparatus, the apparatus should be regarded as a part of the IGUS. The physical construction of the IGUS itself may be used to articulate the universe. We should assume that an IGUS compares FoUs by its own capabilities, and these capabilities might not be distinguished by the IGUS itself (even for this distinction, information might be required). The IGUS might correlate one of the capabilities with one of its own actions, and the capability may be identified indirectly through the action.

Finally, an IGUS is assumed to play the role of the measurer. An IGUS is assumed to have abilities to construct the suitable representations of physical entities that are operated during the information phenomenon.

These revised presuppositions impose physical requirements on IGUSes. Though the outcome might not novel—after all, the IGUS in this chapter has similar construction with information agents already proposed—the reasoning may attract some interest.

---

[3]In this sense, one may call an FoU as a *proto–object.*

a. An IGUS is assumed to be able to directly indicate an FoU by a certain physical interaction i. Only during interaction, the IGUS can indicate a specific FoU just as it is. When the interaction is lost, the IGUS cannot identify the same FoU (if there is no further information). Using a term in linguistics, *diachronically* indistinguishable interactions are assumed to work for indicating FoUs. Such kind of interactions will be called as the *indicative interactions* in this article. The most distinguishable character of the indicative interactions may be their durability. Durability is essentially relative. This durability is related to the similarity detections by the IGUS (see the next item). The "source" of the indicative interaction should be the IGUS. In this sense, metaphorically, the indicative interaction can be regarded as an "active" interaction of an IGUS, and the physical implementation of such an interaction is often called as an *actuator* of the IGUS. An IGUS may have representations of each FoU paired with the indicative interaction i. Note that this representation is rather loose. If the IGUS activate the same interaction i, the exactness of the indication is governed by chance. That is, the IGUS cannot "specify" the FoU by using the indicative interaction alone.

b. Another particular physical interaction p is assumed. An FoU a picked out by an indicative interaction i is compared with another FoU b picked out in the same way (during i is kept). The interaction p tells the IGUS that, from the viewpoint of the IGUS, the compared FoUs are whether indistinguishable or not. Thus this interaction p detects a similarity between a and b. Using a term in linguistics, *synchronically* indistinguishable interactions are assumed to work for detecting the similarities among FoUs. Such kind of interactions will be called as the *comparing interactions* in this article. Most of the case, this interaction can be regarded as, metaphorically speaking, a "passive" interaction of an IGUS, and the physical implementation of such an interaction is often called as a *sensor* of the IGUS.

c. And *memory* is required to store the effects of interactions (physical difference made by another part of the universe), keep them, and utilize them later if necessary.

Thus, an IGUS in this chapter is a kind of so-called *sensorimotor system* (e.g., [107]). But note that we should not use a personification to avoid confusions. An

IGUS in general is, ultimately, a mere bunch of physical interactions[4].

3.3.4   A Model of Extensional Information Articulation Process

The formulation of the extensive measurement tells us the importance of the distinction between the physical world and the mathematical model, and the importance of the linkage of them.

In general information articulation processes, information is related to a universal structure constructed from FoUs. FoUs are tentative entities, but if the IGUS can extract something *universal*, that is, existent or operative almost everywhere and under almost all conditions, this universality may be utilized (as information). In the case of the extensive measurement, the universality is simply condensed in a homomorphism $\varphi : A \to \mathbb{R}$, assuming the Archimedes' condition. But in general information articulations, the differences induced within an IGUS may form various structures, and the IGUS may utilize these various structures. Any models of information articulation process should incorporate such an extraction of universality from rather vague FoUs, at some generic level.

Following the formulation of the extensive measurement, a model of the process of extensional information articulation by an IGUS satisfying the requirements mentioned in Section 3.3.3 is developed. The model is presented, with one-to-one correspondence to the steps in the Section 3.2, as follows:

1. As already mentioned, there is no a priori thing or object. An IGUS may indicate an FoU by an indicative interaction. An FoU is, as it were, proto-information. A collection of FoUs, I, may be made physically or within the IGUS (as representation of a collection of physical FoUs).

2. Comparator (*i.e.*, Sensor) for indifferences $m \in M$ is introduced. It is a pair of the indicative interaction and the comparing interaction. The comparator can compare an FoU with another one. That is, the comparator m shows that the compared FoUs are whether indistinguishable or distinguishable from the comparator's viewpoint. This distinction induces an equivalence

---

[4]In this sense, *memory* also should have been more elaborated in the language of physical interaction. But the authors omit unnecessary details for the present study.

relation $\sim_m^I$. In addition, the FoUs are partitioned into equivalence classes, tentatively. (Recall that an FoU is not identified by the IGUS at this stage. It is only indicated during the indicative interaction, and at best, has a representation loosely related to the actual physical FoU.) There may be a collection of comparators M for an IGUS and the relation $\sim_m^I$ is relative to each comparator $m \in M$.

3. Assume that one can define a physical relation R_M^I between the (tentative) equivalence classes of FoUs. Suffix for the comparator is replaced/generalized by the set of comparators because a combination of the comparators is commonly used for articulating one piece of information.

4. Characterize the physical relation R_M^I by the physical operations/interactions on the equivalence classes of FoUs, if possible.

5. Try to deduce a universal structure constructed physically from the equivalence class of FoUs and R_M^I. If this step is not successful, the universal structure is inquired in the formal side, *i.e.*, in the mathematical model (step 7).

6. Corresponding the above physical situations, relate the equivalence classes of the FoU to a set X in the mathematical model. The set X must have the structure reflecting the physical relation R_M^I. (Thus, they have *operational* meaning in the real physical world.)

7. Introduce the homomorphisms between X and another (known) mathematical structure K, which represents the articulated information. Deduce a universal structure from the K. The universal structure may be a model of the information that is captured (and to be utilized) by the IGUS.

This is a model for describing the process itself, and, simultaneously, this is a model for analyzing information phenomena as information articulation processes. In the latter case, note that the subject of these statements is also an IGUS (typically, human-IGUS) that is investigating the information phenomena.

3.4   Attributes as Limit

In this section, as a significant instance of the extensional information articulation, the attributes extraction from the universe is considered.

We focus on the step 7 in the extensional information articulation process, because this step for extracting the universality is especially worthy to investigate in information science. Other steps in the information articulation process might be investigated in different disciplines. For example, the steps for detecting similarity of FoUs, the step 1 and 2, can be abstracted using *the Curie's principle* in physics. As Joe Rosen [108] explicates, symmetry looms in the foundation of science. Major ingredients of that foundation, reproducibility, predictability and reduction are all symmetry. Hence, physical interactions can be represented at a generic level by the words of symmetry. Thus, we can abstract away the source of equivalence classes of FoUs by assuming certain symmetries.

To focus on the step 7, in this section, following conditions for the IGUS are supposed:

1. The IGUS has collections of FoUs each of which forms an equivalence class. Each of the equivalence relations is induced by certain symmetry of the physical interaction (*comparing interaction*) of the IGUS;

2. The inclusion relation between the collections of FoUs is considered as the relation R_M^I in the step 3;

3. The IGUS has enough physical capabilities to meet the requirements for implementing the model described in the following.

By supposing the first condition, the steps 1 and 2 in the extensional information articulation process are skipped; the step 3 is took by the second condition; the steps 4 and 5 are abandoned; the set X in the step 6 is equipped with an order relation induced by the inclusion relation supposed as the second condition; and the last condition ensures taking the step 7 (Figure.3.1).
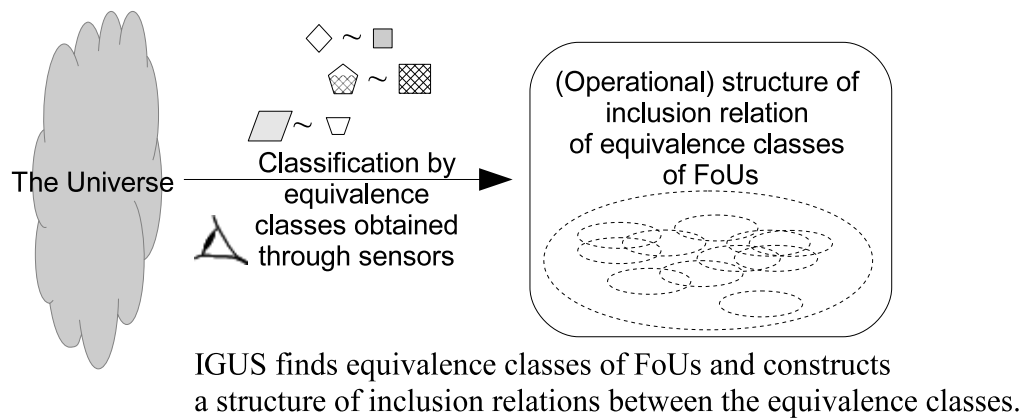
FIGURE 3.1: IGUS constructs a structure from FoUs

Let us return to the situation described in the Section 1.1.1: Suppose that there is a green apple. The question is: how can a *tabula rasa* IGUS articulate this part of the universe as the green apple? Let us assume that the IGUS can capture some similarities of parts of the universe through certain symmetries of the interactions. But this does not mean that a particular attribute, e.g., *green*, can be carved out of the universe by the IGUS. The IGUS must distinguish each attribute by the differences induced within itself.

We, human-IGUSes, often use the Aristotelian concept to specify a certain object out of this uniform universe. In Aristotelian classical category, features of the category members are binary, primitive, universal, abstract and innate, all members of the category have equal status and the categories have clear boundaries (see for example, [63, 67]). A concept is defined via a set of necessary and sufficient properties. That is, a concept has two sides: its *extension*, *i.e.*, a collection of objects, and its *intension*, *i.e.*, a collection of attributes (which represents the equal status of objects), and these two sides reciprocally define the concept (for a formal model of such "concept", see [109]).

But as already mentioned, for a *tabula rasa* IGUS, there is no attributes and objects beforehand. An IGUS must extract them all by itself. A model is required to describe and analyze such a bootstrapping attributes extraction process.

The model presented in this section is developed in three steps.

First, *the projective limit* is used as a mathematical model to attain universality from finite physical interactions between the universe and the IGUS. In some field

in physics such as physical cosmology, researchers have similar problems to explain the situation in which universal information is extracted from finite observations, and our model is along with the line of Sorkin [111].

Second, structural reduction of information is introduced. All of the universal elements obtained in the projective limit are not necessarily used as attributes to specify a part of the universe. Elements for minimal construction of attributes are modeled as *irreducible elements of the lattice* in the limit.

Third, in the same way as in Formal Concept Analysis [109], *Galois correspondence* between the inclusion structure of equivalence classes of FoUs and the inclusion structure of attributes obtained in the limit establishes the objects and attributes for the IGUS.

Though the following description includes minimum mathematical information for interested readers, the use of mathematics in this section is to exemplify more concretely the possibility of the extensional information articulation, and is not intended to present a mathematically rigorous formulation.

3.4.1 Universality

An IGUS, which was initially a *tabla rasa*, incrementally obtains collections of FoUs each of which forming an equivalence class. The collections of equivalence classes of FoUs at the stage t is denoted as S_t. Between the equivalence classes A, B in S_t, there exists a natural inclusion relation $\subseteq$: $A \subseteq B$ if and only if for any $a \in A$, $a \in B$. Here, a is a representation of FoU in the IGUS. It is natural to simply think that a certain uniformity of FoUs, that is, an *attribute* is represented by each equivalence class (as Chechkin). However, it has some drawbacks as mentioned in the Section 3.1.

To obtain a required universality of attributes and reflect the proper relationship between them, we should move to down-sets of S_t, which is denoted as D_t, and the sequence of them. D_t is the down-set of S_t if, whenever $X \in D_t$ and $Y \in S_t$ and $Y \subseteq X$, we have $Y \in D_t$. Recall that X, Y here are equivalence classes of FoUs.

Because there is a mathematical fact that the family of down-sets ordered by set inclusion forms a lattice isomorphic to the original lattice [69], D_t can be considered to be virtually the same as S_t, in the context of this chapter.

Thus an IGUS is assumed to have a sequence $D_0 = \emptyset, D_1, D_2, \ldots$, which incrementally increases according as the IGUS has more experiences, that is, more interactions with the universe. The universality is extracted from this virtually infinite sequence.

From T-indexed family of sets, $\{D_t\}_{t \in T}$ and appropriately defined functions $f_{tu} : D_u \to D_t$, one can construct a mathematical structure called a *projective system* $\left(D_t, (f_{tu})_{t \leqslant u} \right)$. f_{tu} can be defined as:

$$f_{tu}(x) = \bigvee \left\{ \eta : g_{tu}(\eta) \leqslant x, \eta \in D_t \right\}, \tag{3.1}$$

where $g_{tu} : D_t \hookrightarrow D_u$ is the inclusion map preserving the top and the bottom in D_t, and the symbol $\bigvee$ denotes the least upper bound.

For a projective system, a *universal* concept, that is, *the projective limit* is defined:

$$\varprojlim_{t \in T} D_t = \left\{ (x_t)_{t \in T} : x_t \in D_t, t \leqslant u \Longrightarrow x_t = f_{tu}(x_u) \right\}. \tag{3.2}$$

What characterizes the projective limit is its universality. For a projective system $(D_t, f_{tu})_{t,u \in T}$, a set D and a family of maps $(p_t : D \to D_t)_{t \in T}$ are given, and for any $t, u \in T$ satisfying $t \leqslant u$, if $p_t = f_{tu} \circ p_u$, then there exists a *unique* map $\rho_D : D \to \varprojlim_{t \in T} D_t$ that satisfies $p_t = \pi_t \circ \rho_D$ for any $t \in T$, where π_t is the projection map for t–th component of $\varprojlim_{t \in T} D_t$. This is a realization of *universality* in a mathematical model (Figure.3.2).

The projective limit gradually changes as the IGUS is affected by new interactions. But as far as the interactions do not induce a significant structural change in D_t, the limit remains the same. In this sense, the attributes modeled as this structural limit can be said to be *universal*, and this corresponds to the mathematical universality mentioned above.

Finite sequence up to D_t, *i.e.*, $(D_s)_{s \leqslant t \in T}$ can be considered to be a finite approximation of the projective limit $\varprojlim_{t \in T} D_t$. By construction from the family of

Universal structure is obtained by
the projective limit of the sequence of
the inclusion structures of
equivalence classes of FoUs..

Limit structure of
inclusion relation
of equivalence classes
of FoUs

(Operational) structure of
inclusion relation
of equivalence classes
of FoUs

(Operational) structure of
inclusion relation
of equivalence classes
of FoUs

limit
= universal

FIGURE 3.2: Universal structure for attributes

down-sets, the finite sequence up to D_t can be easily retrieved from D_t, hence D_t can be identified with $(D_s)_{s \leqslant t \in T}$ as a finite approximation of $\varprojlim_{t \in T} D_t$.

Metaphorically speaking, suppose that each D_t is an approximate value of the ratio of the circumference of a circle to its diameter, π, that is, $D_0 = 3, D_1 = 3.1, D_2 = 3.14, D_3 = 3.141, \ldots$. Then the projective limit is the value π. The universal value π is represented extensionally as

$$\pi = \{3, 3.1, 3.14, 3.141, \ldots\} \tag{3.3}$$

However, note that the projective limit above in general is for the structures, not simply for values. In the present study, the projective limit is the limit of inclusion structures of (down-sets of) the equivalence classes of FoUs.

3.4.2   Specifiability

In the sense described in Section 3.4.1, each elements $x, y \in \varprojlim_{t \in T} D_t$, is universal.

However, not all of them are required to specify a part of the universe as an attribute[5]. There seems, so to speak, an *economic principle.*

The projective limit obtained in Section 3.4.1 can be assumed to have the structure mathematically called *lattice.* There is a fact that in a lattice, any element of the lattice can be obtained as a certain combination of its specific part of elements [70]. These special elements, which form the basis of the lattice, are mathematically called *the meet irreducible elements* of the lattice.

Therefore, these meet irreducible elements of the projective limit (of the family of down-sets of the collections of equivalence classes of FoUs) may be regarded as *attributes* derived from the structural differences induced within the IGUS by the physical interactions. We will call them as *s-attributes* ("s" for *structural*) if it is required to emphasize on the distinction from other ordinary uses of the word "attributes" (Figure.3.3).
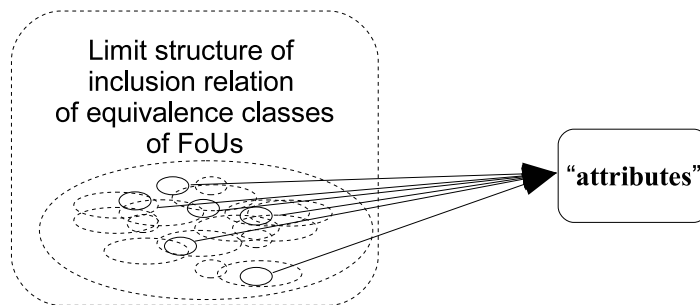
FIGURE 3.3: Several specific elements — irreducible elements — in the limit structure are used as "attributes"

When one of the meet irreducible elements of the projective limit is labeled as attribute a, roughly speaking, all of the FoUs belonging to the element of the projective limit have this attribute a. More precisely, we might be able to say that an FoU *hereditarily succeeds* to the labeling as attribute a, because an element of the projective limit is a set of the down-sets D_t each of which consists of a set of equivalence classes of FoUs.

In this way, an IGUS can equip enough s-attributes to (virtually) specify the equivalence classes of FoUs, *i.e.*, particular parts of universe. Besides, the IGUS is now able to handle attributes extensionally, because the s-attributes are virtually represented extensionally. "Virtually" here is placed for worrying that this model of the attributes seems rather conceptual because the s-attributes are defined "in

---

[5]They may be called as *proto–attributes* corresponding to *proto–objects.*

the limit." However, even if it is merely conceptual, by understanding in this way, we can define a clear concept of the *approximation* of an attribute in D_t, that is, the projection of an s-attribute in $\varprojlim_{t \in T} D_t$. These approximations can clearly be handled extensionally. Moreover, there might be a physical implementation of the limit, which provides extensionality of the attributes directly.

3.4.3   Duality and Galois Connection

As described in the above subsections, an IGUS can obtain a set of s-attributes that are somewhat universal, and the IGUS can handle the s-attributes extensionally. But recall that actual physical FoUs are merely loosely related to the representation within the IGUS. At this stage, there is no specific object with which these s-attributes are associated. As for information phenomena, the known objects are a part of the structure built in an IGUS through the previous information from the external world. *Models* mentioned above might be a subpart of the structure which enable to classify (and, first of all, identify) the *objects* in the external world. In this chapter, the author shows that the universal elements of the structure formed by the known objects, that is, *attributes* may be considered as *models* above, and also shows the way of classifying the objects by the attributes.

In general, to investigate information phenomenologically on the physical basis, we have to face with the opposition between the two systems – an IGUS and the universe. The problem is the way in which an IGUS may induce the physical/structural differences within itself in harmony with the differences found in the universe by the IGUS. An IGUS might induce its internal differences arbitrarily, but what functions as information is the differences induced with this harmonization.

The whole picture of this information capturing from the universe by an IGUS may be drawn as a *duality* between the external world and the system. To be more precise, it may be drawn as a duality between the structure found in the external world and the structure built within the system. In this view, one should use the theoretical framework based on the mathematical concept called *duality* to investigate the interrelations between the invisible, incomprehensible, indivisible universe and the information world which is (turned into) visible, comprehensible for each IGUS.

Duality is a general mathematical idea representing the relationship having following features. A general theoretical structure found in analyzing various phenomena is as follows:

1. prepare known objects that are able to reproduce the essential properties of the structure or the behavior of theoretical (unknown) objects that should be described, analyzed, classified, and interpreted;

2. as for the focusing aspect, classify the known objects as *models* or *samples* that faithfully reconstruct the unknown object under consideration;

3. then interpret the meaning of the original objects from the interrelations of the *models*.

Without such a process, it may be difficult to approach the essentials of the unknown objects in the fickle, evasive real world. Naturally, a certain guarantee should exist to use known objects as proxies of the unknown real objects. There should be a balancing or an adjusting mechanism between the process from the unknown objects to the models and the reverse process from the models to the unknown objects. The *duality* is none other then the requirement for not forming a gap more than assumed when one proceeds back and forth between the models and the unknown objects. [71][6].

Precise duality, that is, one to one correspondence between the unknown objects and the models does not exist by nature. The balancing/adjusting mechanism mentioned above may be well-described by a concept in category theory – *adjunction*, which will be explained in the next chapter. The canonicity and the universality of information might be understood well by *adjunction*.

The connection between the s-attributes obtained and the specific part of the universe may be established by the Galois connection on the model. The Galois connection is an adjunction. Physically, it may be established by the indicative interactions that are correlated with the representation of FoUs within the IGUS. Through this correlation, the indicative interactions are also correlated with the s-attributes. Therefore, the IGUS can activate the indicative interaction correlated with one of the s-attributes and indicate some part of the universe. The part

---

[6]Excerpt and translated with a slight modification by the author.

of the universe may or may not strictly agree with the FoUs initially indicated. But, anyway, this indication is stable for the IGUS in comparison with the initial indication to the FoUs, because it has its ground within the IGUS itself. And if the indication correlated with the s-attributes specify a part of the universe, the part of the universe is/*becomes* the object corresponding to the s-attributes.

This situation is mathematically modeled as the Galois connection between the lattice of equivalence classes of FoUs and the lattice of the s-attributes. See [109] for general formulation in detail (Figure.3.4).

Thus, in this model, a universal structure, *i.e.*, the projective limit, is built from the structures of the inclusion relation of the FoUs, and using extensionality of the universal structure, via the Galois connection, another universal structure—(the inclusion relation of) the pairs of objects and attributes. The IGUS has now got the extensionally specifiable attributes and objects, In this way, the information phenomena may be understood as the transformation of extensional structures, that is, the transformation of extensional differences.

1. Attributes concerned with
the same equivalence class
of FoUs form a set.
These sets form
the inclusion structure.

Each pair of the FoU structure and the attribute structure form
Galois connection.
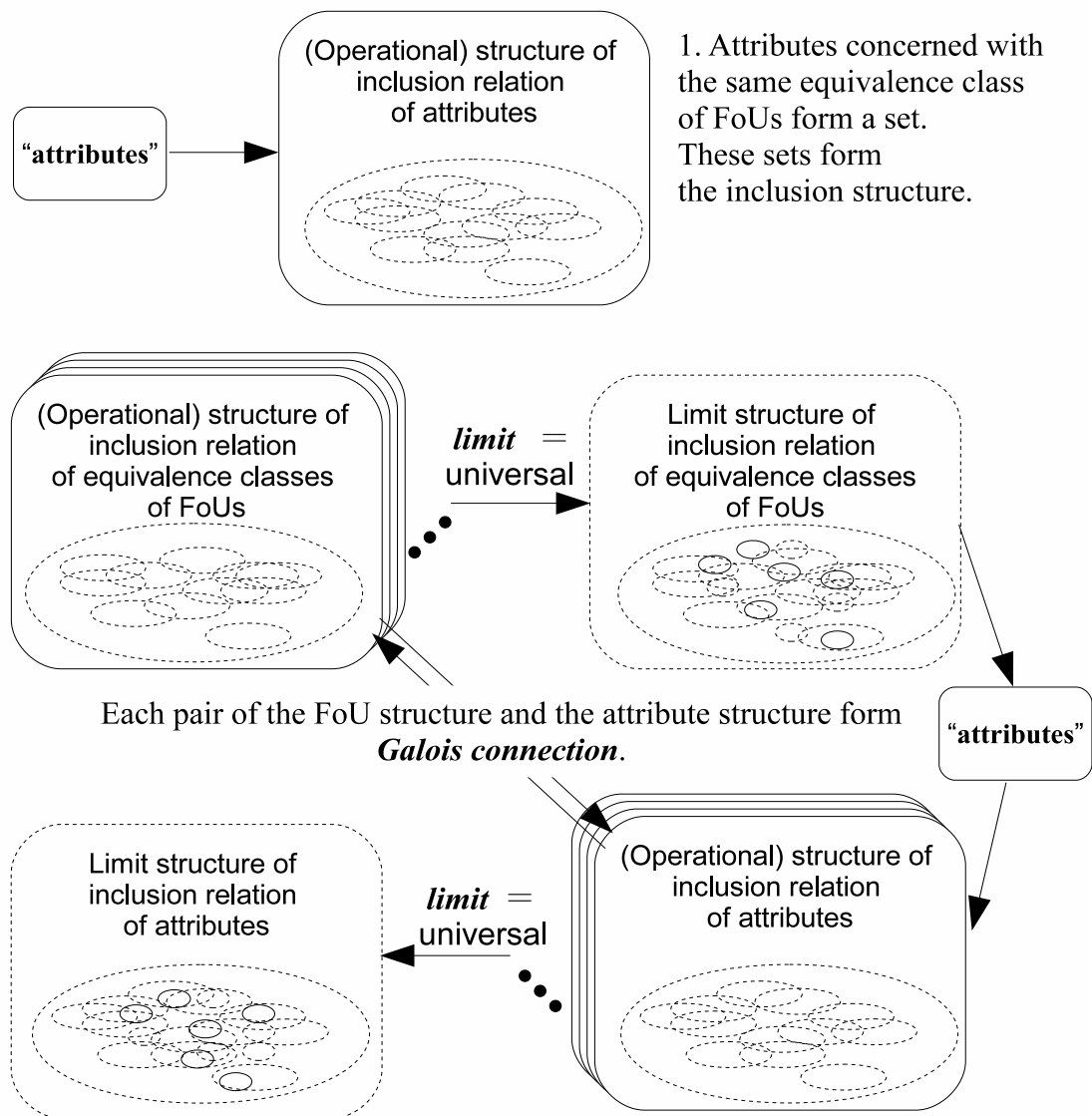
2. Another universal structure is obtained also by the projective limit of
the sequence of the inclusion structures of attributes.
Each pair of the FoU structure and the attribute structure form *Galois connection*,
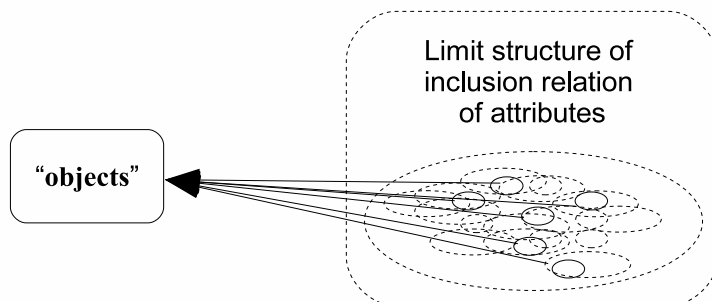
3. Several specific elements — irreducible elements —
in the limit structure are used as "objects".

FIGURE 3.4: Steps to obtain "objects"

3.5   Conclusive remarks on this chapter

It has not been well investigated to understand the fundamentals of information as fully extensional physical phenomena. Studies of information in physics, and a few studies in information theory, e.g., Chechkin's (see [15]) indeed treated information extensionally. However, these treatments have seemed to be not thorough enough. They did not provide any general enough models that could be used in a fundamental level of information phenomena, such as the articulation of information from the universe.

In this chapter, we have developed a model of the process of extensional information articulation, generalizing the theory of the extensive measurement in metrology. In addition to this, a model of the attribute creation process has been presented as a detailed example of a step of the extensional informational articulation process.

The former model provides an abstract model to represent a concrete information articulation process extensionally. By establishing the relation between the components in the model and those in an actual, concrete situation, an information phenomenon may be sorted out and analyzed as an extensional information articulation. Thus, this model may play the similar role as that of the extensive measurement theory in metrology.

The latter model has exemplified the concepts of the former model. It has demonstrated a concrete formulation along the line of the former model, *i.e.*, the model of the process of extensional information articulation, and presented the possibility of a new understanding of the attributes in which attributes are extracted fully extensionally from the universe. The outline of the model is summarized as in Figure.3.5.
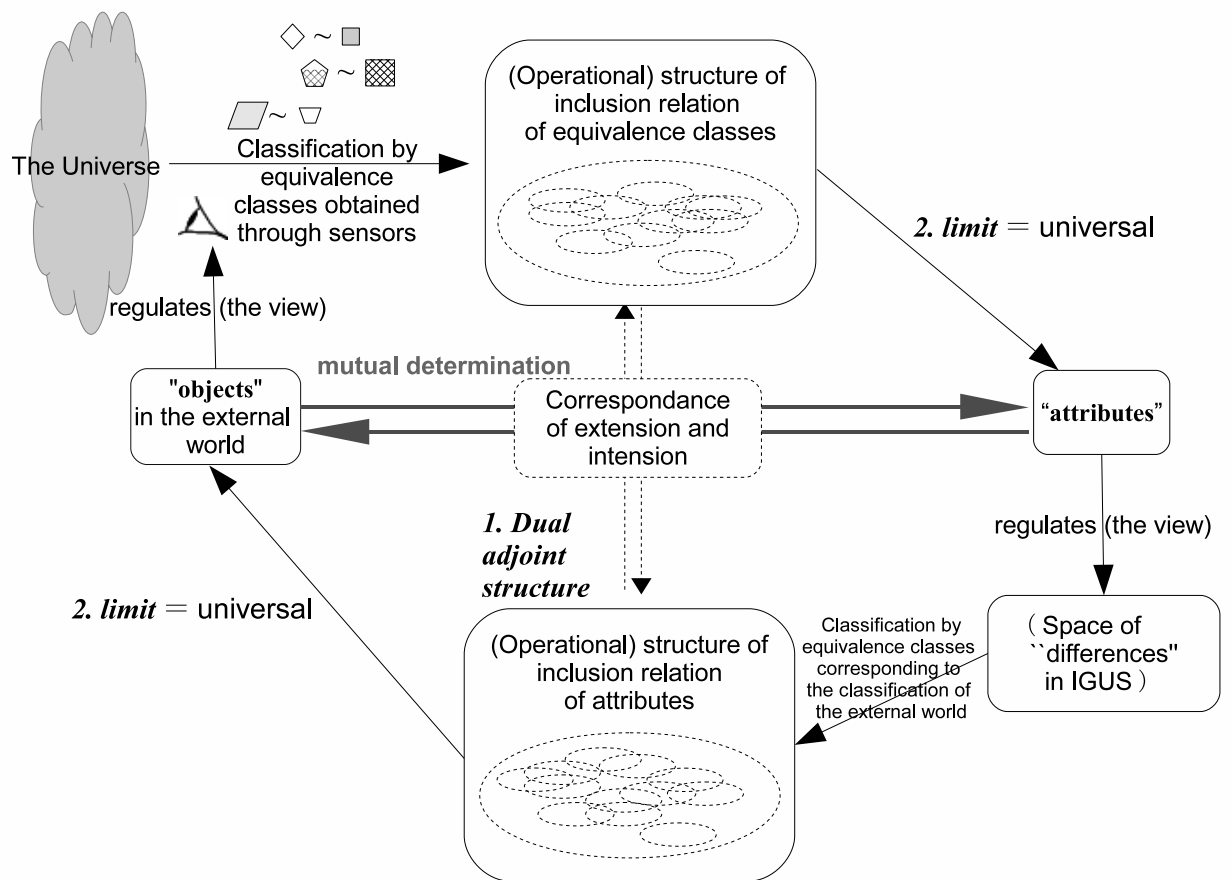
FIGURE 3.5: Model of the process of extensional information articulation

Chapter 4

An internal description of an origin of the attribute

4.1 Introduction

If one places importance on the existence of physical basis underlying information phenomena to understand information fundamentally as in [1] , it is inevitably a physical difference within the system that captures each information from/in the universe. The difference within the system is induced by the physical difference in the universe, and the induction is restricted by the physical ability of the system.

Hence a theory of information which aims at such a fundamental understanding should describe, analyze, and interpret the differences as information and the mechanism of the transfer of the differences. A difference is something in between. Since this *between-ness* or *intermediary-ness* is considered to be one of the main characteristics of information, it is desirable to use a suitable language to describe such a character directly.

One of such a mathematical gadget might be category theory [72–74]. A category consists of the objects and the arrows between the objects. In category theory, the leading role is played by the arrow between the objects. What is in the spotlight is the arrows, *i.e.*, *between-ness*.

In this chapter, the author attempts to show that the usefulness of the language of category theory, especially the *topos*, for investigating the information phenomena. By connecting the requirements for the language to investigate information

phenomena and the features of elements in category theory, and by exhibiting the use of category theory to study a typical information extraction phenomenon, that is, the attribute extraction for classification, the effectiveness of category theory in the study of information will be shown.

Information acquisition by an IGUS may be better formulated in the language of category theory [72–74], topos [75–77] in particular. Though detailed discussions will be evolved in the following sections, the reasons are roughly:

- The components of the category theory are suitable for representing the *relational* aspect of information, *i.e.*, difference which makes difference;

- Elements in a category is treated *contextually*;

- "Information" acquired by/at the informee system is understood as a *structural universality* built within the system from fickle stimuli of outer world. The structural universality is immanent in the category theory;

- There exists a *subobject classifier* (in every topos);

- *Topos* is a generalization of the category sets;

- One may successfully explain a mechanism of the extraction of information called *attribute* from the outer universe in the language of category theory.

The structure of this chapter is as follows: In Section 4.2, the properties of the category theory and topos which seems to be useful for information studies are briefly surveyed. The author clarifies the connection of the viewpoint for information in this study and the features of the category theory in this section. Then, a mechanism of the extraction of attributes by an IGUS is explained in the language of topos in Section 4.3 as a demonstration of the usefulness of the category theory. Section 4.4 concludes this chapter.

4.2 Category theory as a language for information phenomena

Category theory seems to be, though not perfect, suitable for investigating information phenomena. In this section, concepts in category theory are introduced

minimally. Readers should consult [72–74, 78], and other textbooks for detailed definitions.

Definition 4.1. A category $\mathcal{C}$ consists of a family of *objects* $Ob(\mathcal{C})$ and for each pair of objects, A, B in $Ob(\mathcal{C})$, a set $\mathcal{C}(A, B)$ of *arrows*[1] from A to B, together with a way of composing arrows that match

$$\circ : \mathcal{C}(A, B) \times \mathcal{C}(B, C) \to \mathcal{C}(A, C).$$

This data is to satisfy:

1. composition is associative. If $h \circ (g \circ f)$ is defined, $(h \circ g) \circ f$ as well and $h \circ (g \circ f) = (h \circ g) \circ f$.

2. there are identities for each object, so there are *identity* arrows id_A in $\mathcal{C}(A, A)$ for A in $Ob(\mathcal{C})$, such that if $f \in \mathcal{C}(A, B)$,

$$\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A.$$

    (An arrow is also written as $f : A \to B$ if $f \in \mathcal{C}(A, B)$, and the composite of the arrows $f \in \mathcal{C}(A, B)$ and $g \in \mathcal{C}(B, C)$ is written as $g \circ f$ or simply gf.)

The object A of an arrow in $\mathcal{C}(A, B)$ is called the *domain* of the arrow, and the object B is called the *codomain* of the arrow.

In this thesis, we will use the term *arrow* mainly, but sometimes use the term *morphism* interchangeably with *arrow*. And by abuse of notation, $Ob(\mathcal{C})$ is often written as $\mathcal{C}$ when there seems no risk of confusion.

An arrow $f \in \mathcal{C}(A, B)$ is *isomorphism* or *invertible* in $\mathcal{C}$ if there is an arrow $g \in \mathcal{C}(B, A)$, such that $g \circ f = id_A$ and $f \circ g = id_B$. Objects $A, B \in Ob(\mathcal{C})$ are *isomorphic* in $\mathcal{C}$ if there is an arrow $f \in \mathcal{C}(A, B)$ that is an isomorphism in $\mathcal{C}$.

In category theory, isomorphic objects are regarded as "essentially same."[2]

> ...(snip)..., isomorphic objects "look the same". ... (snip) ... these arrows, which establish a "one-one correspondence" or 'matching" between the elements of the two objects, preserve any relevant structure.

---

[1]also called *morphisms*.
[2]This is not a mathematically defined term here.

This means that we can replace some or all of the members of one object by their counterparts in the other object without making any difference to the structure of the object, to its appearance. ...(snip)... An object will be said to be "unique up to isomorphism" in possession of a particular attribute if the only other object possessing that attribute are isomorphic to it. A concept will be "defined up to isomorphism" if its description specifies a particular entity, not uniquely, but only uniquely up to isomorphism. [75].

An "arrow between categories" is called a *functor*.

Definition 4.2. A *functor* $F : \mathcal{C} \to \mathcal{D}$ is given by:

- An object–map, assigning an object FA of $\mathcal{D}$ to every object A of $\mathcal{C}$.

- An arrow–map, assigning an arrow $Ff : FA \to FB$ of $\mathcal{D}$ to every arrow $f : A \to B$ of $\mathcal{C}$, in such a way that composition and identities are preserved:

$$F(g \circ f) = Fg \circ Ff, \ F\mathrm{id}_A = \mathrm{id}_{FA}.$$

An "arrow between functors" is called a *natural transformation*.

Definition 4.3. Let $F, G : \mathcal{C} \to \mathcal{D}$ be functors. A *natural transformation*

$$t : F \to G$$

is a family of morphisms in $\mathcal{D}$ indexed by objects A of $\mathcal{C}$,

$$\{t_A : FA \to GA\}_{A \in Ob(\mathcal{C})}$$

such that, for all $f : A \to B$, the following diagram commutes:

$$
\begin{array}{ccc}
FA & \xrightarrow{Ff} & FB \\
\downarrow{\scriptstyle t_A} & & \downarrow{\scriptstyle t_B} \\
GA & \xrightarrow[Gf]{} & GB
\end{array}
$$

That is, $t_B \circ Ff = Gf \circ t_A$. This condition is known as *naturality*.

Categories, functors between categories, and natural transformations between functors are the main characters in category theory. Various concepts are built on them.

The merits of using category theory as a language for information studies are summarized as follows:

1. Arrows as comparison

2. Being contextual

3. Capability for representing structural universality

4. Existence of the subobject classifier

5. Topos as a generalization of the category of sets

6. Existence of a successful application

Following subsections explain each of these merits in detail.

4.2.1   Arrows are comparison

The components of category theory are suitable for representing the relational aspect of information. First of all, an arrow is often interpreted as comparison.

> "The basic belief of category theory is that whenever we conceive of a collection of 'objects' - things we don't want to take apart - we should, at the same time, decide how these 'objects' are to be compared. We then formalize a category. In other words, for any given category C we should think of the arrows of C as those gadgets which compare the objects." [72].

> "... one of the reasons for the usefulness and importance of Category Theory is that it gives an abstract mathematical setting for analogy and comparison, allowing an analysis of the process of abstracting and relating new concepts." [79].

As mentioned in the above quotes, an arrow in category theory may naturally be interpreted as a *difference*. In category theory, the arrows are in the foreground and the objects are in the background, therefore the arrows in category theory are considered to be suitable for describing the difference carried by information.

4.2.2 Category is contextual

As many researchers in various fields such as linguistics has been addressed, information has *contextual* aspects. For example, in the episode 1.1.2, the meaning of a word is understood in the context of a language.

All the elements in category theory are defined structurally in a context of the category they reside. For example, let us see the definition of a *monic*, a special arrow which is often considered to represent "a part of the target object." A *monic* $f : A \to B$ in a category $\mathcal{C}$ is defined as follows.

Definition 4.4. An arrow $f : A \to B$ in a category $\mathcal{C}$ is called a *monic* when, for every object $C \in Ob(\mathcal{C})$ and every pair of arrows $g, h : C \rightrightarrows A$, the following property holds: $(f \circ g = f \circ h) \Longrightarrow (g = h)$.

As this definition states, a monic is defined within the structure it constructs with other arrows g, h within the same category. Any other featured elements in category theory are defined in similar contextual ways. Thus one can expect that the structural property of category theory enable us to naturally represent the contextual aspect of information.

4.2.3 Structural universality

In the view of this thesis, information acquired by/at the informee system may be understood as a *universality* built within the system from the fickle stimuli of outer world. Information may be considered to be used to cast anchor in the fluctuate, fickle universe to capture and utilize properties of the universe which can be shared among IGUSes. To be utilized timelessly and to be shared as information, there must be a canonical way of capturing the properties of the universe by an IGUS. The universality arises here. The IGUS may articulate the universe in any degree

of *arbitrariness* by its own criteria. But such an arbitrarily articulated fragment of the universe is not reproducible in general. Hence it may not be usable as information with other IGUSes and even by the IGUS itself. However, if there is a canonical way of articulating the universe by the IGUS (and the way is in common with other IGUSes), the phenomena might be reproducible as information phenomena. The canonicity yields the universality required for information. This universality is inevitably structural because it should be understood as something embedded structurally in the physical base of the IGUS.

The structural universality is immanent in category theory. In category theory, one can discuss the universality in a mathematically rigorous manner. Thus category theory may be regarded to capture this universality of information coherently as structural one.

In category theory, universal arrows are defined as follows.

Definition 4.5. Let $G : \mathcal{D} \to \mathcal{C}$ be a functor, and C an object of $\mathcal{C}$. A *universal arrow from C to G* is a pair (D, η) where D is an object of $\mathcal{D}$ and an arrow $\eta : C \to G(D)$ of $\mathcal{C}$, such that, for every pair (D', f) with D' an object of $\mathcal{D}$ and an arrow $f : C \to G(D')$ of $\mathcal{C}$, *there exists a unique* arrow $\hat{f} : D \to D'$ of $\mathcal{D}$ such that $f = G(\hat{f}) \circ \eta$.

Diagrammatically:

$$
\begin{array}{ccc}
C \xrightarrow{\ \eta\ } G(D) & \quad & D \\
\ \ \searrow_{f} \quad \downarrow_{G(\hat{f})} & & \downarrow_{\hat{f}} \\
\qquad G(D') & & D'
\end{array} \quad .
$$

In other words, every arrow f to G factors uniquely through the universal arrow η.

Universal arrows are ubiquitous in mathematics. Most familiar example might be bases of vector spaces. Let Vct_K denote the category of all vector spaces over a fixed field K, with arrows of linear transformations, while $U : Vct_K \to Set$ is a functor sending each vector space V to the set of its elements (such a functor is often called *the forgetful functor*, because it *forgets* the structure of the vector space). For any set X there is a familiar vector space V_X with X as a set of basis vectors. V_X consists of all formal K–linear combinations of the elements of X. Let us consider an arrow $\iota : X \to U(V_X)$, the function which sends each $x \in X$

into the same x regarded as a vector of V_X. For every other vector space W, it is a fact that each function $f : X \to U(W)$ can be extended to a unique linear transformation $\hat{f} : V_X \to W$ with $U(\hat{f}) \circ \iota = f$. This familiar fact states exactly that ι is a universal arrow from X to U. For other examples of the universal arrow or the concept of the universality in mathematics, readers should consult [74] and other appropriate textbooks of category theory.

Closely related to the concept of universality, there is a structural concept called *adjunction*.

Definition 4.6. Let $\mathcal{C}, \mathcal{D}$ be categories. Given $\mathcal{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{D}$, where F and G are functors, an *adjunction* from $\mathcal{C}$ to $\mathcal{D}$ is given by a pair of natural transformations

$$\eta \; : \; 1_{\mathcal{C}} \Rightarrow GF \qquad \text{(unit)}, \tag{4.1}$$

$$\epsilon \; : \; FG \Rightarrow 1_{\mathcal{D}} \qquad \text{(counit)}, \tag{4.2}$$

satisfying the triangle identities axioms:

$$\begin{array}{ccc} F \xrightarrow{\;F\eta\;} FGF & \quad & G \xrightarrow{\;\eta G\;} GFG \end{array} . \tag{4.3}$$

That is, $\epsilon F \circ F\eta = 1_F$ and $G\epsilon \circ \eta G = 1_G$. We say that F is *left adjoint* to G, and G is *right adjoint* to F.

Let us consider comparing the categories $\mathcal{C}$ and $\mathcal{D}$ by functors F, G:

$$\mathcal{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{D}. \tag{4.4}$$

Recall that arrows are comparison in category theory. In this comparison, there are several levels of "similarity" or "sameness" between the two categories.

In the first level, these functors are going back and forth and returning to the "exact starting place":

$$1_{\mathcal{C}} = GF, \tag{4.5}$$

$$FG = 1_{\mathcal{D}}. \tag{4.6}$$

Categories $\mathcal{C}$ and $\mathcal{D}$ are *isomorphism* in this case. This is the first level of similarity

and categories $\mathcal{C}$ and $\mathcal{D}$ may be considered to be "essentially the same."

In the second level, these functors are going back and forth and returning to the starting point *up to isomorphism*:

$$1_{\mathcal{C}} \cong GF, \tag{4.7}$$

$$FG \cong 1_{\mathcal{D}}. \tag{4.8}$$

Categories $\mathcal{C}$ and $\mathcal{D}$ are *equivalent* in this case. This is the second level of similarity. Categories $\mathcal{C}$ and $\mathcal{D}$ might be considered to be "different", but the difference is on the essentially same objects only.

The adjunction, the third level of "similarity" may be interpreted as "Categories $\mathcal{C}$ and $\mathcal{D}$ might be considered to be 'similar' or 'essentially the same' within the permissible level measured by η and ϵ." Comparing with the isomorphism or equivalent case, the unit η and the counit ϵ may be considered to be showing the gap between the $1_{\mathcal{C}}$ and GF, FG and $1_{\mathcal{D}}$, respectively.

The adjunction structure can be found everywhere and, as is shown below, this "gap permitting similarity" is closely related – actually, equivalent – to the universality concept in category theory. To see this equivalence, another version of the definition of adjunction might be appropriate. If the above mentioned definition is regarded as the definition, the following definition becomes a proposition that holds, and vice versa.

Definition 4.7. Let $\mathcal{C}, \mathcal{D}$ be categories. An *adjunction* from $\mathcal{C}$ to $\mathcal{D}$ is a triple (F, G, θ), where F and G are functors

$$\mathcal{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{D} \tag{4.9}$$

and θ is a family of bijections

$$\theta_{C,D} : \mathcal{C}(C, G(D)) \overset{\equiv}{\to} \mathcal{D}(F(C), D), \tag{4.10}$$

for each $C \in Ob(\mathcal{C})$ and $D \in Ob(\mathcal{D})$, *natural* in C and D.

As explained in [112], universality and adjunctions are equivalent. We only mention here the fact that the following propositions hold.

Proposition 4.8. (Universals define Adjunctions) *Let $G : \mathcal{D} \to \mathcal{C}$. If for every object C of $\mathcal{C}$ there exists a universal arrow $\eta_C : C \to G(F(C))$, then:*

1. *F uniquely extends to a functor $F : \mathcal{C} \to \mathcal{D}$ such that $\eta : 1_\mathcal{C} \to GF$ is a natural transformation.*

2. *F is uniquely determined by G (up to unique natural iso), and vice versa.*

3. *For each pair of objects C of $\mathcal{C}$ and D of $\mathcal{D}$, there is a natural bijection:*

$$\theta_{C,D} : \mathcal{C}(C, G(D)) \equiv \mathcal{D}(F(C), D). \tag{4.11}$$

Proposition 4.9. (Adjunctions define Universals) *Let $G : \mathcal{D} \to \mathcal{C}$ be a functor, $D \in Ob(\mathcal{D})$ and $C \in Ob(\mathcal{C})$. If, for any $D' \in Ob(\mathcal{D})$, there is a bijection*

$$\phi_{D'} : \mathcal{C}(C, G(D')) \equiv \mathcal{D}(D, D') \tag{4.12}$$

natural in D' then there is a universal arrow $\eta : C \to G(D)$.

Therefore the following two situations are equivalent in the sense that each determines the other uniquely.

- We are given a functor $G : \mathcal{D} \to \mathcal{C}$, and for each object C of $\mathcal{C}$ a universal arrow from C to G.

- We are given functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{C}$, and a natural bijection

$$\theta_{C,D} : \mathcal{C}(C, G(D)) \equiv \mathcal{D}(F(C), D). \tag{4.13}$$

This means that universality and adjunctions are equivalent in category theory.

Thus by using the adjunction (if possible), one may expect to naturally represent universality and canonicity of information.

4.2.4 Subobject classifier

Information is intimately related to classification. For instance, in Shannon's communication theory, described in 2.2.2, a message received at the message–receiver is destined to be identified with one of the elements of the predefined set of messages. Identification is a kind of classification.

To classify information, an IGUS requires a criterion for the classification, *i.e.*, meta–information for the classification. However, the meta–information itself should also be considered to be information. The message–receiver in Shannon's communication theory is assumed to have a set of all possible messages defined in advance. There is no doubt that the message–receiver has obtained this meta–information[3] through some information mechanism. Similarly, an IGUS in general must have obtained meta–information that enables a classification of current information to be obtained. Then it must also have obtained meta–meta–information that enables a classification of the meta–information for classification. Information demands information for classifying it as information. However, it is unrealistic that an IGUS incorporates such an infinite hierarchy of information required for classification just as it is. Hence the language for describing the fundamental mechanism of information should have a capability to represent such an infinity in a "closed" manner.

In category theory, there is a concept called *subobject classifier*. This is a special kind of object of a category. Intuitively, as the name suggests, what a subobject classifier does is to classify *subobjects* of a given object according to which elements belong to the subobject in question. In short, the subobject classifier is *the object that classifies objects*.

First, let us define the subobject of an object in a category on the analogy of subset. Any monic function $f : M \rightarrowtail S$ determines a subset of S, $\mathrm{Im} f = \{f(x) : x \in M\}$. f induces a bijection between M and $\mathrm{Im} f$, hence $M \cong \mathrm{Im} f$. Thus the domain of a monic function is isomorphic to a subset of the codomain. This leads us to the definition of *subobjects*, the categorical versions of subsets.

---

[3]In this case, the set of all possible messages is used as the meta–information for identifying, that is, "classifying", the message that is received by the message–receiver.

Definition 4.10. A *subobject* of an object X in a category $\mathcal{C}$ is a monic

$$m : M \rightarrowtail X.$$

Two monics $M \xrightarrow{m} X, N \xrightarrow{n} X$ are called *equivalent*, and are written as $m \sim n$, if there is an isomorphism $i : M \cong N$ such that $m \circ i = n$. Clearly $\sim$ is an equivalence relation on the class of all monic arrows in $\mathcal{C}$ with codomain X. Each $m : M \rightarrowtail X$ determines an equivalent class

$$[m] = \{n : m \sim n\}.$$

We should *redefine* a subobject of X to be an equivalence class of monics with codomain X. That is:

Definition 4.11. A *subobject* of an object X in a category $\mathcal{C}$ is an entity

$$[m] = \{n : m \sim n\},$$

where $m : M \rightarrowtail X$ is a monic with codomain X.

When no confusion is likely to occur, we shall identify $[m]$ with its representative m, and even with the domain X of m.

Now it is time to turn to the definition of the classifier of subobjects. Again, we should start with the case of subsets.

In set theory, there is a bijective correspondence between subsets of X and functions $X \to 2 = \{0, 1\}$. Given a subset $M \subseteq X$, the *characteristic function of M*, $\chi_M : X \to 2$ is defined by

$$\chi_M = \begin{cases} 1 & \text{if } x \in M \\ 0 & \text{if } x \notin M. \end{cases} \tag{4.14}$$

The correspondence between subset and characteristic function can be "captured" by a pullback diagram. In set theory, the pullback of two set function f and g is defined by putting

$$D = \{\langle x, y \rangle : x \in A, y \in B, \quad \text{and} \quad f(x) = g(x)\} \tag{4.15}$$

with f' and g' as the projections:

$$f'(\langle x, y \rangle) = x \tag{4.16}$$

$$g'(\langle x, y \rangle) = y. \tag{4.17}$$

If $f : A \to C$ is a function, and B a subset of C, then the inverse image of B under f, denoted $f^{-1}(B)$, is that subset of A consisting of all the f-inputs whose corresponding outputs lie in C, *i.e.*

$$f^{-1}(B) = \{x : x \in A \quad \text{and} \quad f(x) \in B\}. \tag{4.18}$$

The diagram

$$
\begin{array}{ccc}
f^{-1}(B) & \xrightarrow{\ f^* \ } & B \\
\big\uparrow & & \big\uparrow \\
A & \xrightarrow{\ f \ } & C
\end{array}
\tag{4.19}
$$

is a pullback square, where the arrows with curved tails denote inclusions, and $f^*(x) = f(x)$ for $x \in f^{-1}(B)$.

Thus the correspondence between subset and characteristic function can be "captured" by the following diagram:

$$
\begin{array}{ccc}
M & \xrightarrow{\ ! \ } & 1 \\
\big\downarrow & & \big\downarrow {\scriptstyle \top} \\
X & \xrightarrow{\ \chi_M \ } & 2
\end{array}
\tag{4.20}
$$

where the function $\top$ is a function from $1 = \{0\}$ to $2 = \{0, 1\}$ and $!$ is the unique function $! : M \to 1$.

This situation (in the category of sets and functions) can be generalized to categories that have components required, such as 1, pullbacks and 2. The one element set 1 in set theory is corresponding to the special object called *terminal object* in general categories. The terminal object and pullbacks are defined as *limits* in category theory, and they exist when the category has all finite limits.

Definition 4.12. An object 1 is *terminal* in a category $\mathcal{C}$ if for every object $C \in Ob(\mathcal{C})$ there is one and only one arrow from C to 1 in $\mathcal{C}$.

Definition 4.13. Consider two morphisms $f \in \mathcal{C}(A, C)$, $g \in \mathcal{C}(B, C)$ in a category $\mathcal{C}$. A *pullback* of (f, g) is a triple (P, f', g') where

1. P is an object of $\mathcal{C}$.

2. $f' : P \to B$, $g' : P \to A$ are morphisms of $\mathcal{C}$ such that $f \circ g' = g \circ f'$,

and for every other triple (Q, f'', g'') where

1. Q is an object of $\mathcal{C}$.

2. $f'' : Q \to B$, $g'' : Q \to A$ are morphisms of $\mathcal{C}$ such that $f \circ g'' = g \circ f''$,

there exists a unique morphism $q : Q \to P$ such that $f'' = f' \circ q$ and $g'' = g' \circ q$. See the diagram 4.21.

$$\tag{4.21}$$

Now we can define the subobject classifier.

Definition 4.14. Let $\mathcal{E}$ be a category with all finite limits. A *subobject classifier* in $\mathcal{X}$ consists of an object Ω together with an arrow $\top : 1 \to \Omega$ that is a "universal subobject," in the following sense: Given any object X and any subobject $M \rightarrowtail X$, there is a unique arrow $\chi(m) : X \to \Omega$ making the following diagram a pullback:

$$
\begin{array}{ccc}
M & \longrightarrow & 1 \\
{\scriptstyle m}\downarrow & & \downarrow{\scriptstyle t} \\
X & \xrightarrow{\chi(m)} & \Omega
\end{array} \quad .
\tag{4.22}
$$

The arrow $\chi(m)$ is called the *characteristic arrow of the subobject* $m : M \rightarrowtail X$ (or *of* M).

Every (elementary) topos, which is a certain class of a category, has a subobject classifier [75, 76]. Hence, if it is possible to restrict the language that is used for a study of information to the language of topos, the study may enjoy the classification capability provided by the subobject classifier.

4.2.5   Topos as a generalization of the category of sets

In addition to having the subobject classifier, a topos can be, in a sense, regarded as a generalization of the category of sets [75, 76]. The category of sets is a topos, and it is often said that one way to understand the notion of the topos is as a category of generalized sets. This property of the topos generally makes the studies built on a topos easier to be considered as a generalization of earlier studies built on sets. For instance, Shannon's communication theory is evidently built on the theory of sets. If someone may build a communication theory on topos, it might be considered to be a generalization of the Shannon's theory. However, mathematically rigorous comparisons of the topos and the category of sets (e.g., [80]) is far beyond our scope, therefore, in this subsection, we mention merely the definition and features of the topos that might convince the readers of the analogy between the topos and the category of sets.

Definition 4.15. A *topos* is a category with:

a) finite limits and colimits,

b) exponentials,

c) a subobject classifier.

Unfortunately, the above short definition mystifies the readers who are not familiar with some category theory, and requires a further sequence of definitions to bring it back to the basic concepts of category theory[4]. Since we do not need such detailed definitions here, as Baez summarizes in [81], let us see a bit about what these items a)–c) amount to in the category of sets:

a) states that there are:

- an initial object (an object like the empty set),

- a terminal object (an object like a set with one element),

- binary coproducts (something like the disjoint union of two sets),

- binary products (something like the Cartesian product of two sets),

- equalizers (something like the subset of X consisting of all elements x such that $f(x) = g(x)$, where $f, g : X \to Y$),

- coequalizers (something like the quotient set of X where two elements $f(y)$ and $g(y)$ are identified, where $f, g : Y \to X$).

In fact, a) is equivalent to all this stuff. It's a theorem that to have finite limits and colimits is the same as to have all the things listed above.

b) states that for any objects x and y, there is an object y^x, called an *exponential*, which acts like "the set of functions from x to y." This means that, as functions which relate sets are themselves sets, a topos also has function–like being as its objects.

---

[4]In the appendix A.2, we have the definitions for the notions that appears in this subsection and that are not defined in the text.

c) assures the existence of the subobject classifier Ω, mentioned already in 4.2.4. Every topos has the subobject classifier which acts like $\{0, 1\}$ in the category of sets. Thus as in sets, objects in a topos may be classified merely using the object and arrows.

With all these facts, a topos may be considered to behave like the category of sets.

4.2.6   Existence of a successful application

Finally, as the next section will be fully devoted to this, one may successfully explain a mechanism of the extraction of information called "attribute" from outer universe in the language of category theory. Though it might be a mere example, the way in which an IGUS obtains attributes from the external universe is considered to be a significant information phenomenon. In the light of its significance and generality of this example, it may be expected that the language of category theory as used in the example may be applied widely to the other information phenomena.

4.3   An internal description of attributes from the inclusion structure of extension

In this section, a formulation of a specific information phenomenon is developed in the language of category theory. The information phenomenon formulated is the classification of the outer world of an IGUS, which is also greatly concerned with information in general. The process contains the attributes extraction process, which occurs in a boot-strapping manner within the IGUS and enables to classify the outer world. In Chapter 3, the authors ground the phenomenon in the connection to the extensive measurement theory in metrology, and by using somewhat familiar mathematics, Chapter 3 demonstrates the validity of the scheme in the case of the attribute extraction. We get along with the line of Chapter 3 in this section, and expand the understanding by a new formulation.
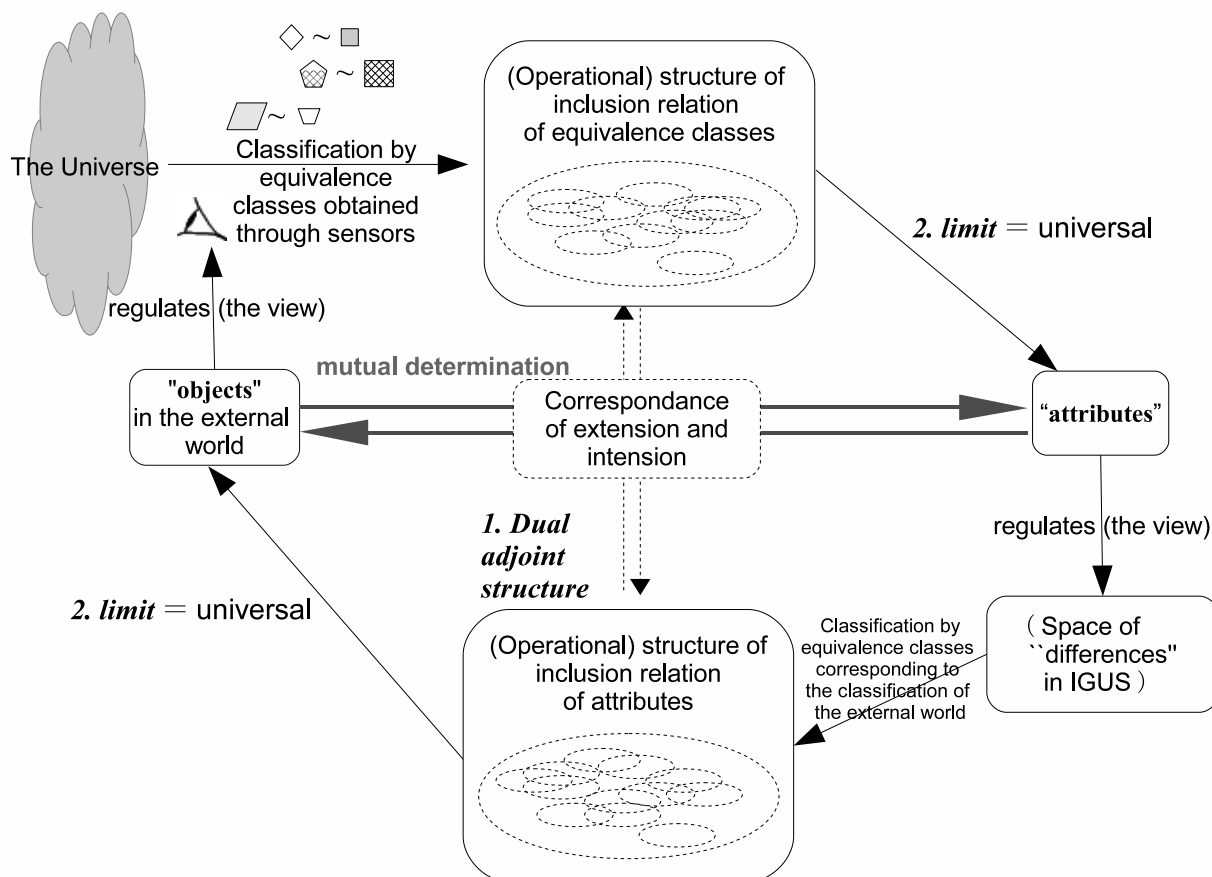
FIGURE 4.1: Model of the process of extensional information articulation

The outline of the idea developed in Chapter 3 is summarized as in Figure.3.5, repeated here as Figure.4.1.

The problems targeted in this scheme are to yield a tolerable classification

- from finite observations;

- with the criterion of classification generated in a boot–strapping manner;

- by a self–generated universal base, *i.e.*, attributes, that may be specify each part of the universe according to the criterion within such a mechanism.

In this scheme, there are two main points.

The first point is the *dual adjoint structure.* The dual structure naturally consists of two structures; the structure of inclusion relation of the equivalent classes of fragments of the universe, and the structure of inclusion relation of attributes. The fragments of the universe are articulated under the restrictions of the capabilities of the sensory motor system of the IGUS under consideration. Indifferences between them are introduced by the sensors that determines the family of equivalence classes of the fragments. This structure is tentative, because the fragments of the universe has transitory nature. By this structure, the objects in the world are established instead of the fragments of the universe, according as the duality is established. In the resultant structure, the attributes extracted from the tentative structure that consists of the fragment of the universe provide a base for classification of objects.

The second point is taking the *limit* to obtain universal, *i.e.*, attributes. Anyway, the IGUS has a structure resulted from sensed stimuli; therefore a possibility is to utilize it for extracting the characteristic values to specify the internal parts of the structure. In Chapter 3, this is accomplished by taking the limit of the (evolutional) sequence of structures. The characteristic values obtained from the limit are called *s–attributes* in Chapter 3. Thus an IGUS may obtain a self–generated universal base *s–attributes* to articulate and classify the universe in a boot–strapping manner, Likewise, objects to be elements that is composing the world of the IGUS may be obtained by taking the limit of the sequence of the s–attribute structures.

Thus the IGUS has two structures, takes the limits of them, establishes the duality, and finally fixes objects and their attributes composing its world.

Readers should consult Chapter 3 or the paper [1] for details.

In this section, the same phenomenon will be formulated by the language of category theory. With this change of formulation, other aspects than what is demonstrated in Chapter 3 are revealed in more obvious manner. A special emphasis is laid on the "local" picture of the contextuality of the classification that is obtained by the language of category theory. The focus in this chapter is the second point mentioned above, that is, the attribute related point. The first point is merely briefly mentioned in Section 4.3.1.

4.3.1   Dual adjoint structure between the tentative classification of the world and attributes

For the formulation of the duality part, there is nothing special to be mentioned in this thesis. This section is essentially a commentary on the well–known facts in mathematics, which have been already applied in various fields other than mathematics.

As in Chapter 3, an IGUS is supposed to be able to articulate the universe to comprehend its environments. *Articulation* by an IGUS means the following steps:

1. An IGUS finds some partial uniformity in the universe;

2. The IGUS divides the universe into fragments accordingly;

3. The IGUS collects the uniform fragments of the universe and distinguishes them from the rest of the universe (The divisions may be done in a multi-layered way);

4. The IGUS forms (a representation of) an inclusion structure of (the representations of) the fragments collected within itself;

5. The IGUS assigns attributes based only on the information which can be utilized by itself, that is, only on the inclusion structure of the fragments of the universe;

6. The IGUS finally obtains concepts and has a capability to refer to the specific fragments of the universe indirectly.

In the mathematical model of this section, the (observed) universe is assumed to be a topological space U, and the third step may be regarded as an introduction of open subsets, *i.e.*, topology, into U. More precisely speaking, we are assuming that the world for an IGUS should be none other than the topological space generated by a certain topology the IGUS itself has settled. In this subsection, as usual in this thesis, we will identify the universe observed by an IGUS with the world for the IGUS.

The structure mentioned in the fourth step is regarded as the lattice $O(U)$ of open subsets of U. The lattice $O(U)$ is complete, and forms a *frame. Frame* is a complete lattice satisfying the infinite distributive law. Frames form a category **Frm** with functions preserving finite meets and arbitrary joins. The category dual of the category of frames is called the category of *locales* **Loc**.

Frm and **Loc** are considered to be adequate representation of the situation, because **Frm** and **Loc** are players in the *pointless topology*, an approach to topology that avoids mentioning points. At this stage of the articulation of the universe, there is no distinguishable object for the IGUS. That is, there is no distinguishable *point* for the IGUS in the model of the universe (for the IGUS). Mathematical formulation should incorporate this situation, and it may be accomplished by the language of pointless topology.

After all the steps above, the IGUS obtains a picture of the universe, *i.e.*, the world for the IGUS, which is full of objects. Mathematically, the objects (which are elementary, that is, not regarded as something like composites) are modeled as points, and the world is modeled as an ordinary topological space that has many points.

The interconnection between the category **Top** of topological spaces whose morphisms are the continuous functions and the category **Frm** is well discussed in mathematics under the name of *Stone–type duality*. The Stone–type duality is the duality between a subcategory of **Top** and a subcategory of **Frm**: the category **Sob** of *sober* spaces with continuous functions and the category **SFrm** of *spatial* frames with appropriate frame morphisms. A *sober* space is a topological space such that every irreducible closed subset of X is the closure of exactly one point of X: that is, this closed subset has a unique generic point.

By this duality, the world U within an IGUS may canonically be restricted to a sober space. Each generic point in the sober space may be identified as an elementary object in the world of the IGUS. Thus the IGUS becomes to be able to canonically refer to the *object* in the world.

4.3.2 Variable sets

In this subsection, for the sake of simplicity, we treat the fragments of universe picked by an IGUS as if they were distinguishable and persistent. In other words, the fragments of the universe are supposed to be *extensive*. Actually, the extension of the universe, *i.e.*, objects for the IGUS should be modeled to be obtained by the duality structure and the dynamics on the structure. But to focus on the characterization of attributes acquired by the new formulation, we will avoid inevitable messy details.

The set P of equivalence classes of the fragments of the universe forms a partially ordered set, also called a *poset*. A partial order on P is introduced as a binary relation $\leq$ on P. $p \leq q$ if and only if the IGUS is required to add some fragments of the universe operationally to p to form q. [5]

Each $p \in P$ may also be considered to be a state of knowledge. If $p \leq q$, q includes more knowledge, that is, roughly speaking, information, than p. We intentionally confuse here the dynamic view with the static view. In other words, there is no distinction between the diachronic view and the synchronic view. On one hand, adding a fragment of the universe to one of the existing equivalent class is dynamic, a change of state. Actually this dynamics builds the poset P. On the other hand, the poset P shows the structure of the classification. This is a static view. In this subsection, the view will be switched freely according to its conveniences.

A poset itself may be regarded as a category $\mathcal{P}$. The objects are the elements of the poset, and an arrow from p to q for $p, q \in P$ exists if and only if $p \leq q$. The category that so arises is called a *poset category*[6].

To represent the contextuality formed by finite observations, the contextuality required for information, one may use a topos, also called a *variable set* [76].

---

[5]Note that, strictly speaking, it is an order relation between sets of operations of the IGUS, at this stage. Though it is irrelevant to the following discussion in this section because the fragments of universe under consideration have been supposed to be distinguishable and persistent, it is important for the reader to be aware that there are delicate difficulties to define the sets of the fragments of the universe by extension, in principle. This difficulty may be removed by the *soberification* induced by the Stone–type duality described in the previous subsection.

[6]A poset category and the underlying poset of it are so similar that we often use the same symbol by an abuse of notation hereafter in this thesis.

Suppose that an IGUS encountered a kind of herring, carpet, tape, card in a classification situation. Each of them is a specific instance, which might be identified by the spatio–temporal occurrence. They seemed similar, hence the IGUS would bundle them into an equivalent class[7]:

$$\varphi = \{herring_{id_2}, carpet_{id_3}, tape_{id_7}, card_{id_9}\}. \qquad (4.23)$$

The IGUS would embody a statement like

$$\varphi(x) = \text{``}x \text{ is red."} \qquad (4.24)$$

which would determine the above set as

$$\varphi = \{x : \varphi(x) \text{ is true.}\} \qquad (4.25)$$

Note that the meaning of the words used in the predicate "is red" should not be taken in ordinary sense here[8]. It is a mere label to represent the similarity among the elements in the set. Precise statement should be such as "*x is found to have some similarity with other elements that may be characterized by an optical effect induced in one of the sensors, and ... .*" The statement has been shortened to "is red" for readability. In ordinary set theory, it is definitely determined whether or not an element belongs to a set by a predicate. However, what an IGUS can utilize to build a predicate is its mere finite observations. But, as is pointed out in 3.1, the consequence brought by a direct connection between the finite observations of the extension and the predicate is entirely improper, or unsatisfactory at least.

The topos, also called the variable set, has a remedy. φ should not be regarded as determining a set per se, but rather as determining the collection

$$\varphi_p = \{x : \varphi(x) \text{ is known at } p \text{ to be true.}\} \qquad (4.26)$$

for each state p. For a given poset P of states of knowledge, the assignment of φ_p

---

[7]Each of the subscript is for the identification of the spatio-temporal occurrence of the situation. These subscripts are written here to emphasize that the situation represented by each of the element is an instance.

[8]and scientifically, such a word of *attribute* cannot have the strictly common meaning that is shared by all IGUSes, as the readers of Chapter 3 well understand now.

to p determines a function $P \to Set$. If $x \in \varphi_p$ and $p \leq q$, it is considered to be natural that $\varphi(x)$ is true also at q, *i.e.*, $x \in \varphi_q$. Thus

$$p \leq q \Longrightarrow \varphi_p \subseteq \varphi_q. \tag{4.27}$$

This means that φ determines a functor from the category of a poset to the category of sets: $\mathcal{P} \to \mathcal{S}et$. The functor assigns the inclusion arrow $\varphi_p \hookrightarrow \varphi_q$ to each $p \to q$ in $\mathcal{P}$. It is known that this functor category $\mathcal{S}et^{\mathcal{P}}$ is a topos [75][76].

4.3.3   Building a variable set step by step

In this section, following [113] which applies the topos theory to the spacetime physics, we describe a topos–theoretic way to model an information phenomenon in a step by step manner. The information phenomenon modeled here is the extensive classification. In this process, an IGUS classifies the fragments of the universe it obtained and yields information, *i.e.*, "attributes" to refer to a component of the classification by only using the fragments themselves, *i.e.*, *extensively*. Mathematically, there is no mystery in the topos building process described in this section. However, the commentary in the context of this study on the mathematical constructions in the process may be of value.

We refer to the case of set–subset relation (in the category $\mathcal{S}et$ of sets). In the set theory familiar to us, a subset can be captured in the words of sets and functions as follows. As already mentioned at the Equation (4.20), a set S is a subset of U, if and only if the following diagram forms the pullback:

$$\begin{array}{ccc} S & \xrightarrow{\ !\ } & 1 \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle T} \\ U & \xrightarrow{\ \chi\ } & 2 \end{array} \quad . \tag{4.28}$$

where 1 is a one element set $\{0\}$, 2 is a two element set $\{0,1\}$, s in the inclusion function of S into U, ! is the unique function from each element of S to 1, T is a mono–function which sends 0 to 1, and χ is a function. The statement that the diagram is pullback implies that $T \circ ! = \chi \circ s$, hence you can easily see that the function χ is the characteristic function of S:

$$\chi(x) = \begin{cases} 1 & \text{for } x \in S \\ 0 & \text{for } x \notin S \end{cases} \tag{4.29}$$

in this situation.

We will expand this understanding to the case of the category of functors from a poset category to the category of sets: $\mathcal{P} \to \mathcal{S}et$, also written as $\mathcal{S}et^{\mathcal{P}}$. In this category, a topos, the set *varies* with the element in $\mathcal{P}$, therefore an object in this category might be thought as a *variable set*. This category provides an interesting generalization of ordinary sets.

In the following steps, we will focus on the *set side* of the functors first. That is, we will slight the *arrow part* for a while. By this emphasis, we will try to make the similarity with the $\mathcal{S}et$ explicit. After the similarity is recognized, we will recover the citizenship of the arrow part and establish the construction as functors.

4.3.3.1   Functor V

Assume that the collection P of the concepts is a finite partially ordered set[9].

The state of knowledge at the concept p within the context P, that is, the idea on the world of the IGUS that has the concept p, ought to be represented by the down–set of p. This defines the functor $V \colon \mathcal{P} \to \mathcal{S}et$, whose object part is

$$\text{for any } p \in \mathcal{P}, V(p) = \{r \in \mathcal{P} : r \leq p\}, \tag{4.30}$$

and arrow part is

$$\text{for any arrow } p \to q \ , V_{p,q} : V(p) \to V(q). \tag{4.31}$$

V is considered to be an object of the category $\mathcal{S}et^{\mathcal{P}}$. Arrows of this category are natural transformations, which will be defined later.

---

[9]As a theoretical framework of a model, we might extend this assumption to be a more general setting, such as a discrete poset and a locally finite poset. However, because all the experiences of an IGUS is finite, it is not so unnatural to assume that its internal model is finite.

This encoding means that in the view[10] of the IGUS in the state p, the lower concepts of p are also in scope. In other words, "if the IGUS knows the concept p, then it also knows the lower concepts included in p"[11].

4.3.3.2   Terminal object $t.o.$

To establish a similar relation in the case of subset (Equation (4.28)), it is required to find the object that plays the same role with $1 = \{0\}$ in the Equation (4.28). In $\mathcal{S}et$, every (sub)set S is provided with exactly one function from S to the one element set, 1. To provide every (sub)object in the same manner, a special object corresponding to the one element set is expected. Such an object is called the *terminal object.*

The terminal object of the category $\mathcal{S}et^{\mathcal{P}}$ is the functor $t.o. : \mathcal{P} \to \mathcal{S}et$ whose object part is

$$\text{for any } p \in \mathcal{P}, t.o.(p) = \{0\} , \tag{4.32}$$

and arrow part is

$$\text{for any arrow } p \to q \, , t.o._{p,q} : \{0\} \to \{0\} \text{ when } p \leq q. \tag{4.33}$$

The terminal object forms the same web of relations as in $\mathcal{P}$, but every concept is replaced with $\{0\}$.

4.3.3.3   Subobject classifier Ω

What the poset P encodes is whether a concept includes another concept or not. Suppose that an IGUS has a concept p, and we would like to discuss the relation between the p and another concept r for the IGUS. If $r \leq p$, then the concept r is included within the concept p. If $r \nleq p$, there seems no information on r that may be represented at p, but it is not true. Even in the case of $r \nleq p$, if there is

---

[10]Functor symbol "V" is for view

[11]Though such a personification will appear frequently in the followings, do not take them too literally.

q that includes both p and r, by considering the inclusion relation $p \to q$ as the "distance" to including r, one can represent the relation between p and r at p, *i.e.*, the truth–value on the inclusion of r at p. This arrow $p \to q$ is a candidate for the truth–value at p, for r to be included. Naturally, all the arrows $p \to q'$ for all $q' \geq q$ should also be included in the truth–value when $p \to q$, because when q includes both p and, say, r, $q' \geq q$ also includes them, Moreover, there will not be a single first concept q in the common *super*-concept of p and r, but a set of such concepts. That is, there might exist many q_i, $i \in \{1, 2, \ldots\}$ that includes both p and r. These, and all the concepts that includes them, should be the inclusion relations in the truth–value at p for r.

Let us formalize this. Let us define $R(p)$ as the set of all inclusion relationships that start at p,

$$R(p) := \{\text{all } p \to p' : p \leq p' \text{ in } \mathcal{P}\} \tag{4.34}$$

for all the objects p. Next, we define subsets $S(p)$ of $R(p)$ such that, if $S(p)$ contains the inclusion relation $p \to q$, it should also contain $p \to q'$, for every $q' \geq q$. $S(p)$ is therefore a left–composable subset of $R(p)$ since the above definition means that if $(p \to q) \in S(p)$, and there is $(q \to q') \in \mathcal{P}$, then $(p \to q') = (q \to q') \circ (p \to q) \in S(p)$. That is, $S(p)$ is closed under left composition. Such a set of arrows $S(p)$ is called as a *sieve on p* or *p–sieve* in some literature[12].

The truth–value at p that we are looking for is precisely represented as a sieve. It is not simply a yes/no truth–value, but also tells us at p when, or how far, r will be included in the same concept with p.

Therefore, the set of all the sieves is the set of truth–values at p. We denote this set as $\Omega(p)$,

$$\begin{aligned}\Omega(p) &:= \{p\text{–sieves}\} \tag{4.35}\\ &= \{S(p) : S(p) \subseteq R(p)\}. \tag{4.36}\end{aligned}$$

Such an Ω constructed in this way, that is, the *subobject classifier*, uniquely exists up to isomorphism.

---

[12]Note that, in the different convention, it is called *cosieve*.

4.3.3.4   Total truth arrow and characteristic arrow

A special subset of the concepts in $\mathcal{P}$ is, naturally, those in $V(p)$. There is a sieve in $\Omega(p)$ that corresponds to these particular concepts.

They are special because at p they have already "known" to the IGUS. While the truth–value for $r \notin V(p)$ may be considered to be a *partially true* value, the sieve for the concepts in $V(p)$ is the *totally true* value. This understanding can be used to find the sieve in $\Omega(p)$ that corresponds to the concepts in $V(p)$ and thus the set $V(p)$ itself. Since these concepts is included at p and at all concepts that include p, the truth–value sought for is the whole $R(p)$, the *maximal sieve* in $\Omega(p)$ (as the general topos theory tells us).

Thus we define the *totally true* arrow at p,

$$T_p : \{0\} \to \Omega(p) \tag{4.37}$$

such that

$$T_p(0) = R(p) = \text{maximal sieve on } p. \tag{4.38}$$

Next, let us consider $V(p) \subseteq \mathcal{P}$, and name this inclusion ι_p,

$$\iota_p : V(p) \hookrightarrow \mathcal{P}. \tag{4.39}$$

As both $V(p)$ and (the object part of) $\mathcal{P}$ are sets, the fact that the former is a subset of the latter is represented as the following pullback diagram:

$$\begin{array}{ccc} V(p) & \longrightarrow & \{0\} \\ \downarrow{\scriptstyle \iota_p} & & \downarrow{\scriptstyle T_p} \\ \mathcal{P} & \xrightarrow{\;\chi_p\;} & \Omega(p) \end{array} \quad . \tag{4.40}$$

That is, $V(p)$ contains those concepts $r \in \mathcal{P}$ that satisfy

$$\chi_p(r) = (\text{maximal sieve on } p) = T_p(0). \tag{4.41}$$

4.3.3.5 Constructing a functor

Now we have the (prospected) set of truth–values $\Omega(p)$ at a given concept p which is the set of sieves on p, and the structure, *i.e.*, pullback diagram, that represents the subset relation. Next, these results are "varied" over the poset category $\mathcal{P}$ so that the functor V may be defined by a generalization of what was done for $V(p)$. Being in harmony with the move from $V(p)$ to $V(q)$, the appropriate truth–values are given by $\Omega(p)$ and $\Omega(q)$, so that $V(q)$ can be found by given $V(p)$, $p \leq q$ and $\mathcal{P}$. That is, the following diagram commutes:

$$\begin{array}{ccc} V(p) & \xrightarrow{\iota_p} & \mathcal{P} \\ {\scriptstyle V_{p,q}}\downarrow & & \downarrow{\scriptstyle id_\mathcal{P}} \\ V(q) & \xrightarrow[\iota_q]{} & \mathcal{P} \end{array} \qquad . \tag{4.42}$$

The lefthand side $V(p) \overset{V_{p,q}}{\rightarrow} V(q)$ shows the components of the functor V. The righthand side may be varied in the same way by constructing the constant functor

$$World : \mathcal{P} \to \mathcal{S}et \tag{4.43}$$

with

$$\begin{aligned} World(p) &= \mathcal{P} & \text{for any } p \in \mathcal{P}, & \tag{4.44} \\ World_{p,q} &= id_\mathcal{P}. & & \tag{4.45} \end{aligned}$$

The diagram is redrawn as

$$\begin{array}{ccc} V(p) & \xrightarrow{\iota_p} & World \\ {\scriptstyle V_{p,q}}\downarrow & & \downarrow{\scriptstyle World_{p,q}} \\ V(q) & \xrightarrow[\iota_q]{} & World \end{array} \qquad . \tag{4.46}$$

The functor V is a *subfunctor* of *World*, that is, there is an inclusion

$$\iota : V \to World, \tag{4.47}$$

which is a natural transformation. That is, ι has components $\iota_p.\iota_q$ that make the above diagram commutes, *i.e.*, $\iota_q \cdot V_{p,q} = World_{p,q} \cdot \iota_p$ for every p, q that are related by an inclusion relation.

The set of truth–values for $W(p)$ is $\Omega(p)$ while for $W(q)$ it is $\Omega(q)$. These are also components of the functor

$$\Omega : \mathcal{P} \to \mathcal{S}et. \tag{4.48}$$

For each concept in $\mathcal{P}$, Ω provides its set of truth–values, *i.e.*, the sieves on that concept. The arrow part of the Ω, the set of functions $\Omega_{p,q} : \Omega(p) \to \Omega(q)$, provides the set of sieves on q for the given set of sieves on p.

The characteristic arrow χ is a natural transformation from $World$ to Ω,

$$\chi : World \to \Omega. \tag{4.49}$$

It has components at each concept that compose as ι did above.

$$
\begin{array}{cc}
p & World(p) \xrightarrow{\chi_p} \Omega(p) \\
\Big\downarrow f & \scriptstyle World(f)= \Big\downarrow \scriptstyle World_{p,q}=id_{\mathcal{P}} \quad \Big\downarrow \scriptstyle \Omega(f)=\Omega_{p,q} \\
q & World(q) \xrightarrow[\chi_q]{} \Omega(q)
\end{array} \qquad . \tag{4.50}
$$

The particular subfunctor V is obtained when, at each concept p in $\mathcal{P}$, χ maps into the maximal sieve at p. In other words, the function T_p above is the p–component of the natural transformation from the terminal object $t.o.$ for $\mathcal{S}et^{\mathcal{P}}$ to Ω.

$$T : t.o. \to \Omega, \tag{4.51}$$

with

$$T_p : t.o.(p) = \{0\} \to \Omega(p). \tag{4.52}$$

Finally, we have everything required to give V as the subfunctor. V, *World*, *t.o.*
and Ω are objects in the category $\mathcal{S}et^{\mathcal{P}}$. And with the natural transformations,
ι, T and χ, we may construct the diagram

$$
\begin{array}{ccc}
V & \longrightarrow & t.o. \\
\downarrow{\scriptstyle \iota} & & \downarrow{\scriptstyle T} \\
World & \underset{\chi}{\longrightarrow} & \Omega
\end{array}
\qquad (4.53)
$$

and ask that the functor V makes this diagram pullback. This means that it
reduces to the diagram of sets consistently at each concept p in $\mathcal{P}$. When this
requirement is met, V is a subfunctor of *World*.

4.3.3.6   Summary of the construction

Overall, we have achieved to express the extensional inclusion structure in terms
of lower concepts, *i.e.*, included concepts, *at each concept*. This new expression
is regarded as more satisfactory in the view of foundations of information than
one with the entire structure. Each concept in the context may be evaluated its
degree of inclusion at each concept, *i.e.*, at each state of knowledge p. This local,
internal view is well suited to describe the informational classification process of
an IGUS contextually.

As something extra, even when the inclusion structure is not a lattice, the variable
sets over it form a lattice. They satisfy a particular algebra, called a Heyting
algebra, whose operations reflect the underlying poset, *i.e.*, context. Thus, we can
give the information structure *algebraically*. Though this matter is not pursued in
this thesis, this algebraic property seems to be very promising.

4.3.4   Concrete example

In this section, the process in the previous section is demonstrated by a concrete example. Let us consider the following inclusion structure:
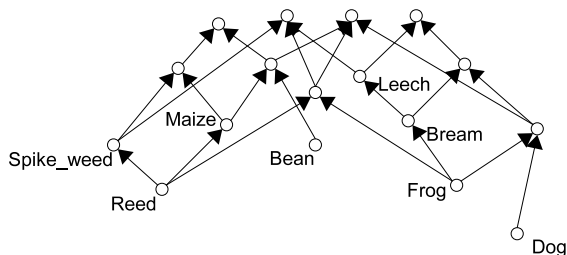
FIGURE 4.2: Living Beings and Water
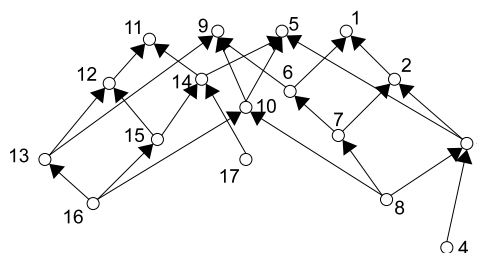
FIGURE 4.3: Example poset

For simplicity, we name each element in the poset by number as Figure.4.3. In this Hasse diagram, members of each concept is included upwards. In Figure.4.2, only additional members are specified in the diagram. All the members in each concept are, for instance,

$$\text{Concept } 4 \quad = \quad \{\text{Dog}\}, \tag{4.54}$$

$$\text{Concept } 8 \quad = \quad \{\text{Frog}\}, \tag{4.55}$$

$$\text{Concept } 3 \quad = \quad \{\text{Dog, Frog}\}, \tag{4.56}$$

$$\text{Concept } 7 \quad = \quad \{\text{Frog, Bream}\}, \tag{4.57}$$

$$\text{Concept } 2 \quad = \quad \{\text{Dog, Frog, Bream}\}, \tag{4.58}$$

and so on.

In the followings, as above, we will only show some instances instead of lengthy complete data. All of the data are exhibited in the tables in the Appendix A.3. As for the above, see Table.A.1.

4.3.4.1   Functor V

In this case, the object part of the functor V is defined as:

$$V(1) = \{1, 2, 3, 4, 5, 6, 7, 8\}, \tag{4.59}$$
$$V(2) = \{2, 3, 4, 7, 8\}, \tag{4.60}$$
$$V(3) = \{3, 4, 8\}, \tag{4.61}$$
$$V(4) = \{4\}, \tag{4.62}$$
$$V(5) = \{3, 4, 5, 8, 10, 14, 15, 16, 17\}, \tag{4.63}$$
$$V(6) = \{6, 7, 8\}, \tag{4.64}$$

and so on. See Table.A.2 in the Appendix A.3 for all the data.

The arrow part of V maps an order relation $f : p \to q$ to the inclusion function $g : V(p) \to V(q)$. For instance, the arrow $f : 3 \to 2$ in the category $\mathcal{P}$ is mapped to an arrow in $\mathcal{S}et$, *i.e.*, the inclusion function $g : V(3) \to V(2)$ defined as

$$g(3) = 3, \tag{4.65}$$
$$g(4) = 4, \tag{4.66}$$
$$g(8) = 8. \tag{4.67}$$

4.3.4.2   Sieves and the subobject classifier Ω

Concrete sieves are rather lengthy. We only mention a few here.

In the following, each $R(p)$ is a maximum p–sieve. $Sx(p) \subseteq R(p)$ are other sieve on p, and the number x suggests the characteristic concept (end–point) in that sieve.

$$R(1) = \{1 \to 1\}, \tag{4.68}$$
$$R(2) = \{2 \to 2, 2 \to 1\}, \tag{4.69}$$
$$S1(2) = \{2 \to 1\}, \tag{4.70}$$

$$R(3) = \{3 \to 3, 3 \to 2, 3 \to 5, 3 \to 1\}, \tag{4.71}$$

$$S1(3) = \{3 \to 1\}, \tag{4.72}$$

$$S2(3) = \{3 \to 2, 3 \to 1\}, \tag{4.73}$$

$$S5(3) = \{3 \to 5\}, \tag{4.74}$$

$$R(4) = \{4 \to 4, 4 \to 3, 4 \to 2, 4 \to 5, 4 \to 1\}, \tag{4.75}$$

$$S1(4) = \{4 \to 1\}, \tag{4.76}$$

$$S2(4) = \{4 \to 2, 4 \to 1\}, \tag{4.77}$$

$$S3(4) = \{4 \to 3, 4 \to 5, 4 \to 2, 4 \to 1\}, \tag{4.78}$$

$$S5(4) = \{4 \to 5\}, \tag{4.79}$$

$$R(5) = \{5 \to 5\}, \tag{4.80}$$

and so on. See Table.A.3 in the AppendixA.3 for all the data.

The subobject classifier for a concept p is the set of p–sieves. For instance,

$$\Omega(4) = \{R(4), S1(4), S2(4), S3(4), S5(4), \emptyset\}, \tag{4.81}$$

where each sieve is defined above, and $\emptyset$ is the empty sieve. See Table.A.4 in the Appendix A.3 for all the data.

Every concept in $\mathcal{P}$ may classified at p by $\Omega(p)$ through the characteristic arrow χ. At 4, for instance,

$$\chi_4(x) = \begin{cases} R(4), & \text{for } x = 4; \\ S1(4), & \text{for } x = 1, 6; \\ S2(4), & \text{for } x = 2, 7; \\ S3(4), & \text{for } x = 3, 8; \\ S5(4), & \text{for } x = 5, 10, 14, 15, 16, 17; \\ \emptyset, & \text{for } x = 9, 11, 12, 13. \end{cases} \tag{4.82}$$

All the concepts in $\mathcal{P}$ are classified into six groups at 4. At 4, the concept 4 is naturally a true subconcept. However, Some of the other elements are also subconcepts partially. For example, the concept 1 and 6 have the same status of partial inclusion to 4, represented by the sieve $S1(4)$.

Thus, by the language of the topos, one may describe the inclusion status locally, from the internal view. The inclusion status may be regarded as the difference induced within the context, that is, information within the context. By the language of the topos, contextual information may be well-described from the internal view.

4.3.5 Attributes revisited

In this subsection, we revisit the attributes studied in Chapter 3 and cast new light on that matter from this internal view.

In Chapter 3, attributes are understood as structural existence obtained from the extensional structure. A certain limit operation explains the universal existence other than the fragments of the universe themselves, and the universality of this new existence, *i.e.*, attributes, is utilized to specify each of the equivalent class of the fragments of the universe. The most important point is that the attributes to be used as information for classification are obtained merely from the structure of extension.

In this subsection, we could have interpret such a discussion in Chapter 3 into a categorical formulation. However, we take another route because merely interpreting the traditional mathematics of this kind in the language of category theory attracts little interest here.

Instead, in this new formulation, let us explore new characterization of the "attributes" in more general structural setting, *i.e.*, in this category theoretical setting. By this category theoretical setting, it will be implied that the concept of attributes may be generalized in the other structure than the inclusion relation. Furthermore, at the end of this chapter, the readers will hopefully be convinced that a *local* characterization of the attributes[13] is possible as an alternative to the *global* characterization of the attributes[14] described in Chapter 3.

We focus on one of the functions of attributes, namely, the function of specifying concepts. For each concept p, IGUS requires (informational) resources to specify it. Since it is unrealistic to require the whole knowledge structure to specify

---

[13]At the stage of this thesis, the reason to pursue such a characterization is a mostly aesthetic one. The implication ought to be explored further in future researches.

[14]Recall that characterizing the attributes by the projective limit requires the whole structure of the concepts.

each of the concepts, we should assume that, at a knowledge state p, the utilizable resources for the IGUS are restricted to the resources around p. Thus the available resources for the IGUS are $V(p), \Omega(p), p \to q$, and so on.

A clue to specify p is naturally in the subobject classifier. The subobjcet classifier classifies p in the whole context. It has sieves in it. As we have seen in previous sections, the sieves are built from concepts and reflect the whole inclusion structure of the concepts. Hence if the IGUS "utilizes" $\Omega(p)$, then the IGUS may specify p with reflecting the whole context of p. A candidate that may be utilized by the IGUS to label the concept p is the *difference* between $\Omega(p)$ and its neighbor $\Omega(q)$[15], because if there is a difference, the difference indicates a specific character of p in the context.

First, we introduce a new symbol:

Definition 4.16. $\ker \Omega_{p,q} = \{S_p \in \Omega(p) : \Omega_{p,q}(S_p) = \{\emptyset\}\}$.

Roughly speaking, $\ker \Omega_{p,q}$ is a set of sieves that are "dropped" when Ω moves from p to q. In a poset, the greater neighbour q of an element p, or, more precisely, $q \geq p$ for which if $p \leq r \leq q$ then $r = p$ or $r = q$ holds, is called a *cover* of p. Intuitively, if

$$\ker \Omega_{p,q} \neq \emptyset \text{ for some } q \text{ covering } p, \tag{4.83}$$

the IGUS has an opportunity to use these "to be dropped" sieves to specify p, because their elimination is specific to p. If

$$\ker \Omega_{p,q} = \emptyset \text{ for any } q \text{ covering } p, \tag{4.84}$$

there is no *difference* to be utilized for the specification of p at p. In this case, the IGUS inevitably needs additional information specific to p, that is, an *attribute* of p. Thus, in the view formulated in this paper, attributes may be understood as *the fillings for lack of difference*.

Actually, the condition Equation (4.84): "$\ker \Omega_{p,q} = \emptyset$ for any q covering p" implies a necessity of an attribute specific to p. To investigate it, we introduce

---

[15]Recall that an arrow in a category is interpreted as a difference between the objects it connects.

another symbol:

Definition 4.17. A binary relation on the elements in $\mathcal{P}$, $\overset{d}{\ll}$ is defined as follows:

$$\left(p \overset{d}{\ll} q \right) : \Longleftrightarrow (\text{ for any } r \in [p) \text{ there exists a } q_r \in [q) \text{ such that } r \leq q_r),$$

(4.85)

where $[p) := \{q : p \leq q\}$.

When $p \overset{d}{\ll} q$, we should say that p *is dangling from q*.

If the poset underlying $\mathcal{P}$ has maximal elements, the situation is clearly described by them. In our case, the poset should be assumed to be finite, so the condition meets. We should introduce the notation for the set of maximal elements in $[p)$.

Definition 4.18.

$$\text{max.} [p) := \{\check{p} \in [p) : [\check{p}) \cap [p) = \{\check{p}\}\} .$$

(4.86)

The following proposition obviously holds.

Proposition 4.19.

$$\left(p \overset{d}{\ll} q \right) \Longleftrightarrow (\text{max.} [p) \subseteq \text{max.} [q)) .$$

(4.87)

The dangling relation is a partial order relation, and we may define an equality on it:

Definition 4.20.

$$\left(p \overset{d}{=} q \right) : \Longleftrightarrow \left(p \overset{d}{\ll} q \right) \text{ and } \left(q \overset{d}{\ll} p \right).$$

(4.88)

Naturally, the following proposition holds.

Proposition 4.21.

$$\left(p \overset{d}{=} q \right) \Longleftrightarrow (\text{max.} [p) = \text{max.} [q)) .$$

(4.89)

In the notation here, a p–sieve S_p is a subset of $[p)$, *i.e.*, $S_p \in \{[\sigma_p) : \sigma_p \in [p)\}$, that is closed "upwards" under $\leq$, *i.e.*, if we have that

$$\text{whenever } q \in S_p \text{ and } q \leq r, \text{ then } r \in S_p.$$

The following proposition is easily seen to hold:

Proposition 4.22. *Let $\mathcal{P}$ be a finite poset category and $p, q \in \mathcal{P}$. For the subobject classifier $\Omega_{p,q}$ of the variable set $\mathcal{S}et^{\mathcal{P}}$, where $\mathcal{S}et$ is the category of sets and functions,*

$$(\ker \Omega_{p,q} = \emptyset \text{ for any } q \text{ covering } p) \iff \left(p \overset{d}{=} q \text{ for any } q \text{ covering } p\right). \quad (4.90)$$

Proof. $\ker \Omega_{p,q} = \emptyset$ means that there is no p–sieve $\left[\sigma_p^q\right)$ which satisfies $\left[\sigma_p^q\right) \cap [q) = \emptyset$. It implies $\max.\left[\sigma_p^q\right) \subseteq \max.[q)$ for every p–sieve $\left[\sigma_p^q\right)$. Especially, $\max.[p) \subseteq \max.[q)$ holds. Since q is a cover of p, every elements of $\max.[q)$ belongs to some p–sieve., *i.e.*, $\max.[q) \subseteq \max.[p)$. Thus, $\max.[p) = \max.[q)$ holds.
Conversely, if $\max.[p) = \max.[q)$ holds for any q covering p, then obviously there is no p–sieve $\left[\sigma_p^q\right)$ which satisfies $\left[\sigma_p^q\right) \cap [q) = \emptyset$.
Therefore,

$$(\ker \Omega_{p,q} = \emptyset \text{ for any } q \text{ covering } p) \iff (\max.[p) = \max.[q)).$$

On the other hand, the Equation (4.89) says that $\left(p \overset{d}{=} q\right) \iff (\max.[p) = \max.[q))$. Thus, the Equation (4.90)

$$(\ker \Omega_{p,q} = \emptyset \text{ for any } q \text{ covering } p) \iff \left(p \overset{d}{=} q \text{ for any } q \text{ covering } p\right)$$

holds. $\qquad\square$

The proposition 4.22 explicates that the condition the Equation (4.84): "$\ker \Omega_{p,q} = \emptyset$ for any q covering p" implies a necessity of an attribute specific to p. Let us interpret the right–hand of the Equation (4.90) in the view of this thesis.

To make the things straight and simple, suppose that the poset category $\mathcal{P}$ forms a finite chopped lattice[16]. As is known well in the lattice theory [70, 82], every

---

[16]A chopped lattice M is obtained from a finite lattice L with unit 1, and chopping the unit element: $M = L - \{1\}$.

elements in a lattice is specified as a meet of *meet–irreducible* elements[17]. Hence if each of meet–irreducible elements bears a corresponding attribute, other elements may be specified by the combination of these attributes[18]. That is, an attribute corresponds to the meet–irreducible element specific to the attribute.

When $p \stackrel{d}{=} q$ for any q covering p, we should consider two cases:

1. there is only one covering q;

2. there are multiple covering $q, q', \ldots$, where $q \neq q'$, and so on.

The first case means that p is meet–irreducible[19]. In the second case, we naturally have $q \stackrel{d}{=} q' \stackrel{d}{=} \cdots$, and $\max. [q) = \max. [q') = \cdots$. However, this never happens in the case of a lattice because if that is the case in a poset, the poset does not form a lattice [20].

Thus, in the case of a poset category formed by a finite chopped lattice,

$$\ker \Omega_{p,q} = \emptyset \text{ for any } q \text{ covering } p$$

actually means that p corresponds to the attribute specific to p.

In the case of general finite posets, let us consider what the second case above means. $p \stackrel{d}{=} q$ for any (more than one) q covering p may be considered to be interpreted that p *has redundant parts of the context to specify itself*. Comparison with the case of the lattice makes it clearer. To form a lattice, every element p of a poset is expected to be specified uniquely by the meet of a pair of meet–irreducible elements that are specific to p. Conversely, a pair of meet–irreducible elements characterizes an element *uniquely*. This may be interpreted in the view of this

---

[17]In a lattice L, an element $x \in L$ is meet-irreducible if $x \neq 0$ (in case L has zero), and $x = a \wedge b$ implies $x = a$ or $x = b$ for all $a, b \in L$.

[18]For instance, suppose that a,b,c are concepts, then the attributes of a concept $c = a \wedge b$ are the union of the attributes of a and b. The readers should consult [109] about these circumstances.

[19]An element in a lattice is meet–irreducible if and only if it is covered by at most one other element.

[20]It is simply understood in the case of q and q' have more than one cover. In this case, q has at least two elements in $\max. [q)$ whose one of maximal lower bounds is q. However, because $\max. [q) = \max. [q')$, the elements also has q' as one of their maximal lower bounds. Thus the greatest lower bound of these two elements in $\max. [q)$ does not exist, and the poset does not form a lattice.

study as follows:

> In a lattice, merely two parts of the context, *i.e.*, elements within
> the lattice, are required to characterize an element, p for instance, in
> the lattice. The parts are called meet–irreducible elements and they
> specify the element p uniquely. Thus the meet–irreducible elements
> are corresponding to the information called "attribute" which an IGUS
> may utilize to characterize the world.

In case of the second case occurs, in parallel to the lattice case, the situation may
be interpreted as follows:

> In a structure in general, more specifically, a partial ordering relations
> in this thesis, there are many parts of the context to characterize an
> element, for instance, p. However, in contrast to the case of lattices,
> the element p might not be uniquely specified by these parts. If an
> IGUS tries to specify p within the structure, the IGUS needs additional
> information to distinguish among the elements characterized by the
> same parts of the context. The role of this information is the same as
> the role of the meet–irreducible elements for p in the case of lattices.

Thus, in both cases, we may say that *information is required to specify at p where
the condition that* $\ker \Omega_{p,q} = \emptyset$ *for any q covering p holds.*

Although, mathematically, the above discussion is merely a change in the wording
and the expressions such as $\ker \Omega_{p,q} = \emptyset$ might be pretentious as just a condition for
the meet–irreducibility of p, by expressing in the language of category theory, they
provide insights into a generalization of the concept of attribute in the structure
other than the inclusion relations. As we have observed, $\ker \Omega_{p,q} = \emptyset$ means the
lack of available contextual resources to specify the element p in the context. The
right–hand side of the Equation (4.90) that is specific to the poset structures may
be interpreted as the condition that states that all the sieves at every covering
have eventually same ends, because the Equation (4.89) tells that is the case
for the poset structures. The meaning of "covers" and "having eventually same
ends" must differ in every structure, and another consideration must be required.
However, thinking in line with it for other concept structures such as the network
structure may be of value.

4.4   Conclusive remarks on this chapter

In this chapter, it is presented that the language of category theory, especially the topos, is effective for investigating the information phenomena. Basics of category theory is introduced, and it is shown that its features, such as the relativity represented by arrows, the structural contextuality of the elements, the structural universality, and the existence of the subobject classifier in some categories, are endowed with the properties suitable to investigate the information phenomena.

Moreover, in this chapter, the author demonstrates an application of the language of category theory to a typical, important information phenomenon. The studied information phenomenon is the attribute extraction for classification. The category used for modeling this phenomenon is the topos. The topos is a kind of category, and it has especially desirable properties for modeling the information phenomena and is considered to be a generalization of the category of sets. By using the language of topos for modeling the attribute extraction, a new, internal description is obtained and this description provides a new understanding of the attribute as information.

Chapter 5

Torsorial Nature of Representations

5.1   Introduction

In the episode 1.1.3, each of the children who is watching a blot on the wall finds a different letter individually. We can say that these children have different *views* for the blot. And even the way articulating the world depends on the one's *view*. That is to say, *the* blot may differ from child to child.

At a certain level of understanding, an IGUS[1] such as the children mentioned above can be considered to be an organizationally closed system. In cybernetics, a system is said to be *organizationally closed* if its internal processes produce its own organization, thus continuously rebuilding the system's identity in a changing environment [41]. An IGUS as a closed information system must interpret the differences of the physical stimuli of the outer world into information for its own by utilizing the differences within itself, and incorporate the information into itself [1]. In other words, an IGUS actively projects its internal view to obtain the information of the outer world. The view is projected on the outer world to articulate and interpret events in the outer world from the IGUS's standpoint. Thus what we (as IGUSes) deal with are always interpretations, *i.e.*, representations, of the outer world through one's own view. The world itself or *thing in itself* cannot be directly dealt with or referred to.

---

[1]Information Gathering and Using/Utilizing System. See 1.5.

Some researchers say that information, or more specifically, representation, should be regarded as something resides in between. To take some instances, Bateson [37] says that "information is··· difference which makes difference," and Heylighen [41] says that "a representation··· constitutes an interface, it stands in between Mind and Nature, in between subject and object, in between Self and World." Bateson gives a vague expression, while Heylighen illustrates what are adjoining the representation.

However, thinking carefully from a certain standpoint, that is, the standpoint of the present chapter, this is found to bring us some difficulty. Mind and Nature, subject and object, Self and World, all of them are representations (for an IGUS) themselves. Each of them also stands in between something. We (as IGUSes) always talk only about something in between what we cannot directly refer to. It is a common difficulty, part of which can easily confirmed by casting a glance at tautologous definitions on a dictionary. To make matters worse, these representations are carved out of the world from different points of view, even when we are talking about superficially an identical thing.

In this chapter, a toy model using an algebraic structure, namely, *torsor*, is introduced to describe and investigate an aspect of such representations. Toy model here means a simple mathematical model that reflects simplified but essential properties of target phenomena. It is used to illustrate a mechanism in order to make the phenomena easier to visualize and investigate. This research constructs the toy model by establishing a connection between the properties of representation and the structure of the algebra. By using the torsor, the toy model aims to describe naturally the representation related phenomena incorporating the *betweenness* of representation.

The structure of this chapter is as follows: Section 5.2 reviews two important concepts in this study: the betweenness of representation and the closure. In Section 5.3, the algebraic structure used in our toy model, namely, torsor, is introduced. The toy model is described in Section 5.4. Section 5.5 exemplifies the toy model by several cases which are simple, but may have broad scope. Section 5.6 concludes this chapter.

5.2   Representation as Janus-Faced Closure

Several studies capture information in ternary relations. In semiology, the central notions are Signe–Signifiant–Signifié (Sign–Signifier–Signified), and in semiotics, they are Sign/Representamen–Object–Interpretant. As briefly quoted in the Introduction 5.1, in systems theoretic study of cognitive science, Heylighen [41] says that:

> a *representation* belongs neither to the realm of matter, of outside objects, of things-in-themselves, nor to the realm of pure mind or Platonic Ideas: it *constitutes an interface*, it *stands in between* Mind and Nature, in between subject and object, in between Self and World. (Emphasis added by the author of this thesis)

In these studies, a representation (which is almost equivalent to a sign in this chapter) is placed between something within an IGUS and (what corresponds to) something in the outer world.

To fix the terms and make the concept of representation discussed in this study clearer, an ontology of representation is presented next in Section 5.2.1.

Another important concept in this study is *closure*. Suppose that "everything is a representation," then a simple thought on a representation mechanism requires *the representation of representations*, *the representation of the representations of representations*, $\cdots$. That is, we need an infinite hierarchy of representations which consists of meta-representations, meta-meta-representations, and so on. This is not suitable for a fundamental mechanism of IGUS as a finite physical existence. To avoid this infinite hierarchy, we introduce the *closure*. Closure also provides a mechanism to extract a stable structure from the outer world that is evasive, indeterminate and hard to figure out directly. To form a closure, the interface nature (*i.e.*, the betweenness) of the representation plays an essential role.

General definition of the closure is described in Section 5.2.2.

Summing up beforehand, in this chapter, we consider a representation as what constitutes an interface that stands in between the outer world and the inner world of the IGUS under consideration. In the term of the ontology, the former corresponds to, or, can be regarded as, *the Representational form*, and the latter corresponds to *the Proposition*. They can be other representations, or, especially the former, can be substrata that are never directly referred to. (The term *substratum* is introduced in the next subsection.)

A solution for how to constitute a representation as an interface between substrata is provided by the closure. A representation can be considered to be a stable closure formed between indeterminate constituents, a part of outer world and the inner world of the IGUS.

5.2.1   Ontology of Representation

Mizoguchi [115] defines a conceptual model of representation (mainly for engineering purposes). In this subsection, the author sketches his model. Because the rough meaning of the terms used in this model such as *class* and *part-of relation* seem evident for our purpose, the precise definitions of the terms are omitted. For the details of his model, readers should consult [115].

A class *representation* is composed of two parts, form and content (as shown below).

```
-- Representation
   p/o "form":Representational form
   p/o "content":Proposition
```

where p/o stands for *part-of* relation/slot, "form" is a slot name and ":Representational form" is a class constraint the slot value has to satisfy. Its identity is inherited from the *form*, which is usually what people sense its existence. On the other hand, the *content* is the hidden part and it is a proposition that the author of the representation would like to convey through the representation.

Representational form has the conceptual structure shown below.

```
-- Representational form
  - Symbol sequence
    - Natural language
      - Spoken language
      - Written language
    - Artificial language
      - Musical symbol
      - Mathematical symbol, etc.
  - Image
  - Speech
```

Proposition is a content of a statement of the result of one's recognition of every phenomenon that occurs in the real world. *Events* and *Facts* are *Proposition* and they are refined as follows:

```
-- Proposition
  - Design Proposition
    - Procedure
    - Music
    - Symbol
    - Specification
  - Product Proposition
    - Novel
    - Poem
    - Painting
```

It is important to clearly distinguish between a representation and a represented thing. Any representation is not embodied or realized unless it becomes a represented thing.

```
-- Represented thing
  - p/o "representation":Representation
  - p/o "medium":Representational medium
```

```
-- Representation medium
  - Physical thing
    - paper
    - canvas
  - Electronic thing
    - CD-ROM
    - etc.
```

Figure 5.1 depicts an example of a model of a representation—the letter "a"—according to this ontology. (Excerpt with minor modifications from Figure 1 of [115])
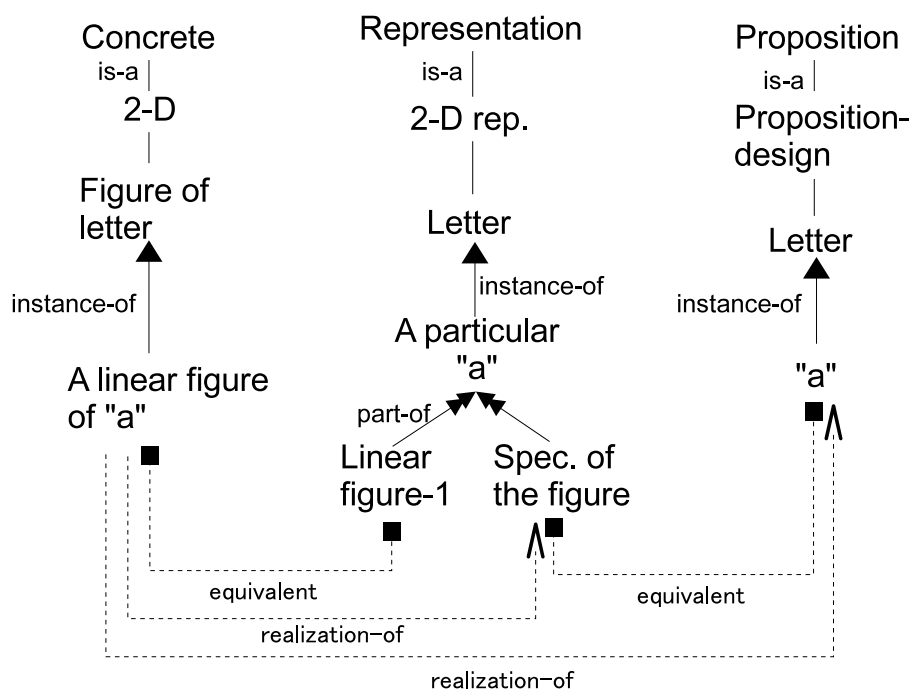
FIGURE 5.1: Ontology of "a".

Since this ontology is mainly for engineering and is constructed from the standpoint that is rather too practical for our purpose, some modifications are required.

For the present study, the focal point of this ontology is that a representation is in between the Concrete, Represented thing and the Proposition, as depicted in Figure 5.1. From the standpoint of this chapter, the Concrete, Represented thing in this ontology does not exist independently of the IGUS, and cannot be dealt by the IGUS directly. It is not even concrete in some sense. We use the word *substratum* instead of this *Represented thing* in the term of this ontology. The word

substratum is used to imply the *vague, indeterminate* material of which something (= *representation*) is made and from which it derives its special qualities. A substratum might be a thing, but in general, it is a more vague and indeterminate "event", *i.e.*, a bunch of physical stimuli before being sensed by a certain IGUS.

The Propositions are what resides in an IGUS. In this chapter, they are considered to be representations of the substrata as a part of the IGUS, or the substrata (in the IGUS) themselves. They are also not concrete against the impression of the word "proposition".

Thus, from the standpoint of our study, the representation connects vague constituents. To be more precisely, as described next in Section 5.2.2, the action to represent can be considered to be an action to yield some stableness from the vague and indeterminate constituents of the representation. The Represented thing and the Proposition in the term of this ontology are, as it were, the result of a projection of this stable representation onto each of the original substrata.

5.2.2   Closure

The importance of the concept of *closure* is repeatedly emphasized in many fields such as the systems theory, cybernetics, and complex systems theories. (See [36, 40], for example.)

Mathematical definition of the closure is as follows:

Definition 5.1. *Closure* is an operation C on sets, $C : A \to A^*$, with the following properties:

1. $A \subseteq A^*$ (monotonicity)

2. $(A^*)^* = A^*$ (idempotence)

3. $A \subseteq B \Longrightarrow A^* \subseteq B^*$ (inclusion preservation)

A set A is called *closed* if $A^* = A$. Intuitively, such a closure of a set means that somehow "missing elements" are added to it, until no more of them are needed.

In this chapter, important is the case in which a closure is formed reciprocally, *i.e.*, adjunctively by two sets. We formalize this using the Chu space [110]. (For

the interpretation of Chu space in this subsection, the author has referred to the description in [83].)

Definition 5.2. A *Chu space* (X, S, R) consists of two sets X, S and a relation R between them. If an element x of X has the relation R with an element s of S, we write

$$x \vDash s. \tag{5.1}$$

Suppose that an IGUS has a Chu space within it. We assume that every element of X represents the IGUS's experience and every element of S represents the nature or property of the experience. $x \vDash s$ is interpreted as representing the relation that a certain experience x has a certain property s. Then we can have the set of all the properties the IGUS remembers when it encounters a particular experience x:

$$S_x = \{s \in S : x \vDash s\} \tag{5.2}$$

and the set of all the common properties for a collection Y of some experiences:

$$S_Y = \{s \in S : \forall x \in Y,\ x \vDash s\}. \tag{5.3}$$

Similarly, for all experiences that have a particular property s in common, we can have the set of all such experiences as

$$X_s = \{x \in X : x \vDash s\}. \tag{5.4}$$

For all the experiences that have all its properties in a collection T of some properties, we can also have the set of all such experiences as

$$X_T = \{x \in X : \forall s \in T,\ x \vDash s\}. \tag{5.5}$$

Using the above sets, define two operations:

$$\mu : PX \to PS, X \supset Y \longmapsto S_Y \tag{5.6}$$

$$\gamma : PS \to PX, S \supset T \longmapsto X_T \tag{5.7}$$

and two composites:

$$Y^* := \gamma\mu(Y) \tag{5.8}$$

$$T^* := \mu\gamma(T) \tag{5.9}$$

where PS and PX are the power set of S and X, respectively. Operations $\gamma\mu$ and $\mu\gamma$ are closure operators. If $Y = Y^*$, then Y is said to be *closed*. Similarly, if $T = T^*$, then T is also said to be closed. The closed experience set Y can be regarded as a stable articulation of experiences, and the closed property set can be regarded as stable composite idea.

All the closed experience sets and all the closed property sets both have a structure called the complete lattice, and determine each other reciprocally (Galois correspondence). Some researchers call the pair of the closed experience set and the closed property set that determine each other in this structure as a *formal concept* [109].

Such a closure formed by two sets plays a significant role in our toy model. Representations can be considered to be, in a sense, a closure of its two constituents. This construction with two sets enables dynamic yielding of views and representations. However, let us start with the introduction of an algebraic structure, namely, *torsor*.

5.3   Torsor

In abstract algebra, there exists a suitable structure to represent the above mentioned facets of representation, that is, a *torsor* [2]. As we will see in the following sections, torsor enables us to avoid directly mentioning the represented thing. This feature is desirable for a fundamental model of information that this thesis explores. For a brief introduction including some history of torsor, the readers should refer to the appendix A of [84] and references cited therein.

Definition 5.3. A *torsor* $(G, (\cdot, \cdot, \cdot))$ is a set G together with a ternary operation $G \times G \times G \to G; (x, y, z) \mapsto [xyz]$ satisfying the identities: for all $x, y, z, u, v \in G$

$$[xy\,[zuv]] = [[xyz]\,uv] \qquad \text{(para–associative law)} \qquad (5.10)$$

$$[xxy] = y = [yxx] \qquad \text{(identity law)} \qquad (5.11)$$

The set of isomorphisms between two isomorphic objects naturally forms a torsor, with the operation $[fgh] = fg^{-1}h$ (here juxtaposition denotes composition of functions). This torsor becomes a group once a particular isomorphism by which the two objects are to be identified is chosen.

Torsors are everywhere [85]. Here are some examples of the torsor.

Groups Any group becomes a torsor under the operation: $[xyz] = xy^{-1}z$ [86].

Integers If x, y, z are integers, we can set $[xyz] = x - y + z$ to produce a torsor.

Miscellaneous Energies, voltages, indefinite integrals, and many others are torsors (see [85]).

In this chapter, the only derived property of the torsor to which we refer is the following.

Proposition 5.4. $[xy[zuv]] = [x[uzy]v]$.

---

[2]There are several equivalent versions, going under different names such as *heap, groud, flock, herd, principal homogeneous space, abstract coset, pregroup* and others. We use what seems to be the most suitable term, namely, torsor (the word *torsor* evidently seems to be originated from *torso*—something that is left unfinished).

5.4   Toy Model – Interpretation of Torsor

In this section, a toy model to describe a certain aspect of representations is built. More specifically, we build a model which describes some kinds of consistency of representations. The toy model is built by fixing the interpretation of the structure of a torsor into the words describing the target representation phenomenon. Because the structures of representation phenomena are not fully understood yet, the interpretation determining the toy model is defined in a rather vague manner, that is, by using the undefined but suggestive words. Precise interpretation must be given formally or informally in each study of a specific phenomenon. Some such examples are provided in the next section.

5.4.1   Interpretations of the Elements and Operation

Let us interpret each element of a torsor H as a *representation*, and the ternary operation $[xyz] \in H$ as *"the representation of a substratum which yields the representation z in the view of y, and the representation $[xyz]$ is obtained in the view of x,"* where the word *substratum* is used here to express *the represented thing* somewhat abstractly. Though this interpretation might seem to introduce two sorts, *view* and *representation*, it is not the case. The word *view* is simply a nickname of representation to suggest its role. It is used for understandability, and all the elements of a torsor are uniformly interpreted as *representations of physical situation in the real physical world*.

Dictionary definition of *view* is a mode or manner of looking at or regarding something. But, as can be seen later on, more active picture may be appropriate in this chapter. A view actually *yields* a representation. The operational meaning of the word *yield* here, and a mechanism of this yield will be described in Section 5.4.4.

Note that there are two choices of standpoint when using this toy model, namely, the local standpoint and the transcendental standpoint. In principle, all the phenomena are local, closely associated with each, particular IGUS. Hence all the elements of the torsor implicitly have an index to represent the IGUS. This is on local standpoints. However, one (as an IGUS) can set all the views by one's own view. This standpoint viewing the overall situation is called the transcendental standpoint. Appropriate standpoint depends on the case.

Also note that though the word *substratum* is used explicitly in above mentioned and following interpretations, the corresponding term does not exist in the formulae of torsor. This is one of the reasons to use the torsor. Because we usually regard representation as "re"-presentation of represented something, the descriptions of the interpretation in this section follow such a custom. However, the fact is, from the standpoint of this chapter, that the represented, *i.e.*, the substratum, which is the fundamental source of a representation, cannot be directly referred to. An IGUS must refer to the substratum only via its representation. Thus for the IGUS, the substratum is an unspecifiable existence. By using torsors, we can argue the matters of representations without directly referring to *a raw event*, "Ding an sich," or a substratum which yields the representations.

5.4.2   Interpretation of the Identity Law

Recall that the identity law of the torsor is:

$$[aax] = x = [xaa], \quad \forall a, x \in H \tag{5.12}$$

The first equation in the Equation (5.12), $[aax] = x$, can be interpreted to mean that *the representation of a substratum that yields the representation x in the view of a is x, and the representation $[aax]$ is obtained in the view of a*. To be short, *the same view yields the same representation*.

The second equation in the Equation (5.12), $[xaa] = x$, can be interpreted to mean that *the representation of a substratum which yields the representation a in the view of a is x, and the representation $[xaa]$ is obtained in the view of x*. This means that *there is (at least) a substratum that yields the representation itself for all representations (or views)*.

These interpretations of the identity law seem quite natural, as required properties of the representation phenomena.

5.4.3   Interpretation of the Para-associative Law

Recall that the para-associative law of the torsor is:

$$[[abc]de] = [ab[cde]], \quad \forall a, b, c, d, e \in H. \tag{5.13}$$

The verbatim interpretation of the requirement for the para-associative law is somewhat messy. This is also one of the reasons for using the language of algebra. Instead of analyzing in the words, we will use diagrams.

To begin with, the situation of $[abc]$ is depicted as Figure 5.2, where b and a are views and X represents a substratum, *i.e.*, the represented, that yields c and $[abc]$. Comparing with the definition of the interpretation of the ternary operation $[abc]$: *the representation of a substratum which yields the representation c in the view of b, and the representation $[abc]$ is obtained in the view of a*, the meaning of the diagram seems clear. The diagram corresponds to the interpretation. A formula form is also used with the figure form interchangeably. For the situation of Figure 5.2, we also write

$$X \xrightarrow{b} c \tag{5.14}$$

$$X \xrightarrow{a} [abc]. \tag{5.15}$$
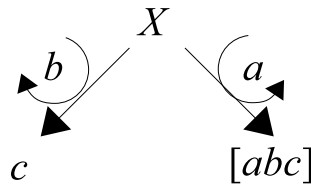
FIGURE 5.2: Ternary operation of torsor.

Using this diagram, the situation of the para-associative law can be depicted as in Figure 5.3. where X_1 and X_2 are the substrata again, and Y is also the substratum. X_1 and X_2 yield representations, but Y yields views. The view c is Y in the view of b, and the view $[abc]$ is Y in the view of a. The point here is the representation $[cde]$ is yielded in the view of c and in the view of b simultaneously. Starting from that point, we will trace how this diagram is built. From the transcendental standpoint, it is easy to understand.
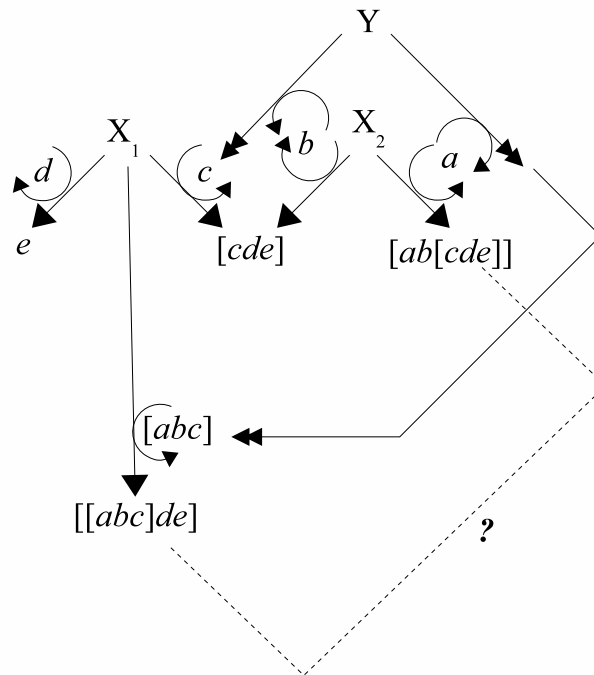
FIGURE 5.3: Para-associative law.

In our interpretation, for the existence of $[ab[cde]]$, there must exist a substratum X_2 that yields both of $[ab[cde]]$ and $[cde]$.

$$X_2 \overset{a}{\to} [ab[cde]] \tag{5.16}$$

$$X_2 \overset{b}{\to} [cde]. \tag{5.17}$$

But $[cde]$ itself is what is define by

$$X_1 \xrightarrow{d} e \tag{5.18}$$

$$X_1 \xrightarrow{c} [cde] \tag{5.19}$$

thus $[cde]$ can be yielded by the substratum X_1 other than X_2. So, more precisely, the situation is: for given $(a,)b, c, d, e \in H$,

$$X_1 \xrightarrow{c} [cde] \tag{5.20}$$

$$X_2 \xrightarrow{b} x \tag{5.21}$$

and

$$x = [cde]. \tag{5.22}$$

Thus there must be a restriction for c and b to make the Equation (5.22) hold.

The coexistence of the views c and b requires the existence of the substratum Y, which yields c in the view of b, because (since there is no other assumption on the existence of the view that may yield the view c and b) *the coexistence of c and b* means that the representation c can be yielded by some substratum (Y) in the view of b (and some other substratum yields b in the view c). With the existence of the substratum Y, $[abc]$ can be yielded in the view of a.

Then, in the view of $[abc]$, we can have $[[abc]de]$ from the substratum X_1. The diagram is completed.

Thus the para-associative law can be considered as a requirement for some kind of consistency among representations.

Naturally enough, it is not evident whether the reality conforms the para-associative law. Whether the law holds or not may depend on the case. When the law does not hold, it may be an indication that the representations lack some consistencies. In that case, a representation used by an IGUS may be incomprehensible by another IGUS. Investigation of such a matter on this toy model is expected to reveal some essential properties of the representation.

5.4.4   Role of the Closure

To make the above mentioned interpretations attainable, the concept of closure described in Section 5.2.2 plays a significant role in this toy model. It provides a mechanism to yield a new representation from a pair of representations or a pair consisting of a substratum and a representation. Especially, it enables yielding views without a meta-representation hierarchy.

As demonstrated previously in Section 5.4.3, in our interpretation, we must have some mechanism to make the following hold

$$X_1 \xrightarrow{c} [cde] \tag{5.23}$$

$$X_2 \xrightarrow{b} [cde] \tag{5.24}$$

$$Y \xrightarrow{b} c. \tag{5.25}$$

The view b yields both a representation $[cde]$ and a view c. Though the substrata used are different, they are constituents only used to explain the interpretation and not the elements in the torsor, *i.e.*, the toy model. Our basic assumption is that all the constituents of the model are representation. Hence we must have a mechanism that can be understood in the language of representation.

Closure reviewed in Section 5.2.2 can provide such a mechanism. Recall that, as reviewed in Section 5.2.1, any representation has two constituents. In accordance with this, a representation is considered to be a closure originated from two substrata, each of which vague and indeterminate. By forming the representation, these substrata become indirectly accessible for an IGUS by referring to the representation.

This two constituent construction enables representations to yield another representation *uniformly*, that is, irrespective of view or representation. It is done by the process like the genetic recombination: one of the constituents separates from the other constituents, then, it links with another substratum (or a constituent of another representation), and finally, this new pair yields another stable representation by forming a closure. A view, namely, a representation which reflects something internal to an IGUS, is considered to be yielded by such a *view*ing process.

Thus, we can regard the interpretation of the toy model as valid by making the two assumptions:

- the representation as a closure formed by two constituents

- the above mentioned yielding process

behind the interpretation of the toy model. They are, as it were, an "*internal degrees of freedom*" of the toy model. In addition, they must be specified, formally or informally, when the model is used.

5.4.5   Summary: How to Use the Toy Model

Following all the construction of the toy model mentioned above, we summarize here the steps to use it.

1. Identify representations to examine, and assume that they satisfy the axioms of the torsor;

2. identify the two constituents of the representations which form a closure;

3. fix the meaning of the equality in the algebra;

4. compare algebraic consequences with the reality of the phenomena.

The step three requires some notes. The equality in the torsor, such as the Equation (5.22), may be often too strong as a condition for the relation of representations, if it is interpreted literally as identity. In many cases, weaker condition (for instance, some equivalence relation) may be more appropriate. Thus we should select and fix the appropriate equality relation.

5.5 Examples

In this section, we exemplify the toy model by three elementary but suggestive cases.

5.5.1 Substitution Cipher-Group Operation

The substitution cipher is exemplified as the first case. Because substitutions or permutations form a group (an algebraic structure in mathematics), as already mentioned, it is evident that they also form a torsor by defining $[xyz] = xy^{-1}z$ [86]. However, this case is explicitly written down here to clarify what we do.

Suppose that the set H is a collection of all the character strings, each of which consisting of different letters in L. For instance, suppose that the set $L = \{1, 2, 3, 4, 5\}$, then $H = \{12345, 12354, \ldots, 54321\}$, which has $5! = 120$ elements. A substitution cipher, $i.e.$, permutation, is often represented by Cauchy's two-line notation, one lists the elements of L in the first row, and for each one its image under the permutation below it in the second row. For instance, a particular permutation of the set $\{1, 2, 3, 4, 5\}$ can be written as: $\sigma = \binom{12345}{32514}$. Suppose that we fix the first row of this notation as 12345. Then one can represent a particular permutation only by a string, e.g., 32514 for above σ. In other words, we consider H is isomorphic to the permutation S_5 of five elements.

Let us verify the para-associative law by a particular example. Suppose that $b = 13524, c = 52413, d = 32541, e = 14523$. We write $a = a_1a_2a_3a_4a_5$. As for $[cde]$, because e is yielded in the view of d, the substratum (which can be explicitly represented in this case) is $\left(\begin{smallmatrix}12345\\32541\end{smallmatrix}\right)^{-1}14523$. Then

$$
\begin{aligned}
[cde] &= \begin{pmatrix}12345\\52413\end{pmatrix}\begin{pmatrix}12345\\32541\end{pmatrix}^{-1}14523\\
&= \begin{pmatrix}12345\\52413\end{pmatrix}\begin{pmatrix}12345\\52143\end{pmatrix}14523\\
&= \begin{pmatrix}12345\\52413\end{pmatrix}54321\\
&= 31425
\end{aligned}
\tag{5.26}
$$

then,

$$
\begin{aligned}
[ab[cde]] &= \begin{pmatrix}12345\\a_1a_2a_3a_4a_5\end{pmatrix}\begin{pmatrix}12345\\13524\end{pmatrix}^{-1}31425\\
&= \begin{pmatrix}12345\\a_1a_2a_3a_4a_5\end{pmatrix}\begin{pmatrix}12345\\14253\end{pmatrix}31425\\
&= \begin{pmatrix}12345\\a_1a_2a_3a_4a_5\end{pmatrix}21543\\
&= a_2a_1a_5a_4a_3.
\end{aligned}
\tag{5.27}
$$

In the same manner,

$$
\begin{aligned}
[abc] &= \begin{pmatrix}12345\\a_1a_2a_3a_4a_5\end{pmatrix}\begin{pmatrix}12345\\13524\end{pmatrix}^{-1}52413\\
&= a_3a_4a_5a_1a_2
\end{aligned}
\tag{5.28}
$$

then,

$$
\begin{aligned}
[[abc]de] &= \begin{pmatrix}12345\\a_3a_4a_5a_1a_2\end{pmatrix}\begin{pmatrix}12345\\32541\end{pmatrix}^{-1}14523\\
&= a_2a_1a_5a_4a_3.
\end{aligned}
\tag{5.29}
$$

Therefore, $[ab[cde]] = [[abc]de]$, as expected. Thus the axiom of torsor holds for this substitution cipher (and substitution ciphers in general).

The property of the torsor, Proposition 5.4, $[ab[cde]] = [a[dcb]e]$ holds naturally as follows:

$$
\begin{aligned}
[dcb] &= \begin{pmatrix} 12345 \\ 32541 \end{pmatrix} \begin{pmatrix} 12345 \\ 52413 \end{pmatrix}^{-1} \begin{pmatrix} 12345 \\ 13524 \end{pmatrix} \\
&= \begin{pmatrix} 12345 \\ 41325 \end{pmatrix}
\end{aligned}
\tag{5.30}
$$

hence,

$$
\begin{aligned}
[a[dcb]e] &= \begin{pmatrix} 12345 \\ a_1 a_2 a_3 a_4 a_5 \end{pmatrix} \begin{pmatrix} 12345 \\ 41325 \end{pmatrix}^{-1} 14523 \\
&= a_2 a_1 a_5 a_4 a_3 \\
&= [ab[cde]].
\end{aligned}
\tag{5.31}
$$

5.5.2 Multiple Views for the Real World

To investigate formally, we assume the situation modeled by a Chu space (Section 5.2.2), namely, the formal context in the FCA [109]. A formal concept in FCA links a set of objects and a set of attributes. It can be regarded as a representation in which the objects and attributes correspond to the Representational form and Proposition, respectively.

In the situation where there are multiple IGUSes, their stable articulations of experiences and stable composite ideas, namely, formal concepts, naturally differ for every IGUS. In other words, all IGUSes have different views for the world. These multiple views for the real world may yield different formal concepts and their (inclusion) structures. But this incommensurability is overcome to some extent in the actual world. For instance, if two of the structure of formal concepts yielded from different views are isomorphic (though the isomorphism is rather strong condition), one can say that these two views are *compatible*. The representations based on the compatible views are also compatible, and the compatible representation is comprehensible by each IGUS not having the original view but having a compatible view.

5.5.2.1 Known Facts on FCA

First, some related concepts and results on FCA [109] are summarized.

Definition 5.5. A *formal context* $\mathbb{K} = (G, M, I)$ consists of two sets G and M and a relation I between G and M. The elements of G are called the *formal objects* and the element of M are called the *formal attributes* of the context. When an object g is in the relation I with an attribute m, we write $g \vDash m$ or $(g, m) \in I$ and read it as "the object g *has* the attribute m."

In the above and following definitions, the word "*formal*" is often omitted where it is obvious.

Definition 5.6. For a set $A \subseteq G$ of objects we define

$$\mu(A) := \{m \in M | g \vDash m \text{ for all } g \in A\} \tag{5.32}$$

(the set of attributes common to the objects in A). Correspondingly, for a set $B \subseteq M$ of attributes we define

$$\gamma(B) := \{g \in G | g \vDash m \text{ for all } m \in B\} \tag{5.33}$$

(the set of objects which have all attributes in B).

Definition 5.7. A *formal concept* of the context (G, M, I) is a pair (A, B) with $A \subseteq G, B \subseteq M, \mu(A) = B$ and $\gamma(B) = A$. We call A the *extent* and B the *intent* of the concept (A, B). $\mathfrak{B}(G, M, I)$ denotes the set of all concepts of the context (G, M, I).

The two operators μ and γ form a Galois connection between the power-set lattices of G and M. Hence we obtain two closure systems on G and M, which are dually isomorphic to each other.

We consider a part of the whole context.

Definition 5.8. If (G, M, I) is a formal context and if $H \subseteq G$ and $N \subseteq M$, then $(H, N, I \cap H \times N)$ is called a *subcontext* of (G, M, I).

Subcontexts that have some *compatibility* are important.

Definition 5.9. A subcontext $(H, N, I \cap H \times N)$ is called *compatible* if the pair $(A \cap H, B \cap N)$ is a concept of the subcontext for every concept $(A, B) \in \mathfrak{B}(G, M, I)$.

In the case of finite contexts, the compatible subcontexts can be identified by means of the arrow-relations.

Definition 5.10. If (G, M, I) is a context, $g \in G$ an object, and $m \in M$ an attribute, we write

$$g \nearrow m \quad :\Leftrightarrow \quad \begin{cases} g \not\models m \text{ and} \\ \text{if } \gamma(\{m\}) \subseteq \gamma(\{n\}) \text{ and } \gamma(\{m\}) \neq \gamma(\{n\}), \text{ then } g \models n \end{cases} \quad (5.34)$$

$$g \swarrow m \quad :\Leftrightarrow \quad \begin{cases} g \not\models m \text{ and} \\ \text{if } \mu(\{g\}) \subseteq \mu(\{h\}) \text{ and } \mu(\{g\}) \neq \mu(\{h\}), \text{ then } h \models m \end{cases} \quad (5.35)$$

Thus, $g \swarrow m$ holds if and only if g does not have the attribute m, but m is contained in the intent of every proper subconcept of $(\gamma\mu(\{g\}), \mu(\{g\}))$.

Definition 5.11. A subcontext $(H, N, I \cap H \times N)$ of a clarified context (G, M, I) is *arrow-closed* if the following holds: $h \nearrow m$ and $h \in H$ together imply $m \in N$, and $g \swarrow n$ and $n \in N$ together imply $g \in H$.

The next proposition holds.

Proposition 5.12. *Every compatible subcontext is arrow-closed. Every arrow-closed subcontext of a finite context is compatible.*

5.5.2.2 A Model

Using known facts mentioned above, a multiple views situation is modeled by the toy model.

In this model, we interpret the equality in the formulae of the torsor as the compatibility of the formal concept structure, *i.e.*, the compatibility of corresponding formal context. This interpretation is legitimate because if the contexts behind the formal concept structure are compatible, then the structures are isomorphic as for the common concepts. Thus IGUSes which have different views can agree on the identification of concepts.

For a given formal context $\mathcal{C} = (G, M, I)$, an operation $\widetilde{\alpha_G} : G \rightarrow G \times M$ is defined as extracting the minimum arrow-closed subcontext containing any specified subset H_0 of the formal object G. Similarly, an operation $\widetilde{\alpha_M} : M \rightarrow G \times M$ is defined as extracting the minimum arrow-closed subcontext containing any specified subset N_0 of the formal attribute M. These operations, combined with projection, are defined for later convenience:

$$\alpha_G^1 := \pi_1 \circ \widetilde{\alpha_G} : G \rightarrow G, \ \alpha_G^2 := \pi_2 \circ \widetilde{\alpha_G} : G \rightarrow M \tag{5.36}$$

$$\alpha_M^1 := \pi_1 \circ \widetilde{\alpha_M} : M \rightarrow G, \ \alpha_M^2 := \pi_2 \circ \widetilde{\alpha_M} : M \rightarrow M \tag{5.37}$$

The view π_{H_0} generated with any specified subset H_0 of the formal object G is defined as

$$\pi_{H_0}(\mathcal{C}) = (H, N, I \cap H \times N) = (\alpha_G^1(H_0), \alpha_G^2(H_0), I \cap \alpha_G^1(H_0) \times \alpha_G^2(H_0)) \tag{5.38}$$

$$i.e., \quad \pi_{H_0} = (- \cap \alpha_G^1(H_0), - \cap \alpha_G^2(H_0), I \cap \alpha_G^1(H_0) \times \alpha_G^2(H_0)) \tag{5.39}$$

Similarly, the view ρ_{N_0} generated with any specified subset N_0 of the formal attribute M is defined as

$$\rho_{N_0}(\mathcal{C}) = (H, N, I \cap H \times N) = (\alpha_M^1(N_0), \alpha_M^2(N_0), I \cap \alpha_M^1(N_0) \times \alpha_M^2(N_0)) \tag{5.40}$$

$$i.e., \quad \rho_{N_0} = (- \cap \alpha_M^1(N_0), - \cap \alpha_M^2(N_0), I \cap \alpha_M^1(N_0) \times \alpha_M^2(N_0)) \tag{5.41}$$

They satisfies the following formulas: $\pi_H \circ \pi_H = \pi_H, \rho_N \circ \rho_N = \rho_N$, hence each of them is a kind of projection.

Each element of the model is associated with any subset H_0 of G or N_0 of M. But as a view, it works as π_{H_0} or ρ_{N_0}, that is, it induces an arrow-closed subcontext including the associated subset.

We can have canonical correspondence between the view defined above and the formal context (G, M, I) and the set of formal concepts $\mathfrak{B}(G, M, I)$. Hence by corresponding a set of formal concepts with representation, we can handle the view and the representation uniformly. Formal contexts brought by these views

are arrow-closed subcontexts, hence they are comptible with the original context $\mathcal{C} = (G, M, I)$. On these assumptions, the axioms of the torsor are automatically satisfied.

5.5.2.3   A Concrete Example

Suppose that we have a context $\mathcal{C} = (G, M, I)$, where $G = \{g_1, g_2, g_3, g_4\}$, $M = \{m_1, m_2, m_3, m_4, m_5\}$ and I is defined by the Table 5.1 (excerpted from [109]).

TABLE 5.1: Global context.

	m_1	m_2	m_3	m_4	m_5
g_1	X		X		
g_2	X	X	X		X
g_3		X			X
g_4			X	X	X

The Table 5.1 shows $g_1 \vDash m_1, g_1 \nvDash m_2, g_1 \vDash m_3$, and so on.

We adopt the transcendental view for simplicity. In other words, the context $\mathcal{C}$ is assumed to be global.

The arrow-closed subcontext can be easily obtained from the graph of arrows (Figure 5.4) drawn by Definition 5.10. For example, if we start with a subset $\{m_2\}$, by chasing the links and collecting all the reachable elements, we have

$$(\{g_1, g_3, g_4\}, \{m_1, m_2, m_3, m_5\}, I \cap \{g_1, g_3, g_4\} \times \{m_1, m_2, m_3, m_5\}) \tag{5.42}$$

as the arrow-closed subcontext. Note that in the Figure 5.4, the links m_2 to g_1 and m_1 to g_3 are unidirectional, and the others are bidirectional. Thus if we start with a subset $\{g_1\}$, the arrow-closed subset induced by the subset is $(\{g_1\}, \{m_5\}, I \cap \{g_1\} \times \{m_5\})$.
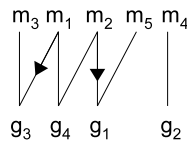

FIGURE 5.4: The graph of arrows.

Suppose that each subset of G or M is assigned to an element of the torsor as in Table 5.2.

TABLE 5.2: Model assignment.

Element	a	b	c	d	e
H_0	$\{g_4\}$	$\{g_1, g_2, g_3\}$	$-$	$\{g_1, g_2\}$	
N_0	$-$	$-$	$\{m_3, m_5\}$	$-$	
H	$\{g_1, g_3, g_4\}$	$\{g_1, g_2, g_3\}$	$\{g_1, g_3\}$	$\{g_1, g_2\}$	Obtained from d
N	$\{m_1, m_2, m_3, m_5\}$	$\{m_3, m_4, m_5\}$	$\{m_3, m_5\}$	$\{m_4, m_5\}$	Obtained from d

Then, what is the interpretation of $[cde]$? Since the context $\mathcal{C}$ is assumed to be global, it can be regarded as the substratum for all representations. Thus e is obtained by applying $\rho_{\{g_1, g_2\}}$ to $\mathcal{C}$. As a result, e is a formal concept lattice of subcontext:

$$(\{g_1, g_2\}, \{m_4, m_5\}, I \cap \{g_1, g_2\} \times \{m_4, m_5\}) \tag{5.43}$$

(It is a chain of $\{g_1, g_2\}$ and $\{g_2\}$ and $\emptyset$. See Figure 5.5 d.) Recall that there is a canonical correspondence between the view and the formal context and the set of formal concepts.

In the same way, a formal concept lattice of subcontext in the view of c is:

$$(\{g_1, g_3\}, \{m_3, m_5\}, I \cap \{g_1, g_3\} \times \{m_3, m_5\}). \tag{5.44}$$

(It is an anti-chain of $\{g_1\}$ and $\{g_3\}$. See Figure 5.5 c.) It corresponds to $[cde]$.
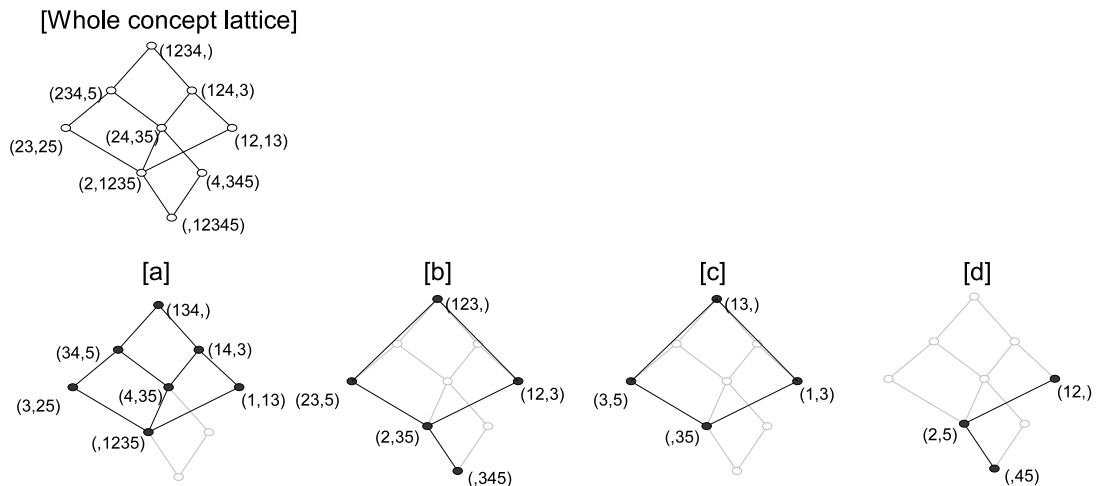

FIGURE 5.5: Concept lattices for full and subcontexts.

The context(= substratum) $\mathcal{C}$ in the view of b is also the formal concept lattice of subcontext:

$$(\{g_1, g_2, g_3\}, \{m_3, m_4, m_5\}, I \cap \{g_1, g_2, g_3\} \times \{m_3, m_4, m_5\}). \qquad (5.45)$$

(See Figure 5.5 b.) This lattice embeds the lattice of $[cde]$, thus "*modulo*" irrelevant elements, the Equation (5.22) holds.

The context in the view of a is the formal concept lattice of subcontext:

$$(\{g_1, g_3, g_4\}, \{m_1, m_2, m_3, m_5\}, I \cap \{g_1, g_3, g_4\} \times \{m_1, m_2, m_3, m_5\}). \qquad (5.46)$$

This is a seven element lattice of $\{\emptyset, \{g_1\}, \{g_3\}, \{g_4\}, \{g_1, g_4\}, \{g_3, g_4\}, \{g_1, g_3, g_4\}\}$ (see Figure 5.5a). This corresponds to $[ab[cde]]$.

Similarly, since $[abc]$ is yielded by the view a, the induced formal concept lattice $[abc]$ must be embedded into the lattice of a (Figure 5.5 a). Also, the formal concept lattice $[[abc]de]$ induced by $[abc]$ must be embedded into the lattice of $[a]$—that is, in the $[ab[cde]]$. Hence, the axiom $[ab[cde]] = [[abc]de]$ holds "*modulo*" irrelevant elements, as expected,

Note that although the context $\mathcal{C}$ is assumed to be global for simplicity in this section, this is not mandatory. As can be understood from above procedures, a context as a substratum can omit the irrelevant objects or attributes to the views applied to the substratum.

The property of the torsor, Proposition 5.4, $[a[dcb]e] = [ab[cde]]$, can verified in the same manner.

5.5.3 Letter Recognition

The use of the toy model is not necessarily restricted to mathematically concrete models. It can be used for more abstract situations.

Let us consider the episode described in 1.1.3. A child see a blot as the letter "pee", another child as "er" and the other child "rho". One part of the situation

can be described as *what is a letter "rho" in the view of G is regarded as "pee" in the view of E.* Thus, in the term of the toy model,

$$[EG\langle\langle\text{rho}\rangle\rangle] = \langle\langle\text{pee}\rangle\rangle \tag{5.47}$$

Another part of the situation can be described as *what is a letter "pee" in the view of E' is regarded as "er" in the view of R.* In the term of the toy model,

$$[RE'\langle\langle\text{pee}\rangle\rangle] = \langle\langle\text{er}\rangle\rangle \tag{5.48}$$

therefore, in combination with the previous equation,

$$[RE'[EG\langle\langle\text{rho}\rangle\rangle]] = \langle\langle\text{er}\rangle\rangle \tag{5.49}$$

In these equations, G is a *Greek letter-related* view, E, E' (not necessarily different) are *English or Latin letter-related* views, R is a *Cyrillic letter-related* view, and $\langle\langle\text{rho}\rangle\rangle, \langle\langle\text{pee}\rangle\rangle, \langle\langle\text{er}\rangle\rangle$ are (possibly, *utterance-related*) representations.

In the axiom of the toy model,

$$[RE'[EG\langle\langle\text{rho}\rangle\rangle]] = [[RE'E]G\langle\langle\text{rho}\rangle\rangle] \tag{5.50}$$

must be hold. Let us consider what it means in this situation.

The point is the term $[RE'E]$ in the right hand of the equation. It must be considered to be a view for *what is a letter "rho" in the view of G.* This term is interpreted as *the representation of a substratum that yields the representation E in the view of E', and the representation $[RE'E]$ is obtained in the view of R.* Extremely roughly speaking, since the view $[RE'E]$ is something obtained in the view of R, a *Cyrillic letter-related* view, $[RE'E]$ might be also a Cyrillic letter-related view, and hence $[[RE'E]G\langle\langle\text{rho}\rangle\rangle] = \langle\langle\text{er}\rangle\rangle$. But a closer examination is required.

In fact, the construction of the toy model provides an insight. In the construction, all representations and views consists of two constituents. One is concrete, belonging to the outer world, and the other is belonging to the inner world of the IGUS. The constituents of the views like G, E, E', R can be considered to be something like a space of topological configurations of the blots and something like a space of (to be) written languages. The space of topological configurations

of the blots is articulated enough to represent letters of a language, in accordance with an articulation of the space of differences required to represent utterances of a language. The latter spaces naturally vary from language to language, and both of the spaces vary from IGUS to IGUS. As can be known, the criteria for allowable forms of a certain letter in a certain language may slightly vary from person to person.

Thus the view $[RE'E]$ can be understood to be constructed as a closure of a space of topological configuration of blots and a space of (to be) written Russian language, and the former space also yields a view E together with a space originated in E'.

Needless to say, this is one of the possibilities. More studies in the linguistics and other related fields are required for clearer understanding. However, investigations using the toy model in this study must be helpful in such a case.

5.6 Conclusive remarks on this chapter

In this chapter, an algebraic toy model for representation is developed. By using torsor as the algebraic structure of the model, one can avoid directly mentioning the represented thing, which is unrealistic at a fundamental level of understanding.

This chapter presents how to interpret a representation phenomenon to the algebraic structure of the model. The correspondence relations of the elements, the two-sidedness of representation and the closure concept play significant roles in the construction of the toy model. Through the examples, it is shown that the axioms of torsor tend to hold and correspond to some consistency requirement for the representation.

The mathematical formulation of the toy model presented in this chapter is naturally not definitive. As is mentioned in Section 5.3, the torsorial concept has been formulated in various equivalent forms. In addition to them, even in math-

ematically more foundational levels, one may replace the set theory, which is the base of the formulation of the closure and torsorial concept in the toy model in the present chapter, with the language of category theory, the named sets, or the like[3]. Each of such formulations has its own value for a toy model of representation phenomena. The point of the present study is that the representation has a torsorial nature. Toy models in various formulations may also be used to describe phenomena of representations. There is much room to be investigated for formulating richer structures suitable for individual representation phenomenon to be modeled.

---

[3]An instance of categorically-oriented articles discussing torsors is [87]. Regarding the named sets, the readers should refer to the appendix A of [15] and references therein. Chu spaces, which are used in Section 5.2.2, also are described in the same appendix as a special case of named sets.

Chapter 6

Conclusion

6.1 Contributions to information sciences

An understanding of the fundamentals of information is explored in this study. In the view of this thesis, information is

- a physical phenomenon,

- universality extracted from the surrounding universe by canonical methods,

- relative to the IGUS, and

- contextual existence.

This study grounds information phenomena on physics, introduces an appropriate language for studying information, and approaches the essential relativity of information.

In this thesis, the author:

- defines information phenomena operationally by modeling the process of extensional information articulation;

- demonstrates the model in the case of attribute extraction from the universe, clarifying essential universality of attributes as information and canonicity of the extraction process;

- proposes the language of category theory as a description language of information phenomena;

- demonstrates the validity of the proposed language in the case of attribute extraction from the universe, and characterizes the attributes by an internal description; and,

- provides an understanding of *information dismantled its content* by using the algebra known as torsor.

To define *information in this thesis* operationally, a model of the process of extensional information articulation has been developed as a generalization of the theory of the extensive measurement in metrology. As a realization of it, the attribute extraction/creation from the universe has been investigated as a significant instance of the information phenomena. We have mathematically understood that, in the attribute extraction process, the attribute may be obtained in a boot–strapping manner by the duality of extensively built structures, and that the universality of the attribute as information may be realized as a limit. Galois connection, or the adjoint structure, that forms information relieves us from the vicious circularity hidden in the preceding understanding of information phenomena, mentioned in Section 1.2[1]. These fruits may be considered to be a validation of the model of the process of extensional information articulation.

---

[1]Most of the preceding understanding of information seem to place information to provide distinction as equal to information formed by the distinction. The author considers that it is the origin of the vicious circularity. The adjoint structure may be considered to be a relaxation of equality as is mentioned in Section 4.2.3.

The language of category theory may illuminate these and some of the fundamental features of information. It has been demonstrated that the characteristics of category theory are well suited to describe the information phenomena. Its structural relativity and contextuality, in particular, may enable to reflect those of information naturally. The validity of using the language of category theory has been demonstrated in an investigation of an information phenomenon, namely, the attribute extraction for classification. By using the language of category theory, especially topos, for modeling the attribute extraction, a new, internal description is obtained and this description provides a new characterization of the attribute as information: "an attribute exists where there is a lack of available contextual resource to specify a certain element in the context."

The torsor, a kind of algebraic structure, sheds a new light on the *information dismantled its context.* The relativity and contextuality hitherto studied should be understood in connection with the significance of information, that is, *the difference made by difference.* This is one side of information phenomena. There is the other side of information phenomena: *the difference which makes difference.* As the author has shown in this thesis, information, which is constructed with both these differences, is heavily made up with their contexts. Some essence of information is behind this cosmetics. This thesis provides an answer to the problem of what information really is like when it is dismantled of its context. An answer in the view of this thesis is given from the algebraic perspective. That is the torsor. Difference which makes a difference may be described adequately in the language of torsor. Though this is not a completely novel view to see the world as Baez [85] indicates, the torsorial model presented in this thesis provides a new perspective on information phenomena.

Information as fully extensional physical phenomena has not been well investigated in preceding studies, despite of its fundamentality and inevitability. The present thesis tackles this problem squarely, and has shown that there may be an extensional understanding of the fundamentals of information, by demonstrating a concrete mechanism. This extensional understanding suggests that the relativity and contextuality of information are inherent because, extensionally, differences in physical interactions are the source of information, and the differences are considered to be necessarily represented by the relations between somethings extensional. This relations constitute a certain structure of extension, and extension characterizes each other through the structure. Thus, information may not escape from the surrounding circumstances, and be inherently relative and contextual.

Through the extensional understanding, this thesis qualitatively extends the prior studies of information mentioned in Chapter 2. This might not be a full extension of those prior studies[2], but at least, the present study may stretch the limit of them. Many of the prior studies assume the distinction of the occurences of the units of information such as microstates, messages, symbols in a string, statements, *etc.* to calculate the "information". Each of them mainly focuses on only one part of the context for the information phenomena, that is, either the information carrier part or the received information part. This thesis considers both parts of the context simultaneously to avoid vicious circularity of the distinction and information. In the information process in our study, the distinction becomes clearer as the boundary of the distinguished objects become clearer[3]. Distinction is woven into the structure that the distinguished objects form, and the structure becomes the context of the information based on the distinction[4]. If one disregards this evolutional process, the distinction becomes God–given and the structure becomes flat, as in the prior studies. And the flat structure is a special case of the structure in this study. In this sense, the present study may be considered to be a qualitative extension of the prior works.

---

[2]That is, this study might not incorporate prior studies completely.

[3]This might sound tautologous to some readers. Note that the author associates the distinction with the attributes that characterizes the object under consideration.

[4]And vice versa. The objects is woven into the structure that the distinction forms.

The quantitative outcome is out of the scope of this thesis, it might be promising. Valuations in a structure, in a finite distributive lattice which appears in Chapter 3 at least, have been studied out of different motives. For example, there is a mathematical theorem proven by Gian–Carlo Rota [116], which states that a valuation in a finite distributive lattice $\mathcal{L}$ is uniquely determined by the values it takes on the set of join–irreducibles of $\mathcal{L}$, and these values can be arbitrarily assigned. Knuth [117], for instance, utilizes this theorem to compute generalized entropies, which are generalizations of familiar information-theoretic quantities such as mutual information. In the same manner, information might be discussed quantitatively on the result of this thesis. In this aspect also, this thesis may be regarded as an extension of the prior information theories referred in Chapter 2.

6.2   Future prospects

This study has exploratory nature and the demonstrations of concrete information phenomena are limited. The models developed in this thesis are undoubtedly still in its infancy. Their virtue should be displayed and validated in further various applications. Some points that are not within the scope of this thesis remain to be addressed in future investigations.

The demonstrations such as in Chapter 3 and in Chapter 4 should be extended in future studies. Besides the attribute extraction, there might still be other information phenomena that may be well investigated in the similar view of this thesis. Besides the inclusion structure, there also might still be other structures that may induce the other type of information than the attribute. For instance of the latter, the network structure might be a good candidate.

The toy model developed in Chapter 5 is also doubtlessly far from reaching full maturity. The model is demonstrated by only a few selected cases, and the algebraic consequences of the model still has much opportunity to be widely explored. However, it seems that the toy model can be expected to capture some essential properties of the representation and have values for studying the representation phenomena. The value of the toy model depends on valuable applications for various studies. Although the application of the model might be restricted to theoretical settings, there will be many potential application fields of the toy model because it is an abstract, formal model of representations that makes few assumptions[5] about actual participants of the information phenomenon under consideration. Body languages or gestures often cause problems that originate from the two-sidedness of representation — one gesture may have multiple meanings that can be determined by the each witness. Cross-cultural/intercultural communications may suffer from similar problems. To study these problems, the toy model may be of help. The model may also be applicable to the study of information exchanges between animals, such as ants and bees, when the researcher tries to describe the researcher's—human's—view explicitly. Besides these fields in which representation problems come to the fore, studies of our daily information exchanges, such as information theory, linguistics, communication theories, studies of man-machine interfaces, *etc.*, may be able to utilize the toy model, when they deal with something semantical.

Explication of the construction of an IGUS is restricted in this thesis. Requirements for the IGUS have been clarified just for our purpose (3.3.3). The essential functions of the IGUS in more general setting should be more fully described in an appropriate abstraction level in future researches. The system identification of an IGUS is also beyond the reach of this study. In the view of this thesis, an IGUS should be considered as simply a collection of physical interactions. How can a part of the universe *recognize* another part of the universe as an IGUS? In this thesis, the author sometimes rely on the intuitive understanding of the human-IGUSes. After all, the language of physics is (basically) for human-IGUSes. However, the information theory, which is based on the physical interactions, *i.e.*, extensional entities, should be more fully developed without assuming human-IGUSes.

---

[5]Assumptions for the participants to be IGUSes.

In this extensional view, the distinction between information and the physical phenomena becomes further obscured. The distinctive characteristics of physical phenomena as information might be worth pursuing from the extensional information viewpoint. Superficially, the information phenomena might be associated with the existence of retarded physical interactions and a choice of reactions of the IGUS. Future work should explore the distinction, or, the difference of *the intensity of informationality*, of physical phenomena from this extensional information viewpoint.

There are many other approaches for investigating the problem of interrelation of physical and informational reality. The most contrasting approach to this study is the approach that tries to explain the reality with information only. As already mentioned in the Section 3.1, many researches adopt such an approach from various motives. It is true, even in the view of this study, that the physical reality for us may be considered to be constructed by *information for us humans*. But it seems also true that there exists something (*i.e.*, the universe) in which we do not have our own way in everything, something that is the never–exhausting source of information. Though it is not the case that the authors had fully evaluated every study adopting a purely informational approach, the physical approach (such as this study) and the informational approach seem to, so to speak, complement or adjoin each other. Future studies should address such matters to clarify the relationship between these two approaches.

In topos and category theory in general, there are many mathematical properties that might touch the hearts of the researchers of information. These properties are waiting for an opportunity to be examined. Though just one application exhibited in Chapter 4 might not be enough convincing, the language of category theory must be promising because of its natural correspondences to the properties of information, as shown in Chapter 4. Future studies should consider the use of the language of category theory, and also investigate further the properties of category theory that might contribute to the information study.

The models developed in this thesis may have its applications in the field of exploratory data analysis by their extensional nature. Immanent universal property, e.g., the s–attribute, of collected data can be represented structurally, *i.e.*, *extensionally* by the collected data themselves. Because this information extracted from the data can be manipulable through its extensionality, the models may be valuable in the field of exploratory data analysis.

The understanding of information in this study might be a matter of interest of the researchers in other disciplines such as linguistics, communication theory, robotics, and many others. With the characteristics of information understood in this thesis, such as the essential contextuality and the bootstrapping capturing process *at the information level*, they might have the opportunity to reconsider their understanding of the phenomena in their field and may reach deeper understanding in their own discipline.

The author expects the understanding in this thesis to be developed further, and the models to be applied in theory and in practice, to fully understand the extensional aspects of fundamentals of information.

Bibliography

Papers constituting this dissertaion

[1] Makoto Yoshitake and Yasufumi Saruwatari,
"Extensional information articulation from the universe," *Information*,
vol. 3, no. 4, pp. 644–660, 2012.

[2] Makoto Yoshitake, "An internal description of an origin of the attribute."
In preparation.

[3] Makoto Yoshitake, "A toy model for torsorial nature of representations,"
Information, vol. 3, no. 4, pp. 546–566, 2012.

General references on information

Concept of information

[4] William Aspray, "The scientific conceptualization of information: a survey,"
IEEE Annals of the History of Computing, vol. 7, no. 2, pp. 117–140, 1985.

[5] John D. Collier, "Intrinsic Information," in *Information, Language and Cognition*, vol. 1, pp. 390–409, Oxford University Press, 1990.

[6] Peter J. Denning, "Can there be a science of information?," *ACM Computing
Surveys*, vol. 27, no. 1, pp. 23–25, 1995.

[7] Rafael Capurro and Birger Hjorland, "The Concept of Information," in *Annual Review of Information Science and Technology*, vol. 37, pp. 343–411,
American Society for Information Science and Technology, 2003.

[8] Olimpia Lombardi, "What is information?," *Foundations of Science*, vol. 9,
pp. 105–134, June 2004.

[9] Manuel Bremer and Daniel Cohnitz, *Information and Information Flow*, Ontos Verlag, 2004.

[10] Scott J. Muller, *Asymmetry : The Foundation of Information*, Springer Berlin Heidelberg, 2007.

[11] Luciano Floridi, "A defence of informational structural realism," *Synthese*, vol. 161, pp. 219–253, 2008.

[12] Luciano Floridi, "Modern Trends in Philosophy of Information," in *Philosophy of Infomation*, pp. 113–132. Handbook of the Philosophy of Science, Elsevier, 2008.

[13] Luciano Floridi, "Against digital ontology," *Synthese*, vol. 168, no. 1, pp. 151–178, 2009.

[14] Luciano Floridi, *INFORMATION – A Very Short Introduction*, Oxford University Press, 2010.

[15] Mark Burgin, *Theory of Information – Fundamentality, Diversity and Unification*, World Scientific: Singapore, 2010.

Thermodynamics

[16] Nicolas L. S. Carnot, *Reflection on the Motive Power of Fire*, Dover, 1990.

[17] Rudolf J. E. Clausius, *The Mechanical Theory of Heat*, translated by Walter R. Browne, MacMillan and Co.: London, 1879.

[18] Ludwig Boltzmann, "Über die beziehung zwischen dem zweiten hauptsatz der mechanischen wärmetheorie und der wahrscheinlichkeitsrechnung respektive den sätzen über des wärmegleichgewicht," *Wiener Berichte*, vol. 76, pp. 373–435, 1877.

[19] Leó Szilárd, "Über die entropieverminderung in einem thermodynamischen system bei eingriffen intelligenter wesen," *Zeitschrift für Physik*, vol. 53, pp. 840–856, 1929.

[20] Enrico Fermi, *Thermodynamics*, Dover Publications, 1937.

[21] Léon N. Brillouin, *Science and Information Theory*, Academic Press, 1957.

[22] Herbert B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, John Wiley & Sons, 2nd ed., 1985.

[23] Leff S. Harvey and Andrew F. Rex, eds., *Maxwell's Demon 2 – Entropy, Classical and Quantum Information, Computing*, Institute of Physics Publishing, 2003.

Mathematical theory of communication

[24] Claude E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[25] Claude E. Shannon and Warren Weaver, *Mathematical Theory of Communication*, University of Illinois Press, 1963.

[26] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2nd ed., 2006.

Algorithmic information theory

[27] Ray J. Solomonoff, "A preliminary report on a general theory of inductive inference," Technical Report ZTB-138, Zator Co., Cambridge, Mass., 1960.

[28] Ray J. Solomonoff, "A formal theory of inductive inference, part i," *Information and Control*, vol. 7, pp. 1–22, 1964.

[29] Ray J. Solomonoff, "A formal theory of inductive inference, part ii," *Information and Control*, vol. 7, pp. 224–254, 1964.

[30] Andrey N. Kolmogorov, "Logical basis for information theory and probability theory," *IEEE Transactions on Information Theory*, vol. 14, pp. 662–664, 1968.

[31] Gregory J. Chaitin, "On the length of programs for computing finite binary sequences: statistical considerations," *Journal of the ACM*, vol. 16, pp. 145–159, 1969.

[32] Gregory J. Chaitin, "A theory of program size formally identical to information theory," *Journal of the ACM*, vol. 22, no. 3, pp. 329–340, 1975.

[33] Richard von Mises, *Probability, Statistics and Truth*, translated by Jerzy Neyman, et al., Macmillan, New York, 1939.

[34] Richard von Mises, *Probability, Statistics and Truth*, George Allen & Unwin Ltd., 3rd ed., 1957. 2nd revised English edition prepared by Hilda Geiringer, Dover, 1981.

Cybernetics

[35] Norbert Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, Paris, Hermann et Cie – MIT Press, Cambridge, MA., 1948.

[36] W. Ross Ashby, *AN INTRODUCTION TO CYBERNETICS*, CHAPMAN & HALL LTD: London, UK, 1957.

[37] Gregory Bateson, *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*, University Of Chicago Press: Chicago, IL, USA, reprinted ed., 1972,1999.

[38] Stewart Brand, "For God's Sake, Margaret – Conversation with Gregory Bateson and Margaret Mead," in *CoEvolution Quarterly*, Issue no. 10, pp. 32-44., 1976. Available as http://www.oikos.org/forgod.htm (accessed September 16, 2013).

[39] Humberto Maturana and Francisco Varela, *Autopoiesis and Cognition: The Realization of the Living*, Reidel, 1980.

[40] Francis Heylighen, "Self-organization, emergence and the architecture of complexity," *Proceedings of the 1st European Conference on System Science, (AFCET, Paris)*, pp. 23–32, 1989.

[41] Francis Heylighen, *Representation and Change. A Metarepresentational Framework for the Foundations of Physical and Cognitive Science*, Communication & Cognition: Ghent, Belgium, 1990.

[42] John Mingers, *Self-Producing Systems: Implications and Applications of Autopoiesis*, Springer, 1995.

[43] Francis Heylighen and Cliff Joslyn, "Cybernetics and Second-Order Cybernetics," in *Encyclopedia of Physical Science & Technology*, vol. 4, pp. 155–169, Academic Press, New York, 3rd ed., 2001.

[44] Heinz von Foerster, *Understanding Understanding: Essays on Cybernetics and Cognition*, Springer, 2003

[45] Pier Luigi Luisi, "Autopoiesis: a review and a reappraisal," *Naturwissenschaften*, vol. 90, pp. 49–59, 2003.

[46] Toru Nishigaki, *Fundamental Informatics — from life to society*, NTT Publishing, Co.,Ltd., 2004. (In Japanese.)

[47] Toru Nishigaki, *Fundamental Informatics II — for Vital Organization*, NTT Publishing, Co.,Ltd., 2008. (In Japanese.)

[48] Bruce Clarke and Mark B.N. Hansen, eds., *Emergence and Embodiment — New Essays on Second–Order Systems Theory*, Duke University Press, 2009.

[49] Kazuya Yamashita, *Autopoiesis*, Minerva-shobou, 2010. (In Japanese.)

[50] Pablo Razeto-Barry, "Autopoiesis 40 years Later. A Review and a Reformulation," *Origins of Life and Evolution of Biospheres*, vol. 42, pp. 543–567, 2012.

Semantic theory of information

[51] Rudolf Carnap and Yehoshua Bar-Hillel, "An outline of a theory of semantic information," *MIT Reseach Laboratory of Electronics Technical Report*, no. 247, 1952.

[52] Yehoshua Bar-Hillel and Rudolf Carnap, "Semantic information," *The British Journal for the Philosophy of Science*, vol. 4, no. 14, pp. 147–157, 1953.

[53] Fred Dretske, *Knowledge and the Flow of Information*, The MIT Press, 1981.

[54] Kenneth J. Barwise and John Perry, *Situations and Attitudes*, MIT Press, 1983.

[55] Kenneth J. Barwise and Jerry Seligman, *Information Flow*, Cambridge University Press, 1997.

[56] John Perry, "Situation Semantics," in *Routledge Encyclopedia of Philosophy*, ed. Edward Craig, Routledge, 1998.

[57] Kozo Sugiyama, et al., ed., *Knowledge Science*, Kinokuniya Shoten, 2002. (In Japanese.) Available as http://www.kousakusha.com/ks/ (accessed May 26, 2008).

[58] Thomas J. Hickey, *History of Twentieth-Century Philosophy of Science*, 2005.
Available as http://www.philsci.com/ (accessed April 28, 2008).

Linguistics, Semiology, Semiotics

[59] Ferdinand de Saussure, *Saussure's Third Course of Lectures in General Linguistics (1910 – 1911)*, Oxford: Pergamon, 1993.

[60] Umberto Eco, *A Theory of Semiotics*, Indiana University Press, 1976.

[61] John L. Austin, *How to Do Things with Words*, Harvard University Press, 2nd ed., 1975. Japanese translation: *Gengo–to–Koui*, TAISHUKAN Publishing Co., Ltd., 1978.

[62] John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, 1983. Japanese traslation: *Shikousei*, Seishinshobo Publishers Co.Ltd, 1997.

[63] Ronald W. Langacker, *Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites*, Stanford University Press: Palo Alto, CA, USA, 1987.

[64] Herbert P. Grice, *Studies in the Way of Words*, Harvard University Press, 1989. Japanese translation: *Ronri–to–Kaiwa*, KEISO–SHOBO, 1998.

[65] John R. Searle, *The Rediscovery of the Mind*, Bradford Books, 1992.

[66] Dan Sperber and Deirdre Wilson, *Relevance. Communication and Cognition*, Blackwell, 2nd ed., 1995.

[67] John R. Taylor, *Linguistic Categorization: Prototypes in Linguistic Theory*, Oxford University Press: Oxford, UK, 2nd ed., 1995.

[68] Daniel Chandler, *SEMIOTICS: THE BASICS*, Routledge, 2nd ed., 2007.

Mathematics used in this thesis

[69] Garrett Birkhoff, *Lattice Theory*, AMS Colloquium Publications Vol.25, American Mathematical Society: Providence, RI, USA, 3rd ed., 1967.

[70] Brian A. Davey and Hilary Priestley, *Introduction to Lattices and Order*, Cambridge University Press: Cambridge, UK, 2002.

[71] Izumi Ojima and Kazuya Okamura, *Mugen Ryoushikei no Butsuri to Suuri (Physics and Mathematics of Infinite Quantum Systems)*, Saiensu-sha, 2013. (In Japanese.)

[72] Harold Simmons, *An Introduction to Category Theory*, Cambridge University Press, 2011.

[73] Steve Awodey, *Category Theory*, Oxford University Press, 2006.

[74] Saunders MacLane, *Categories for the Working Mathematician*, Springer, 1971.

[75] Robert Goldblatt, *Topoi: The Categorial Analysis of Logic*, Dover Publications, 1983.

[76] John L. Bell, *Toposes and Local Set Theories: An Introduction*, Dover Publications, 1988.

[77] Saunders MacLane and Ieke Moerdijk, *Sheaves in Geometry and Logic – A First Introduction to Topos Theory*, Springer, 2nd ed., 1994.

[78] Francis Borceux, *Handbook of Categorical Algebra 1*, Cambridge University Press, 1994.

[79] Ronald Brown and Tim Porter, "Category Theory: an abstract setting for analogy and comparison," in *What is Category Theory?*, pp. 257–274. Polimetrica s.a.s.: Milano, Italy, 2006.

[80] Steve Awodey, Carsten Buz, Alex Simpson, and Thomas Streicher, "Relating topos theory and set theory via categories of classes," Technical Report CMU–PHIL–146, Carnegie Mellon: Pittsburgh, Pennsylvania, 2003.

[81] John Baez, "Topos theory in a nutshell." Last modified April 2006, `http://math.ucr.edu/home/baez/topos.html`, (accessed June 04, 2013).

[82] George Grätzer, *General Lattice Theory*, Birkhäuser, 2003.

[83] Toru Tsujishita, "Life and compex systems," in *Mathematics*, pp. 75–225. SEIDOSHA Inc.: Tokyo, Japan, 1998. (In Japanese.)

[84] Wolfgang Bertram and Michael Kinyon, "Associative geometries i: Torsors, linear relations and grassmannians," *Journal of Lie Theory*, vol. 20, no. 2, pp. 215–252, 2010.

[85] John Baez, "Torsors made easy." Lat modified December 2009, `http://math.ucr.edu/home/baez/torsors.html`, (accessed June 22, 2012).

[86] Jeremiah Certaine, "The ternary operation $(abc) = ab^{-1}c$ of a group," *Bulletin of the American Mathematical Society*, vol. 49, no. 12, pp. 869–877, 1943.

[87] Thomas Booker and Ross Street, "Torsors, herds and flocks," *Journal of Algebra*, vol. 330, pp. 346–374, 2011.

[88] Mathematical Society of Japan and Kiyoshi Ito, eds., *Encyclopedic Dictionary of Mathematics*, The MIT Press, 2nd ed., 1987.

[89] Nicolas Bourbaki, *Elements of mathematics: Theory of Sets*, Springer, 1968.

[90] "nLab," `http://ncatlab.org/nlab/show/HomePage`, (accessed July 16, 2013).

Miscellaneous

[91] Michael J. Reddy, "The conduit metaphor: A case of frame conflict in our language about language," in *Metaphor and Thought*, pp. 284–310, Cambridge: Cambridge University Press, 1979.

[92] Brent Berlin and Paul Kay, *Basic Color Terms: Their Universality and Evolution*, CSLI publications: Stanford, CA, USA, 1999.

[93] Dan C. Marinescu, *Classical and Quantum Information*, Elsevier Inc.: Amsterdam, Netherlands, 2011.

[94] Wojciech H. Zurek, "Algorithmic Information Content, Church Turing Thesis, Physical Entropy and Maxwell's Demon," in *Complexity, Entropy and the Physics of Information*, ed. Wojciech Zurek, pp. 73–89, Addison-Wesley: Redwood City, CA, USA, 1990.

[95] Murray Gell-Mann and James B. Hartle, "Quantum mechanics in the light of quantum cosmology," in *Complexity, Entropy and the Physics of Information*, ed. Wojciech Zurek, pp. 425–458, Addison-Wesley: Redwood City, CA, USA, 1990.

[96] Konrad Zuse, *Rechnender Raum*, vol. 1, Friedrich Vieweg & Sohn, Braunschweig, 1969. English translation: *Calculating Space*, MIT Technical Translation AZT–70–164–GEMIT, Massachusetts Institute of Technology: Cambridge, MA, 1970.

[97] Seth Lloyd, *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*, Random House: New York, NY, USA, 2006.

[98] Carl F. von Weizsäcker, *Aufbau der Physik*, Hanser, Munich, Germany, 1985.

[99] John A. Wheeler, "Information, physics, quantum: The search for links," in *Complexity, Entropy and the Physics of Information*, ed. Wojciech H. Zurek, pp. 3–28, Addison-Wesley: Redwood City, CA, USA, 1990.

[100] Roy B. Frieden, *Physics from Fisher Information*, Cambridge University Press, Cambridge, 1998.

[101] Andrei Khrennikov, "Classical and quantum mechanics on information spaces with applications to cognitive, psychological, social, and anomalous phenomena," *Foundations of Physics*, vol. 29, pp. 1065–1098, 1999.

[102] David H. Krantz, R. Duncan Luce, Amos Tversky, and Patrick Suppes, *Foundations of Measurement Volume I: Additive and Polynomial Representations*, Dover: Mineola, NY, USA, 2007. Reprint of the Academic Press, New York and London, 1971 edition.

[103] Herman von Helmholtz, "Zählen und messen, erkenntnistheoretisch betrachtet," in *Philosophische Aufsätze, Eduard Zeller zu seinem fünfzigjährigen Doctorjubiläum gewidmet.*, pp. 17–52, Fues' Verlag: Leipzig, Germany, 1887. English translation: *Counting and Measuring*, translated by C.L.Bryan, Van Nostrand: Princeton, NJ, USA, 1930.

[104] Otto L. Hölder, "Die axiome der quintität und die lehre vom mass," *Mathematisch–Physikalische Klasse*, vol. 53, pp. 1–64, 1901.

[105] Ryozo Yoshino, Naohito Chino, and Kimihiko Yamagishi, *Suurishinrigaku (Mathematical Psychology)*, Baifukan Co., Ltd.: Tokyo, Japan, 2007. (In Japanese.)

[106] Davis Baird, *Thing Knowledge: A Philosophy of Scientific Instruments*, University of California Press: Berkeley, CA, USA, 2003. Japanese translation: *Mono no katachi wo shita chishiki*, translated by Shunsuke Matsuura, SEIDOSHA Inc.: Tokyo, Japan, 2005.

[107] Stevan Harnad, "To cognize is to categorize: Cognition is categorization," in *Handbook of categorization in cognitive science*, ed. Henri Cohen and Claire Lefebvre, pp. 19–43, Elsevier Inc.: Amsterdam, Netherlands, 2005.

[108] Joseph Rosen, *Symmetry in science: an introduction to the general theory*, Springer-Verlag New York, Inc.: New York, NY, USA, 1995.

[109] Bernhard Ganter and Rudolf Wille, *Formal Concept Analysis : Mathematical Foundations*, translated by Cornelia Franzke, Springer-Verlag Berlin Heidelberg: Berlin, Germany, 1999.

[110] Vaughan Pratt, "Chu Spaces," Technical Report, Stanford University, 1999. Available as http://boole.stanford.edu/pub/coimbra.pdf, (accessed November 26, 2007).

[111] Rafael D. Sorkin, "Finitary substitute for continuous topology," *International Journal of Theoretical Physics*, vol. 30, no. 7, pp. 923–947, 1991.

[112] Samson Abramsky and Nikos Tzevelekos, "Introduction to Categories and Categorical Logic," in *New Structures for Physics*, pp. 3–94, Springer-Verlag: Berlin Heiderberg, 2011.

[113] Fotini Markopoulou, "The internal description of a causal set," *Communications in Mathematical Physics*, vol. 211, pp. 559–583, 2000.

[114] Niklas Luhmann, *Soziale Systeme: Grundlis einer allgemeinen Theorie*, Suhrkamp Verlag KG, 1984.
English translation: *Social systems*, translated by John Bednarz, Jr. with Dirk Baecker, Stanford University Press, Stanford, CA, 1995.

[115] Riichiro Mizoguchi, "Tutorial on ontological engineering – part 3: Advanced course of ontological engineering," *New Generation Computing*, vol. 22, no. 2, pp. 193 – 220, 2004.

[116] Gian-Carlo Rota, "On the combinatorics of the Euler characteristic," in *Studies in pure mathematics: papers in combinatorial theory, analysis, geometry, algebra, and the theory of numbers presented to Richard Rado on the occasion of his sixty-fifth birthday*, Academic Press: London, 1971.

[117] Kevin H. Knuth, "Lattice duality: The origin of probability and entropy," *Neurocomputing*, vol. 67, pp. 245–274, 2005.

Appendix A

Preliminaries and detailed data

A.1 Mathematical Preliminaries for Chapter 3

For details of the terms used here, readers should consult books like [88] and references listed there. Standard books on the order theory such as [70] and [82], and books on the set theory such as [89] may also be of help.

Definition A.1 (Partially Ordered Set). [70] Let P be a set. A *partial order* on P is a binary relation $\leqslant$ on P such that, for all $x, y, z \in P$,

$$
\begin{align}
&(1)\ x \leqslant x, \tag{A.1} \\
&(2)\ x \leqslant y \text{ and } y \leqslant x \text{ imply } x = y, \\
&(3)\ x \leqslant y \text{ and } y \leqslant z \text{ imply } x \leqslant z,
\end{align}
$$

A set P equipped with an order relation $\leqslant$ is said to be a *partially ordered set.*

Strict inequality $<$ is also introduced. $x < y$ in P if and only if $x \leqslant y$ and $x \neq y$. When the order should be explicitly presented, a partially ordered set is denoted as $(P, \leqslant)$.

Definition A.2 (Dual of an Ordered Set). [70] Given any ordered set P we can form a new ordered set P^{∂} (the dual of P) by defining $x \leqslant y$ to hold in P^{∂} if and only if $y \leqslant x$ holds in P.

To each statement about the ordered set P there corresponds a statement about P^{∂}. In general, given any statement Φ about ordered sets, we obtain the *dual*

159

statement Φ^∂ by replacing each occurrence of $\leqslant$ by $\geqslant$ and vice versa. It is known that if a statement Φ about ordered sets is true in all ordered sets, then the dual statement Φ^∂ is also true in all ordered sets.

Definition A.3 (Covering Relation). [70] Let P be an ordered set and let $x, y \in P$. We say x is covered by y (or y covers x) if $x < y$ and $x \leqslant z < y$ implies $z = x$.

Definition A.4 (Down-set). [70] Let P be a partially ordered set and $Q \subseteq P$. Q is said to be a *down-set* or an *order ideal* if, whenever $x \in Q, y \in P$ and $y \leqslant x$, we have $y \in Q$.

Definition A.5 (Upper/Lower Bounds). [70] Let P be an ordered set and let $S \subseteq P$. An element $x \in P$ is an *upper bound* of S if $s \leqslant x$ for all $s \in S$. A *lower bound* is defined dually.

If all upper bound has a least element x, then x is called the *least upper bound* of S. The *greatest lower bound* is defined dually. The least upper bound of $\{x, y\}$ is denoted as $x \vee y$ (read as 'x join y' and the greatest lower bound of $\{x, y\}$ is denoted as $x \wedge y$ (read as 'x meet y'. Similarly for the set S, we write $\bigvee S$ for the join of S, $\bigwedge S$ for the meet of S.

Definition A.6 (Lattice). [70] Let P be a non-empty ordered. If $x \vee y$ and $x \wedge y$ exist for all $x, y \in P$, then P is called a *lattice*.

Definition A.7 (Meet Irreducible Element). [82] An element a of an partially ordered set is called *meet-irreducible*, if $a = b \wedge c$ implies that $a = b$ or $a = c$; it is called *join-irreducible*, if $a = b \vee c$ implies that $a = b$ or $a = c$.

Definition A.8 (Directed set). [88] An ordered set (or in general a preordered set) in which every finite subset is bounded from above is called a *directed set*.

Definition A.9 (Projective System). [88] Let I be a directed set. Suppose that we are given a set X_i for each $i \in I$ and a mapping $\psi_{ij} : X_j \to X_i$ for each $i \leqslant j$, such that $\psi_{ii} = 1_{X_i}$ and $\psi_{ik} = \psi_{ij} \circ \psi_{jk} \, (i \leqslant j \leqslant k)$. Then we denote the system by (X_i, ψ_{ij}) and call it a *projective system* (or *inverse system*) of sets over I. Let P be the subset of the Cartesian product ΠX_i defined by $P = \{(x_i) : \psi_{ij}(x_j) = x_i \, (i \leqslant j)\}$, and let $p_i : P \to X_i$ be the canonical mappings. Then we have P(1) $\psi_{ij} \circ p_j = p_i \, (i \leqslant j)$; P(2) for any set X, and for any system of mappings $q_i : X \to X_i$ satisfying $\psi_{ij} \circ q_j = q_i \, (i \leqslant j)$, there exists a unique mapping $p : X \to P$ such that $p_i \circ p = q_i \, (i \in I)$. We call (P, p_i) the *projective limit* (or *inverse limit*) of the projective system (X_i, ψ_{ij}) over I and denote it by $\varprojlim X_i$.

A.2 Mathematical Preliminaries for Chapter 4

Readers should consult [72–74, 78, 90], and other textbooks for detailed definitions. As for topos, [75–77]. The descriptions in this section are greatly indebted to these references.

In this section, the basic definitions appeared in Chapter 4 are reproduced for the readers' convenience, and the definitions of notions that are mentioned but not strictly defined in the text are also included for completeness.

A.2.1 Definitions appeared in Chapter 4

Definition A.10. A category $\mathcal{C}$ consists of a family of *objects* $Ob(\mathcal{C})$ and for each pair of objects, A, B in $Ob(\mathcal{C})$, a set $\mathcal{C}(A, B)$ of *arrows* or *morphisms* from A to B, together with a way of composing arrows that match

$$\circ : \mathcal{C}(A, B) \times \mathcal{C}(B, C) \to \mathcal{C}(A, C).$$

The object A of an arrow in $\mathcal{C}(A, B)$ is called the *domain* of the arrow, and the object B is called the *codomain* of the arrow. This data is to satisfy:

1. composition is associative. If $h \circ (g \circ f)$ is defined, $(h \circ g) \circ f$ as well and $h \circ (g \circ f) = (h \circ g) \circ f$.

2. there are identities for each object, so there are *identity* arrows id_A in $\mathcal{C}(A, A)$ for A in $Ob(\mathcal{C})$, such that if $f \in \mathcal{C}(A, B)$,

$$\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A.$$

(An arrow is also written as $f : A \to B$ if $f \in \mathcal{C}(A, B)$, and the composite of the arrows $f \in \mathcal{C}(A, B)$ and $g \in \mathcal{C}(B, C)$ is written as $g \circ f$ or simply gf.)

An "arrow between categories" is called a *functor*.

Definition A.11. A *functor* $F : \mathcal{C} \to \mathcal{D}$ is given by:

- An object–map, assigning an object FA of $\mathcal{D}$ to every object A of $\mathcal{C}$.

- An arrow–map, assigning an arrow $Ff : FA \to FB$ of $\mathcal{D}$ to every arrow $f : A \to B$ of $\mathcal{C}$, in such a way that composition and identities are preserved:

$$F(g \circ f) = Fg \circ Ff, \ F\mathrm{id}_A = \mathrm{id}_{FA}.$$

An "arrow between functors" is called a *natural transformation*.

Definition A.12. Let $F, G : \mathcal{C} \to \mathcal{D}$ be functors. A *natural transformation*

$$t : F \to G$$

is a family of morphisms in $\mathcal{D}$ indexed by objects A of $\mathcal{C}$,

$$\{t_A : FA \to GA\}_{A \in Ob(\mathcal{C})}$$

such that, for all $f : A \to B$, the following diagram commutes:

$$
\begin{array}{ccc}
FA & \xrightarrow{Ff} & FB \\
t_A \downarrow & & \downarrow t_B \\
GA & \xrightarrow{Gf} & GB
\end{array}
\ .
$$

That is, $t_B \circ Ff = Gf \circ t_A$. This condition is known as *naturality*.

Definition A.13. An arrow $f : A \to B$ in a category $\mathcal{C}$ is called a *monic* when, for every object $C \in Ob(\mathcal{C})$ and every pair of arrows $g, h : C \rightrightarrows A$, the following property holds: $(f \circ g = f \circ h) \Longrightarrow (g = h)$.

In category theory, universal arrows are defined as follows.

Definition A.14. Let $G : \mathcal{D} \to \mathcal{C}$ be a functor, and C an object of $\mathcal{C}$. A *universal arrow from C to G* is a pair (D, η) where D is an object of $\mathcal{D}$ and an arrow $\eta : C \to G(D)$ of $\mathcal{C}$, such that, for every pair (D', f) with D' an object of $\mathcal{D}$ and an arrow $f : C \to G(D')$ of $\mathcal{C}$, *there exists a unique* arrow $\hat{f} : D \to D'$ of $\mathcal{D}$ such that $f = G(\hat{f}) \circ \eta$.

Diagrammatically:

$$
\begin{array}{ccc}
C \xrightarrow{\ \eta\ } G(D) & & D \ . \\
\ \searrow_{f} \quad \Big\downarrow G(\hat{f}) & & \Big\downarrow \hat{f} \\
\qquad G(D') & & D'
\end{array}
$$

In other words, every arrow f to G factors uniquely through the universal arrow η.

Definition A.15. Let $\mathcal{C}, \mathcal{D}$ be categories. Given $\mathcal{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{D}$, where F and G are functors, an *adjunction* from $\mathcal{C}$ to $\mathcal{D}$ is given by a pair of natural transformations

$$
\eta \ : \ 1_{\mathcal{C}} \Rightarrow GF \qquad \text{(unit)}, \tag{A.2}
$$

$$
\epsilon \ : \ FG \Rightarrow 1_{\mathcal{D}} \qquad \text{(counit)}, \tag{A.3}
$$

satisfying the triangle identities axioms:

$$
\begin{array}{ccc}
F \xrightarrow{\ F\eta\ } FGF & & G \xrightarrow{\ \eta G\ } GFG \ . \\
\ \searrow_{1_F} \quad \Big\downarrow \varepsilon F & & \ \searrow_{1_G} \quad \Big\downarrow G\varepsilon \\
\qquad F & & \qquad G
\end{array} \tag{A.4}
$$

Adjunction may be defined differently.

Definition A.16. Let $\mathcal{C}, \mathcal{D}$ be categories. An *adjunction* from $\mathcal{C}$ to $\mathcal{D}$ is a triple (F, G, θ), where F and G are functors

$$
\mathcal{C} \underset{G}{\overset{F}{\rightleftarrows}} \mathcal{D} \tag{A.5}
$$

and θ is a family of bijections

$$
\theta_{C,D} : \mathcal{C}(C, G(D)) \overset{\equiv}{\rightarrow} \mathcal{D}(F(C), D), \tag{A.6}
$$

for each $C \in Ob(\mathcal{C})$ and $D \in Ob(\mathcal{D})$, *natural* in C and D.

Definition A.17. A *subobject* of an object X in a category $\mathcal{C}$ is a monic

$$m : M \rightarrowtail X.$$

Subobjects may be redefined to be an equivalence class of monics:

Definition A.18. A *subobject* of an object X in a category $\mathcal{C}$ is an entity

$$[m] = \{n : m \sim n\},$$

where $m : M \rightarrowtail X$ is a monic with codomain X.

Definition A.19. An object 1 is *terminal* in a category $\mathcal{C}$ if for every object $C \in Ob(\mathcal{C})$ there is one and only one arrow from C to 1 in $\mathcal{C}$.
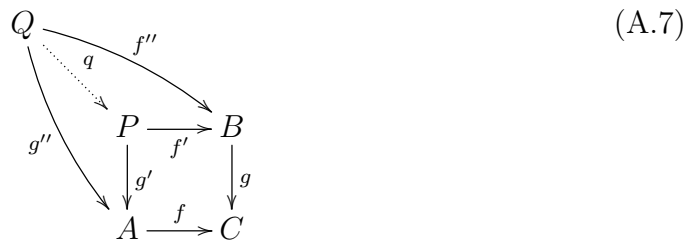
Definition A.20. Consider two morphisms $f \in \mathcal{C}(A, C)$, $g \in \mathcal{C}(B, C)$ in a category $\mathcal{C}$. A *pullback* of (f, g) is a triple (P, f', g') where

1. P is an object of $\mathcal{C}$.

2. $f' : P \to B$, $g' : P \to A$ are morphisms of $\mathcal{C}$ such that $f \circ g' = g \circ f'$,

and for every other triple (Q, f'', g'') where

1. Q is an object of $\mathcal{C}$.

2. $f'' : Q \to B$, $g'' : Q \to A$ are morphisms of $\mathcal{C}$ such that $f \circ g'' = g \circ f''$,

there exists a unique morphism $q : Q \to P$ such that $f'' = f' \circ q$ and $g'' = g' \circ q$. See the diagram A.7.

$$\begin{matrix} Q & & & (\text{A.7}) \\ & q \searrow \; \searrow f'' & & \\ g'' \downarrow & P \xrightarrow{f'} B & \\ & g' \downarrow \quad \downarrow g & \\ & A \xrightarrow{f} C & \end{matrix}$$

Definition A.21. Let $\mathcal{E}$ be a category with all finite limits. A *subobject classifier* in $\mathcal{X}$ consists of an object Ω together with an arrow $\top : 1 \to \Omega$ that is a "universal subobject", in the following sense: Given any object X and any subobject $M \rightarrowtail X$, there is a unique arrow $\chi(m) : X \to \Omega$ making the following diagram a pullback:

$$
\begin{array}{ccc}
M & \longrightarrow & 1 \\
{\scriptstyle m}\downarrow & & \downarrow {\scriptstyle t} \\
X & \xrightarrow{\ \chi(m)\ } & \Omega
\end{array}
\qquad (A.8)
$$

The arrow $\chi(m)$ is called the *characteristic arrow of the subobject* $m : M \rightarrowtail X$ (or *of* M).

Definition A.22. A *topos* is a category with:

a) finite limits and colimits,

b) exponentials,

c) a subobject classifier.

A.2.2 Notions not defined strictly in the text

The duality principle [78] ought to be mentioned:

Definition A.23. Given a category $\mathcal{C}$, the dual category $\mathcal{C}^*$ is defined in the following way:

1. $Ob(\mathcal{C}^*) = Ob(\mathcal{C})$;

2. for all objects C, D in $\mathcal{C}^*$, $\mathcal{C}^*(C, D) = \mathcal{C}(D, C)$;

3. the composition law of $\mathcal{C}^*$ is given by $f^* \circ g^* = (g \circ f)^*$.

In category theory, the following metatheorem – *duality principle* – holds: *Suppose the validity, in every category, of a statement expressing the existence of some objects or morphisms or the equality of some composites. Then the "dual statement" is also valid in every category; this dual statement is obtained by reversing the direction of every arrow and replacing every composite $f \circ g$ by the composite $g \circ f$.*

Definition A.24. An object 0 is initial in category $\mathcal{C}$ if for every $C \in Ob(\mathcal{C})$ there is one and only one arrow from 0 to C in $\mathcal{C}$.

The initial object is the dual of the terminal object.

Definition A.25. A *product* in a category $\mathcal{C}$ of two objects C and D is an object $C \times D \in Ob(\mathcal{C})$ together with a pair $(\pi_C \in \mathcal{C}(C \times D, C), \pi_D \in \mathcal{C}(C \times D, D))$ of arrows such that for any pair of arrows of the form $(f \in \mathcal{C}(X, C), g \in \mathcal{C}(X, D))$ there is exactly one arrow $\langle f, g \rangle \in \mathcal{C}(X, C \times D)$ making

$$
\begin{array}{c}
X \\
\swarrow^{f} \quad \downarrow^{\langle f,g \rangle} \quad \searrow^{g} \\
C \xleftarrow{\pi_C} C \times D \xrightarrow{\pi_D} D
\end{array}
\tag{A.9}
$$

commute, that is, such that $\pi_C \circ \langle f, g \rangle = f$ and $\pi_D \circ \langle f, g \rangle = g$. $\langle f, g \rangle$ is the product arrow of f and g with respect to the projections π_C, π_D.

Coproducts are defined dually.

Definition A.26. An arrow $i \in \mathcal{C}(E, C)$ in $\mathcal{C}$ is an *equalizer* of a pair $f, g \in \mathcal{C}(C, D)$ of arrows if

1. $f \circ i = g \circ i$, and

2. whenever $h \in \mathcal{C}(X, C)$ has $f \circ h = g \circ h$ in $\mathcal{C}$ there is exactly one arrow $k \in \mathcal{C}(X, E)$ such that $i \circ k = h$.

$$
\begin{array}{c}
E \xrightarrow{i} C \underset{g}{\overset{f}{\rightrightarrows}} D \\
\uparrow^{k} \quad \nearrow^{h} \\
X
\end{array}
\tag{A.10}
$$

Coequalizers are defined dually.

Initial/terminal object, products/coproducts, equalizers/coequalizers are special cases of the more general notions of *limit* and *colimit* of a diagram in a category.

Definition A.27. A *diagram* in a category $\mathcal{C}$ is a set of objects of $\mathcal{C}$ together with a set of arrows between these objects.

We have already drawn diagrams many times.

Definition A.28. Let D be a diagram with vertices $\{D_i : i \in I\}$ in a category $\mathcal{C}$. A *cone* over the diagram D is a family $\left\{ A \xrightarrow{f_i} D_i : i \in I \right\}$ of arrows from a fixed object A to the objects in D such that, for any arrow $D_i \xrightarrow{d} D_j$ in D, the diagram

$$
\begin{array}{ccc}
 & A & \\
f_i \swarrow & & \searrow f_j \\
D_i & \xrightarrow{\ \ d\ \ } & D_j
\end{array}
\qquad (A.11)
$$

commutes, *i.e.*, $d \circ f_i = f_j$. The object A is called the *vertex* of the cone.

Definition A.29. A *limit for the diagram* D is a cone $\left\{ A \xrightarrow{f_i} D_i : i \in I \right\}$ such that for all cones $\left\{ A' \xrightarrow{f'_i} D_i : i \in I \right\}$ there is a unique arrow $A' \xrightarrow{g} A$ for which the diagram

$$
\begin{array}{ccc}
A' & \xrightarrow{\ g\ } & A \\
 & f'_i \searrow & \downarrow f_i \\
 & & D_i
\end{array}
\qquad (A.12)
$$

commutes for each $i \in I$. That is, for each $i \in I$, $f'_i = f_i \circ g$.

Colimits are defined dually.

Definition A.30. Let C and D be objects of a category $\mathcal{C}$ such that all binary products with D exist. (Usually, $\mathcal{C}$ actually has all binary products.) Then an *exponential object* is an object C^D equipped with an *evaluation map* $ev \in \mathcal{C}(C^D \times D, C)$ which is universal in the sense that, given any object X and map $e \in \mathcal{C}(X \times D, C)$, there exists a unique map $u \in \mathcal{C}(X, C^D)$ such that

$$
\begin{array}{ccc}
C^D \times D & \xrightarrow{\ ev\ } & C \\
u \times id_D \uparrow & \nearrow e & \\
X \times D & &
\end{array}
\qquad (A.13)
$$

commutes, *i.e.*, $e = ev \circ (u \times id_D)$.

Definition A.31. A *frame* is a lattice L in which every (even infinite) subset $\{a_i\}$ of L has a supremum $\vee a_i$ such that

$$b \wedge \left(\bigvee a_i \right) = \bigvee (a_i \wedge b) \tag{A.14}$$

for all b in L.

Frames form a category with functions preserving meets and joins.

Definition A.32. The category **Frm** of *frames* is the category whose objects are complete lattices satisfying the infinite distributive law (Equation.A.14), and whose morphisms are functions preserving finite meets and arbitrary joins.

A.3 Concrete Example for Chapter 4

TABLE A.1: All the members in each concept

Concept	Member	Concept	Member
1	{Dog, Frog, Bream, Leech}	10	{Frog, Reed}
2	{Dog, Frog, Bream}	11	{Reed, Maize, Spike_weed, Bean}
3	{Dog, Frog}	12	{Reed, Maize, Spike_weed}
4	{Dog}	13	{Reed, Spike_weed}
5	{Dog, Frog, Reed, Maize, Bean}	14	{Reed, Maize, Bean}
6	{Frog, Bream, Leech}	15	{Reed, Maize}
7	{Frog, Bream}	16	{Reed}
8	{Frog}	17	{Bean}
9	{Frog, Bream, Leech, Reed, Spike_weed}		

TABLE A.2: Object part of the functor V

p	$V(p)$	p	$V(p)$
1	$\{1, 2, 3, 4, 5, 6, 7, 8\}$	10	$\{8, 10, 16\}$
2	$\{2, 3, 4, 7, 8\}$	11	$\{11, 12, 13, 14, 15, 16, 17\}$
3	$\{3, 4, 8\}$	12	$\{12, 13, 15, 16\}$
4	$\{4\}$	13	$\{13, 16\}$
5	$\{3, 4, 5, 8, 10, 14, 15, 16, 17\}$	14	$\{14, 15, 16, 17\}$
6	$\{6, 7, 8\}$	15	$\{15, 16\}$
7	$\{7, 8\}$	16	$\{16\}$
8	$\{8\}$	17	$\{17\}$
9	$\{6, 7, 8, 9, 10, 13, 16\}$		

Table A.3: Sieves

$R(1)$	$\{1 \to 1\}$	$R(10)$	$\{10 \to 10, 10 \to 9, 10 \to 5\}$
$R(2)$	$\{2 \to 2, 2 \to 1\}$	$S5(10)$	$\{10 \to 5\}$
$S1(2)$	$\{2 \to 1\}$	$S9(10)$	$\{10 \to 9\}$
$R(3)$	$\{3 \to 3, 3 \to 2, 3 \to 5, 3 \to 1\}$	$R(11)$	$\{11 \to 11\}$
$S1(3)$	$\{3 \to 1\}$	$R(12)$	$\{12 \to 12, 12 \to 11\}$
$S2(3)$	$\{3 \to 2, 3 \to 1\}$	$S11(12)$	$\{12 \to 11\}$
$S5(3)$	$\{3 \to 5\}$	$R(13)$	$\{13 \to 13, 13 \to 9, 13 \to 12, 13 \to 11\}$
$R(4)$	$\{4 \to 4, 4 \to 3, 4 \to 2, 4 \to 5,$	$S9(13)$	$\{13 \to 9\}$
	$\quad 4 \to 1\}$		
$S1(4)$	$\{4 \to 1\}$	$S11(13)$	$\{13 \to 11\}$
$S2(4)$	$\{4 \to 2, 4 \to 1\}$	$S12(13)$	$\{13 \to 12, 13 \to 11\}$
$S3(4)$	$\{4 \to 3, 4 \to 2, 4 \to 1, 4 \to 5\}$	$R(14)$	$\{14 \to 14, 14 \to 5, 14 \to 11\}$
$S5(4)$	$\{4 \to 5\}$	$S5(14)$	$\{14 \to 5\}$
$R(5)$	$\{5 \to 5\}$	$S11(14)$	$\{14 \to 11\}$
$R(6)$	$\{6 \to 6, 6 \to 1, 6 \to 9\}$	$R(15)$	$\{15 \to 15, 15 \to 12, 15 \to 14,$
			$\quad 15 \to 5, 15 \to 11\}$
$S1(6)$	$\{6 \to 1\}$	$S5(15)$	$\{15 \to 5\}$
$S9(6)$	$\{6 \to 9\}$	$S11(15)$	$\{15 \to 11\}$
$R(7)$	$\{7 \to 7, 7 \to 6, 7 \to 1, 7 \to 9,$	$S12(15)$	$\{15 \to 12, 15 \to 11\}$
	$\quad 7 \to 2\}$		
$S1(7)$	$\{7 \to 1\}$	$S14(15)$	$\{15 \to 14, 15 \to 11, 15 \to 5\}$
$S2(7)$	$\{7 \to 2, 7 \to 1\}$	$R(16)$	$\{16 \to 16, 16 \to 10, 16 \to 9, 16 \to 5,$
			$\quad 16 \to 13, 16 \to 12, 16 \to 11,$
			$\quad 16 \to 15, 16 \to 14\}$
$S6(7)$	$\{7 \to 6, 7 \to 1, 7 \to 9\}$	$S5(16)$	$\{16 \to 5\}$
$S9(7)$	$\{7 \to 9\}$	$S9(16)$	$\{16 \to 9\}$
$R(8)$	$\{8 \to 8, 8 \to 3, 8 \to 2, 8 \to 5,$	$S10(16)$	$\{16 \to 10, 16 \to 9, 16 \to 5\}$
	$\quad 8 \to 1, 8 \to 7, 8 \to 6, 8 \to 9,$		
	$\quad 8 \to 10\}$		
$S1(8)$	$\{8 \to 1\}$	$S11(16)$	$\{16 \to 11\}$
$S2(8)$	$\{8 \to 2, 8 \to 1\}$	$S12(16)$	$\{16 \to 12, 16 \to 11\}$
$S3(8)$	$\{8 \to 3, 8 \to 2, 8 \to 5, 8 \to 1\}$	$S13(16)$	$\{16 \to 13, 16 \to 12, 16 \to 11, 16 \to 9\}$
$S5(8)$	$\{8 \to 5\}$	$S14(16)$	$\{16 \to 14, 16 \to 5, 16 \to 11\}$
$S6(8)$	$\{8 \to 6, 8 \to 1, 8 \to 9\}$	$S15(16)$	$\{16 \to 15, 16 \to 5, 16 \to 12,$
			$\quad 16 \to 11, 16 \to 14\}$
$S7(8)$	$\{8 \to 7, 8 \to 2, 8 \to 1, 8 \to 6,$	$R(17)$	$\{17 \to 17, 17 \to 14, 17 \to 5, 17 \to 11\}$
	$\quad 8 \to 9\}$		
$S9(8)$	$\{8 \to 9\}$	$S5(17)$	$\{17 \to 5\}$
$S10(8)$	$\{8 \to 10, 8 \to 5, 8 \to 9\}$	$S11(17)$	$\{17 \to 11\}$
$R(9)$	$\{9 \to 9\}$	$S14(17)$	$\{17 \to 14, 17 \to 5, 17 \to 11\}$

TABLE A.4: Object part of the subobject classifier

p	$\Omega(p)$
1	$\{R(1), \emptyset\}$
2	$\{R(2), S1(2), \emptyset\}$
3	$\{R(3), S1(3), S2(3), S5(3), \emptyset\}$
4	$\{R(4), S1(4), S2(4), S3(4), S5(4), \emptyset\}$
5	$\{R(5), \emptyset\}$
6	$\{R(6), S1(6), S9(6), \emptyset\}$
7	$\{R(7), S1(7), S2(7), S6(7), S9(7), \emptyset\}$
8	$\{R(8), S1(8), S2(8), S3(8), S5(8),$ $S6(8), S7(8), S9(8), S10(8), \emptyset\}$
9	$\{R(9), \emptyset\}$

p	$\Omega(p)$
10	$\{R(10), S5(10), S9(10), \emptyset\}$
11	$\{R(11), \emptyset\}$
12	$\{R(12), S11(12), \emptyset\}$
13	$\{R(13), S9(13), S11(13), S12(13), \emptyset\}$
14	$\{R(14), S5(14), S11(14), \emptyset\}$
15	$\{R(15), S5(15), S11(15), S12(15), S14(15), \emptyset\}$
16	$\{R(16), S5(16), S9(16), S10(16), S11(16),$ $S12(16), S13(16), S14(16), S15(16), \emptyset\}$
17	$\{R(17), S5(17), S11(17), S14(17), \emptyset\}$