

140310

**A STUDY ON THE RELATION
BETWEEN LOGIC AND INFORMATION**

**A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY**

BY

TANKUT BEYGU

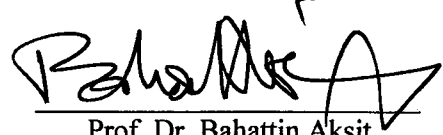
T.C. YÖKSEKÖĞRETİM KURULU
TEZ YERLEŞTİRME VE
DOKÜMANASYON BÖLÜMÜ

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF PHILOSOPHY**

JANUARY 2003

140310

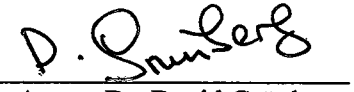
Approval of the Graduate School of Social Sciences


Prof. Dr. Bahattin Akşit
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.


Prof. Dr. Akin Ergüden
Head of the Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.


Assoc. Dr. David Grünberg
Supervisor

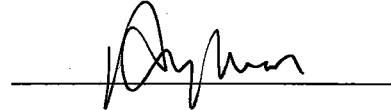
Examining Committee Members

Assoc. Prof. Dr. David Grünberg

Assoc. Prof. Dr. Onur Demirörs

Dr. Ayhan Sol





ABSTRACT

A STUDY ON THE RELATION BETWEEN LOGIC AND INFORMATION

Beygu, Tankut

M.Sc., Department of Philosophy

Supervisor: Assoc.Dr. David Grünberg

January 2003, 109 pages

Gaining prominence at first as a technological concept, information has found a significant role in the study of various problems in philosophy as well as in the diverse fields of science and technology.

Though widely employed and studied, information is a complex notion, as it were, resisting a compact description of its characteristics, appearing non-uniformly in such miscellaneous ways as an organizing principle, a form of knowledge, and a connotation of entropy. Apparently, there is a need for further study to clarify the notion of information and its possible applications.

Situation theory is a competent frame of investigation responding to this need. It allows us to view logical, linguistic and computational issues in a consistent and unified manner which is open to formalization and based on information, referring not only to syntactic aspects or to semantic truth-values, but also to other semantic aspects, e.g., meaning and intension.

On the other side, relevant logics have been opening up promising perspectives for semantical connections of logic.

This study aims at drawing a parallelism between the inference in logic and information flow. The connections between relevant logics and situation theory are examined and modeling of information flow grounded on them is shown to be satisfactory to considerable extent in both syntactic and semantic respects, though not yet complete.

Key Words: Information, Situation Theory, Relevant Logics



ÖZ

MANTIK VE ENFORMASYON ARASINDAKİ İLİŞKİ ÜZERİNE BİR ÇALIŞMA

Beygu, Tankut

Yüksek Lisans, Felsefe Bölümü

Tez Yöneticisi: Doç.Dr. David Grünberg

Ocak 2003, 109 sayfa

Önce teknolojik bir kavram olarak öne çıkan enformasyon bilim ve teknolojinin farklı alanlarının olduğu gibi felsefenin çeşitli problemlerinin incelenmesinde belirgin bir role sahip olmuştur.

Genişçe kullanılması ve üzerinde çalışılmasına karşın, enformasyon kısa, özlü bir tanımı verilmesi güç, örgütleyici bir ilke, bilginin bir formu, entropinin bir konotasyonu gibi türlü şekillerde düzensiz olarak karşılaşılan karmaşık bir nosyondur. Görünen o ki, enformasyon nosyonu ve olası uygulamalarını açıklamak için daha fazla çalışma gerekmektedir.

Durum teorisi bu gereksinime karşılık veren yetkin bir araştırma çerçevesidir. Mantıksal, dilsel ve kompütasyonel konuların formalizasyona açık ve enformasyona dayalı ve yalnızca sentaktik yönleri veya semantik doğruluk değerlerini değil, fakat anlam ve işlem gibi başka semantik yönleri de ele alarak tutarlı ve birleşik bir tarzda görmemizi sağlar.

Öte yanda, ilgililik mantıkları mantığın semantik bağlantılarına yönelik umut verici perspektifler açmaktadır.

Bu alıřma, mantıkta ıkarım ile enformasyon akıřı arasında paralellik kurmayı hedeflemektedir. İlgililik mantıkları ile durum teorisi arasındaki baėlantılar incelenmekte ve bunların temelinde enformasyon akıřı modellemesinin hem sentaktik, hem semantik yönlerden, henüz tamamlanmıř olmamakla birlikte, önemli ölçüde tatmin edici olduėu gösterilmektedir.

Anahtar Kelimeler: Enformasyon, Durum Teorisi, İlgililik Mantıkları.



ACKNOWLEDGMENTS

I express my sincere appreciation for my supervisor Dr. David Grünberg for his encouragement for the thesis subject and his provision of the most favorable circumstances for my study. I am greatly indebted to Dr. İsmail Gül, Dr. Onur Demirörs and Dr. Süleyman Önal for the mathematical and information theoretical insights they provide, otherwise which my study would not be possible. I also want to thank Dr. Ayhan Sol for his friendly support he showed throughout my study.

TABLE OF CONTENTS

ABSTRACT	iii
ÖZ	v
ACKNOWLEDGMENTS.....	vii
TABLE OF CONTENTS.....	viii
LIST OF FIGURES.....	x
CHAPTER	
1. INTRODUCTION	1
2. A CONCEPTUAL EXAMINATION OF INFORMATION.....	3
3. SYNTACTIC DESCRIPTION OF INFORMATION.....	8
3.1. Information Theory.....	8
3.2. Algorithmic Information Theory.....	10
3.3. Assessment of Relevance and Information.....	11
4. DRETSKE'S SEMANTIC THEORY OF INFORMATION.....	13
4.1. An Overview.....	13
4.2. Basic Ideas in the Conceptualization of Information.....	14
4.3. Xerox Principle.....	19
4.4. Relativity and Objectivity of Information.....	19
4.5. <i>De re</i> Report of Informational Content.....	21
4.6. Intensionality of Information.....	21
4.7. Analog and Digital Coding.....	23
4.8. Sensory versus Cognitive Processes.....	24
4.9. Conception of Belief.....	25
4.10. Knowledge.....	25
4.11. Against Skeptical Thesis.....	27
4.12. A Conflation in Dretske's Model.....	29
5. SITUATION THEORY.....	31
5.1. Framework of Situation Theory.....	31
5.2. Ontology of Situation Theory.....	32

5.3.	The Notion of Constraint.....	35
5.4.	Situation-Types.....	36
5.5.	Statements, Propositions and Infons.....	38
5.6.	Partiality.....	38
5.7.	Situation Semantics.....	39
5.8.	Situation-Theoretic View of Conditionals.....	41
5.9.	Comparison of Possible Worlds Semantics and Situation Semantics.....	41
5.10.	The Issue of Well-Foundedness.....	43
6.	A SITUATION-THEORETIC MODELING OF INFORMATION FLOW.....	45
6.1.	Situations in a Perspective.....	45
6.2.	Information Flow.....	46
6.3.	Principles of Information Flow.....	46
6.4.	The Source/Receiver Perspective.....	48
6.5.	Information Channels.....	51
7.	RELEVANT LOGICS.....	60
7.1.	Paradoxes of Implication.....	60
7.2.	Implication Theory of Classical Logic.....	62
7.3.	Main Themes of Relevant Logic.....	63
7.4.	Model-Theoretical Semantics for Relevant Logic.....	64
7.5.	Proof Theory of Relevant Logic.....	68
7.6.	Overview of Relevant Logic Systems.....	73
8.	SITUATION SEMANTICS FOR RELEVANT LOGICS.....	76
8.1.	Overview.....	76
8.2.	Mares's Interpretation.....	78
8.3.	Restall's Interpretation.....	83
8.4.	Complementary Arguments to the Models.....	87
	CONCLUSION.....	88
	REFERENCES.....	90
	APPENDIX	
	REVIEW OF SOME CONCEPTS IN PROBABILITY THEORY.....	93

LIST OF FIGURES

FIGURES

1. Basic communication system.....14
2. Information flow backwards causality.....17
3. The modular law for channels.....55



CHAPTER 1

INTRODUCTION

There is a Leibnizian vision in logic which seeks a harmony of “logic of events” and “logic of arguments.” It is an arduous project of making logic descriptive while keeping it normative. This vision is conflated with another Leibnizian vision, that of “logic by computation.” The bridging notion is information.

It is not accidental that the conflation of visions occurs beyond metaphysical arguments in a world, which Floridi rightly calls *infosphere*, where storage, processing and circulation of information has become a building block of the human being’s ecosystem (see [18] for an extensive explication of the concept). Information reflects the rising consciousness about the epistemic contents that have extra-mental existence and that are needed by the enormous variety of systems in a similar sense that mind needs knowledge.

Following that it was originated in communications engineering and found a counterpart in thermodynamics, the notion of information has developed through the works in diverse disciplines, such as biology, psychology, physiology, linguistics, computer science, and philosophy, beside physics and engineering.

Though information enables us to view various principles and phenomena studied in this broad range of disciplines in one picture, making many underlying relations between them explicit, and providing generalized explanations, it is by no means a uniform notion, mostly it comprises ideas and methods inspired in one discipline and applied in another. However, this does not create an obstacle to the extent that it would prevent an integrated information-theoretic approach to a rich spectrum of interesting problems, rather this situation should be seen as a normal state for a seminal framework of investigation.

Information-theoretic approach can be said to be a paradigm of research by itself. It is apt to develop rich formalism to express adequately new as well as many old problems and to be supported by the progress in various disciplines. The boundaries of this approach is varying; there are views extending the role of information to that of an elementary physical

entity at the same rank with matter and energy (see [30] for the claims of such a view), or to being the principle grounding the mental processes.

As to the scope of validity in order to apply the notion of information, we will remain within the area common to any system of intentional character to some degree. Within this area, we will deal particularly with a rigorous model of information flow.

Flow is the principal act for information, though information may be conserved in a static state, as in DNA molecules of organisms, or on the printed paper, it is the flow of information that provides its essence. By the metaphor of flow, two aspects of information should be understood:

First, information is *communicable*; it is received from the physical environment, from another system, or from one part of a system destined to another. This aspect contains mostly technological issues of signal transmission, and is covered mainly under the communication theory. Many of the syntactic properties of information are shaped by the communicational requirements. However, it contributes to our understanding of semantical properties of information as well, actually, the notion of information has been originated in the communication theory studies. As it will be exposed, the syntactic properties are strongly connected to the semantic properties in the case of information.

Second, information is *processable*. It can be transformed into different semiotic forms, combined with other pieces of information and give out new information. There is a growth and loss of information.

Information will be taken to be a primitive notion; there will not be an attempt to reduce it to some other notions. Information simply exists in virtue of the world comprising regularities and systematic relations. Otherwise, in a world of complete chaos, no information would exist to communicate and to process.

Though it can be hardly said that the information-theoretic studies follow a path signed by well-defined problems, the topics will be presented roughly in a linear progression in order to explicate better the connection between them. After the examination of the information notion and situation theory, giving emphasis on information flow, first, a formal modeling of information flow in situation-theoretic terms, and then, modeling from the point of relevant logics' view is surveyed, and in that context, situation semantics is asserted as a suitable semantics for relevant logics.

The present work owes much to many works; the particularly benefited ones are given in the references. The integrated character of our subject-matter and the same manner the benefited material handles it, supplementing each other, lead us to give references under broad coverage, rather than at specific points. This should be understood as that the references are used frequently and in a scattered fashion.

CHAPTER 2

A CONCEPTUAL EXAMINATION OF INFORMATION

Our world is populated with systems. There is a tremendous diversity of biological systems that have been losing its members day by day because of our tactless activities. There are artificial systems; some of which take over the jobs that have been formerly done by the human being and perform them better than their producers, some of which do the jobs beyond our capability of our naked being. Systems show set-theoretical behavior; they interact with others and their environments in multifarious ways, composing different systems. An ecological system, for instance, demonstrates a harmony of complex relationships among the organisms that inhabit the system and the environment -the environment itself can be composed of biological systems beside physical circumstances. Let us call these systems *objective systems* in contrast to the *subjective systems*, by which we apply the fruitful idea of system onto many bodies of knowledge to model entities and phenomena, despite that they do not have goal-oriented structure, such as the astronomical systems. The distinctive property of the objective systems is their *information processing* capability as their essential ontic basis.

The diversity of the manifestations of information and its behavior within the systems lead us to construct the definition of information as a semantic field taking into account the contexts of applications. Here, we will deal with some prominent aspects of information.

A crucial property of information is *objectivity*. Let us consider an experienced seaman who gazes at the horizon of red color in the evening and the low clouds; these compose the signal carrying him the information that the next morning will be much colder. The *signal* is non-mental, it is objective in the sense that even if there is no cognitive agent, like the seaman, that receives it or knows how to extract it, there is the signal (of information).

The concept of information is closely connected with causality. Indeed, it may become blurry in some cases to distinguish the informational effect and the physical cause-effect relation. Any information flow through the systems can be explicated as a collection of

causes and effects. Certainly, all the relations within a frog can be described in the same manner as a solar system in purely physical terms, however, both the frog and the seaman demonstrate what is generally referred to as *intentional* activities. There is an intrinsic relation between the agent of intentional activities and *intensionality* of information.

Let us suppose there are four deer in a prairie that is a hunting area of a tiger T_E which is a cognitive agent that perceives the prairie purely in an extensional way, and of another tiger T_I which is a cognitive agent that perceives the prairie purely in an intensional way. Furthermore, let us suppose that the tigers react differently for each change of situation. The deer can be represented as a set that can be described extensionally as

$$E_4 = \{\text{deer}_1, \text{deer}_2, \text{deer}_3, \text{deer}_4\}$$

The same set can be described intensionally as

$$I_4 = \{d : d \text{ is a deer in the prairie}\}$$

Let us suppose that the third deer leaves the prairie, then we obtain in the same manner

$$E_3 = \{\text{deer}_1, \text{deer}_2, \text{deer}_4\}$$

$$I_3 = \{d : d \text{ is a deer in the prairie}\}$$

The distinction between two imaginary tigers is clear: While for T_E , the two situations are different ($E_4 \neq E_3$), accordingly it reacts differently, there is no change ($I_4 = I_3$) for T_I , and therefore, there is no change in its behavior, though causally perceives the change. This sense of intensionality is the basis for its logical characteristic that its causal characteristic subvenes. There is another sense of intensionality, which Dretske mentions in [13] as the property of information. The latter sense is the *non-substitutability* of information, which emphasizes the causal characteristic. That is, the “ S is F ” and “ S is G ” do not have the same informational content, even if F and G are extensionally equivalent.

As an example of objective and intensional aspects, let us consider an ancient inscription on a piece of stone lost and rubbed by the passing years. Suppose the stone is unearthed later; though the physical conditions of the stone have changed much since the time of the people that it was originally addressed and it is difficult to see what has been

written on the stone, it is still possible to read it and get the informational content the inscription was intended to carry. Furthermore, the information the stone carries has not disappeared and, when found, reappeared; it was there since the time it was inscribed.

Upon the system-relatedness and intentionality, we should establish the fact that existence of information presupposes a *nomic* (regular with respect to natural laws) universe, no system could exist, let alone information.

In general, information can be represented in the declarative form that situation theory provides, such that a relation (property) R either holds or does not hold (determining polarity) for a set of objects (parameters) $a_1, \dots, a_n$. We can use the representation $\langle\langle R, a_1, \dots, a_n, i \rangle\rangle$, where i is polarity, of an *infor*. Infor is the concept introduced by Devlin [10] referring to the smallest unit of information irrespective of its implementation, determined by the context of use.

Given the formalism, we can analyze an infor as a collection of *data*. Data is usually construed as *what makes difference*. For example, in the statement “Empire State Building is 81 m. taller than Eiffel Tower,” the names of the buildings and the height difference are data. The generally used unit to measure information, the *bit* (*binary digit*) is virtually the simplest representation of difference, taking values 0 or 1. What counts as data and not data is determined by the context and when the context is well-formed, we have the information containing data. Hence, information is an entity complete in itself with respect to semantic content, it can be said that it is composed of data synthesized so as to be meaningful.

A principal property of information is its veridicality, that is, an infor of that kind exists within its truth conditions. In other words, its semantic content satisfies always truth conditions. Therefore, “true information” is redundant in a similar manner held in the deflationary theory of truth. There can be false propositions and incorrect data, but not “false information” (*misinformation* and *disinformation*); it is not information at all

In order to clarify the phrase, a conceptual analysis is given by Floridi [17], which underlines the logical distinction between *attributive* and *predicative* uses of an adjective. A usual example is the uses of the adjectives French, as predicative, and good, as attributive, sentences “Brigitte Bardot is a French actress” and “Brigitte Bardot is a good actress”. If an adjective in a compound is attributive, the compound cannot be divided into a conjunctive form. That is:

Brigitte Bardot is a French actress $\equiv$ Brigitte Bardot is French &

Brigitte Bardot is an actress

whereas

Brigitte Bardot is a good actress $\neq$ Brigitte Bardot is good &
Brigitte Bardot is a French actress

Thus, “ σ is a false infon” cannot be stated as “ σ is false” and “ σ is an infon,” “false infon” is merely a confusion in *attributive* and *predicative* uses of an adjective.

There are several justifications for the veridicality (nondefeasibility) of information. First, information is in the same epistemic category with knowledge, and requires truth just as knowledge does. Let us consider the contrary, the case pointed out by Floridi, that false information were a variety of information, then there would be no loss of information, either deliberate or indeliberate, any change in a message would produce new information that would add up to the present body of information, a result which is unacceptable in the framework of our notions of system, information and regularity. We need that informative content can be lost both physically, syntactically and semantically in order to be able to apply the information notion in diverse contexts.

The second point is that the causal characteristic of information leads to this, just as a series of interdependent physical phenomena cannot have an alethic value, it merely exists.

There is not a consensus on the view of veridicality; among the authors whose works are referred in this study, Devlin, Barwise and Seligman hold that misinformation is possible, however, this variation does not affect the arguments within the boundaries of the present study.

Thus, we can arrive at the following definition of infon σ , hence, information ([16], p.10):

σ qualifies as objective semantic information if and only if:

1. σ consists of a *non-empty* set (D) of data (d);
2. The data in D are *well-formed* (wfd);
3. The wfd in D are *meaningful* (mwfd = δ);
4. The δ in D are *truthful*.

D can contain different types of δ s, but it cannot be empty.

Information can be classified in various ways. One classification that we will follow is to characterize a message (in the general sense; any item that conveys information) in terms of the semiotic aspects by which it obtains the informative value:

- *Syntactic information* arises from the symbols of which messages are composed and the symbolic interrelations. Syntactic information gives the capacity to carry semantic content, hence, the more syntactic information, the less structural ambiguity is.
- *Semantic information* is due to the referential aspect of information.
- *Pragmatic information* is related to the usage; determines how much information certain message carries for a cognitive agent, depending on the context.

Thus, syntactic information mainly deals with the form of information, whereas semantic and pragmatic information are related to the information content. Let us consider the following sentences:

1. There is a bus to my hometown every day of the week except Sunday.
2. There is a bus to my hometown on Monday, Tuesday, ..., Saturday.
3. Electricity is cut off in the southern part of the city on Wednesday.
4. Electricity and water supply are cut off in the southern part of the city on Wednesday.

The information stated in (1) and (2) are syntactically different, while they are semantically and pragmatically identical. They have the same meaning and are both equally informative. (4) is semantically more informative than (3). (3) and (4) can be considered not differing significantly in syntax, however, they may differ pragmatically much; for example, the information contained in the sentences (3) may be of no use for someone who does not have much to do with electricity, but is interested in water supply, therefore, (4) is relevant to him. (3) and (4) pragmatically has no significance for people living in other cities.

We will deal with semantic information and to the extent that it contributes to our understanding of semantic aspects, with syntactic information. For the connection between semantic and syntactic aspects, an analogy can be made with the Diophantine equations of which the solutions are found just from the number-theoretic properties of integers.

Quantitative treatment of information is usually associated with syntactic information, however, there are studies to develop quantitative arguments on semantic information as well (see [17] which provides an overview of this subject).

CHAPTER 3

SYNTACTIC DESCRIPTION OF INFORMATION

Syntactic aspects of information will be examined under the two main topics: Shannon's communication theory which uses the idea of entropy as the measure of information (see [29]), and algorithmic information theory which bases its quantitative analyses on the Kolmogorov complexity (see [21]).

3.1. Information Theory

Information theory has originally been developed to address the communications engineering issues, providing mathematical modeling and analysis of communication systems that carries information. In this context, the theory regards information as only those symbols that are to be transmitted by the source and *uncertain* to the receiver, and does not consider the semantic aspects of information.

One of the main virtues of information theory is that it brings formalism, based on the idea that information is a quantity resolving uncertainty by elimination of *relevant* possibilities.

Thus defined, the amount of information, produced by an information source is a measure of its *entropy*, depicting the probabilistic behavior of a source of information. The concept of entropy is connected with its use thermodynamics in the aspect of order and disorder, which is the original field of this concept. Entropy turns out to be a measure of disorder in an environment.

Let us suppose a random experiment that involves the observation of messages emitted by a discrete source per unit of time. The messages are modeled as a discrete random variable, S , which takes symbols from a fixed finite alphabet set as its values. The symbols are assumed to be emitted in a successive order and are statistically independent of each other.

Let the alphabet set be $S = \{s_0, s_1, \dots, s_{n-1}\}$ with probabilities

$$P(S = s_k) = p_k, k = 0, 1, \dots, n-1$$

which depicts the emission of symbols s_k by the source with probability p_k . This set of probabilities constitute the sample space of the experiment, thus

$$p_0 + p_1 + \dots + p_{n-1} = 1$$

In case that the event $S = s_k$ occurs, if $p_k = 1$ and $p_i = 0$ for all other symbols, then there is no *information* gained by the receiver, since the receiver already knows the message. On the other hand, if the source symbols occur with different probabilities, then there is ambiguity in the message to be resolved by the transmitted information, which makes the difference between before and after the occurrence of the event $S = s_k$. The less probable an event is, the more information we gain when it occurs.

We define the amount of information gained after observing the event $S = s_k$, which occurs probability p_k , as the logarithmic function

$$\begin{aligned} I(s_k) &= \log_2(1/p_k) \\ &= -\log_2 p_k \quad k = 0, 1, \dots, n-1 \end{aligned}$$

The base of the logarithm in equation is conventional; it is settled as the standard practice to use a logarithm to base 2 due to the wide application of two-valued systems. The resulting unit of information is the *bit*. For example, if $p_k = 1/2$, then $I(s_k) = 1$ bit, that is, one bit is the amount of information gained when one of two possible and equally likely events occurs.

By above discussion, the formulation of information holds the following properties that can be intuitively interpreted:

1. $I(s_k) = 0$ for $p_k = 1$
2. $I(s_k) \geq 0$ for $0 \leq p_k \leq 1$
3. $I(s_k) > I(s_i)$ for $p_k < p_i$
4. $I(s_k s_i) = I(s_k) + I(s_i)$, where s_k and s_i are statistically independent.

If an object composed of a string of symbols, for example a message, is random, there is a sense in which the *descriptive complexity* of the event $S = s_k$ is $\log_2(1/p_k)$ upon the observation that $\lceil \log_2(1/p_k) \rceil$ ($\lceil \cdot \rceil$ denotes the smallest integer value equal to or greater than the argument $\log_2(1/p_k)$) is the number of bits required to describe s_k by a binary encoding. For example, suppose that there are three balls colored as green, red, and yellow. If one of the balls is to be taken out under equal probability, and the outcome is to be send to a receiver, two bits are needed to describe that piece of information; a representation scheme can be as follows:

- 01 for the green,
- 10 for the red, and
- 11 for the yellow.

In the binary code approach:

$$1/p_k = 1/3, k = 0, 1, 2 \text{ for green, red and yellow.}$$

$$\log(1/p_k) = 1.585$$

$$\lceil \log_2(1/p_k) \rceil = 2 \text{ bits.}$$

3.2. Algorithmic Information Theory

Indeed, various types of description language can be used in the definition of information. Selecting which language to use affects the characteristics of the definition. Another fundamentally important description language is based on algorithms. As independently observed by Kolmogorov, Solomonoff and Chaitin, the algorithm that reproduces an information content can describe the same content.

The central notion of this approach is the *algorithmic complexity*, which is usually referred to as the *Kolmogorov complexity*. Kolmogorov complexity of a string object is the length of the shortest binary computer program that describes that object. Thus, the Kolmogorov complexity dispenses with the probability distribution.

It can be shown that the definition of complexity is essentially computer independent, i.e., that the length of the shortest program does not depend on the computer, up to an additive constant. Therefore, the expected length of the shortest binary computer description acts as a universal code, which persists uniformly for all probability distributions.

It is not generally a decidable question to find such a minimal program that gives the complexity in an absolute manner; it may take infinitely long time. However, very short, not necessarily minimal, programs can be found in practice. The idea of finding such short programs leads to universal codes as well as a good basis for inductive inference and a formalization of Occam's Razor.

In order to formalize the notion of algorithmic complexity, the *universal Turing machine* model of computation can be used, as the conceptually simplest universal computer, providing a standard basis for the following definition (see [21] for a thorough analysis of the subject):

Suppose that x is any member of a set S of objects, and $n(x)$ is a natural number that enumerates x . Let $n = f(p)$, where f is a partial function over nonnegative integers and p is a natural number with which $l(p)$, the length of the binary string representing p , is associated. Then, the *complexity* of x with respect to the method f that specifies x , is

$$C_f(x) = \min\{l(p) : f(p) = n(x)\}$$

Hence, $C_f(x)$ is the length of the minimal description of x . As an illustrative application, p can be taken as a computer program, f as a computer and $C_f(x)$ becomes the minimal length of a program that runs on f in order to compute x without additional input. Therefore, $C_f(x)$ indicates, in the given sense, the amount of information contained in the object x .

An interesting study that contributes to our understanding the relation between logic and information is given in [9]. In that study, an information-theoretic approach to Gödel's incompleteness theorem is presented on the basis of algorithmic information theory, making use of the idea that information contained in a theorem to be derived from a set of axioms cannot exceed the information contained in the axioms.

3.3. Information and Assessment of Relevance

Beside the definitional features of information, syntactic aspects of information contribute to the assessment of relevance in a group of informational contents. In the next chapters, we will deal with the constraints of relevance that the inferences in logic should satisfy by referring to the semantic concepts. However, the syntactic properties of information supply an interesting and wide horizon to measure and enforce intensional criteria on the logical statements in a quantitative manner, so we will touch upon them briefly.

In this line, two approaches can be mentioned as example. One (see [1]) is a method to measure the information in a formally expressible theory in terms of the possible logical models of the theory and assess the case when new information is added to the theory from the point of a decision maker's point of view.

The other one is the information distance based the Kolmogorov complexity (see [21]). The idea can be described as measuring the shortest program that transforms a syntactically defined object into another one and vice versa.



CHAPTER 4

DRETSKE'S SEMANTIC THEORY OF INFORMATION

4.1. An Overview

A plausible base to investigate the semantic dimension of information is Fred Dretske's theory, which is sketched mainly in [13] and is supplemented later by several of his papers ([14] and [15] among others). The theory comes to the fore as a culmination of information-theoretic approaches that have flourished with the success of the communications and systems engineering studies, and attributes a primary role to the notion of information in any collection of entities that can be identified as a system.

The theory's conceptualization of information is grounded on some of the elementary ideas in Shannon's mathematical theory of information. The mathematical theory is purely quantitative; it provides a measure of information to be associated with a given state of affairs, and a measure of this information when it is distributed to the communication system, however, it contributes to our understanding of the information concept through its syntactic relation to information. The propositional content of a signal (events, structure, or state of affairs) is described accordingly. It takes information source and channel into an account of how much a particular signal carries about a source and what information this is.

Information is an objective commodity that can be produced and processed by, transmitted from any system, independent of the existence of any cognitive agent. Information eliminates equivocation within a given state of affairs, thus, the more a message gets closer to be univocal, the more informative it is.

The theory attempts to generalize its arguments to all kinds of systems characterized by some intentional behavior (organisms including the human being, computers, and mechanical devices). According to Dretske, the higher-level accomplishments associated with intelligent life can then be seen as manifestations of progressively more efficient ways

of handling and coding of information. Meaning, and the constellation of mental attitudes that exhibit it, are manufactured products of which raw material is information.

Eventually, the theory of information is applied to propositional attitudes that are in the focus of epistemology: belief and knowledge. Perception and skepticism are examined in this context. Intensionality suggests one of Dretske's theses, that the intentionality of our cognitive attitudes is a manifestation of their underlying information-theoretic structure; knowledge is information-produced, perception is a coding of information.

The exposition is intended to determine such aspects of Dretske's theory that bear significance to the purposes of this study; it is not a commentary, nor a thorough examination of the theory. The examples are taken from Dretske and slightly modified in order to provide uniformity and ease of reading.

4.2. Basic Ideas in the Conceptualization of Information

The theory assumes a model of information flow that adopts the basic communication scheme: There is a source of information, a receiver, and a communication channel over which the information is conveyed. There may be more than one source or receiver and channel linking them.

The source is any event, structure or state of affairs that generates information. The receiver is a system that has the disposition to discern the incoming information. The channel between them is theoretically their correlation; usually, it is a medium for the information-carrying signal to travel; it does not append new information, however, it can cause partial or total loss of information generated at the source.

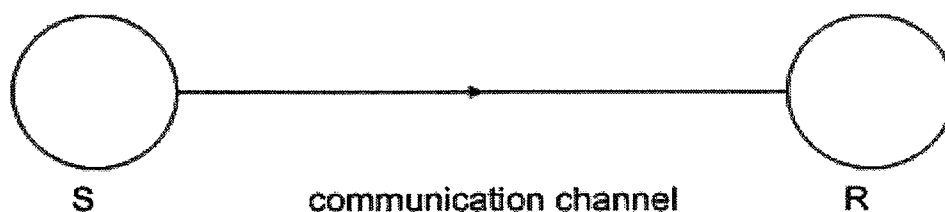


Figure 1: Basic communication system.

Since many issues of which the theory aims to give account are examined from the receiving system's point of view, and within the boundary of the theory the signal can be safely assumed as to reproduce the information at the source, the source will be ignored, unless otherwise is stated.

Inheriting the idea in the communication theory that information is what eliminates comparable possibilities that a signal possesses, measurement of information becomes a

definitional component of the term, however, Dretske's theory deals with the amount of information contained in a particular signal, not as in Shannon's theory, in the average amount of information conveyed.

The amount of informational content a signal S carries is denoted by $I(S)$. The exclusive method to measure the amount of the content is to calculate how many binary decisions are needed to indicate a particular possibility. For example, it takes 5 binary decisions to locate a chess piece on a checkerboard. Thus, each square, which corresponds to a possibility of the piece's location, can be enumerated, for instance, as 00000 for A1, 00001 for A2, and proceeding through the other squares, 11111 for H8, where each digit of a number represents a binary decision. Therefore, a signal carrying the information of where the piece stands possesses an amount of 5 bits of information. A general formula for a measurement of information can be given as:

$$I(S) = \log_2 P$$

where P is the number of possibilities. It should be noted that the amount need not be an integer, but can be a real number. This implies that there is an amount of informational content, a quantity that can be determined, absolutely or relatively, on this capability. Then, the amount of information transmitted by S to the receiver R becomes

$$I_S(R) = I(S) - \text{Equivocation}$$

Equivocation is the loss of informational content, adding ambiguity to the informational content at R . For example, let us suppose this situation: It occurs that only 4 bits could be reliably sent to the receiver to locate a chess piece on the board. In this situation, instead of 64 parts, only 32 parts of the checkerboard can be indicated. An ordinary method that is enumerating 2 squares with one bit results in an ambiguity of 2 squares; the receiver, though gets the information that the piece is not on the 62 squares, cannot pick out on which one of the remaining pair of squares. If there is no equivocation, then the informational content at R is the same as on the signal S , which is the case of univocation.

It should be noted that the environment depicted conforms also to a causal chain of events, and thus, to causality reconstructed in a different perspective. Let us consider for example the case in which each square has a weight sensor that closes a circuit when a piece is located on the square, thus turning on a lamp dedicated to the square on a remote site, so

that someone could perceive where a piece is standing without seeing the checkerboard. At one level, which can be described as physical, the process can be regarded as a series of mechanical and electrical causes and effects.

The theory views the events on another level, which can be described as systemic in distinction to physical, and asserts for information a relatively independent ontological status from causality, which is a justification for information-theoretical analysis.

Information occupies a causal role as the cause of something, on a more abstract level than the physical event carrying the information. Let us consider the more complex information processing systems for which there are an indefinitely large number of sensory inputs with the same cognitive outcome, but having no identifiable physical (non-relational) property in common. What makes them have the same cognitive outcome is the information they bear. Such relevant causal regularities induce a characterization of the cause in terms of its relational (informational) properties.

For this ontological argument of the theory, it suffices to take causality as a relationship of regular succession between the events, while not requiring determinism; every event can have a cause without every event can be determined, hence, in principle not predictable. The theory accepts the simple assumption in claiming that causality is not necessary to the flow of information ([13] p.32):

If C is the cause of E , then C is an essential part of some nomically sufficient condition for E .

In other words, causality is the manifestation of a law-like, succession between events of type C and events of type E under relevantly similar conditions.

The crucial point lies in the distinction between being an effect of an event and having disposition of receiving carried by an event. A chandelier may tremble because of the sound waves produced by an opera singer, but no song is communicated to it. The phenomenon of language is a prominent example for the relative independence of informational relationship from causal relationship among objects that have linguistic capability and do not have.

Conversely, there are situations in which information is communicated even though causal relationship is not established.

Let us consider the situation in which a television station A broadcasts to both television receivers S and R over physical links (for instance, by electromagnetic propagation through atmosphere or cable network). Though S and R are isolated from each other, there is an informational link between them, since it is possible to learn something about R by looking at S , and vice versa. The correlation between the events that occur upon the screens of these two receivers is not accidental.

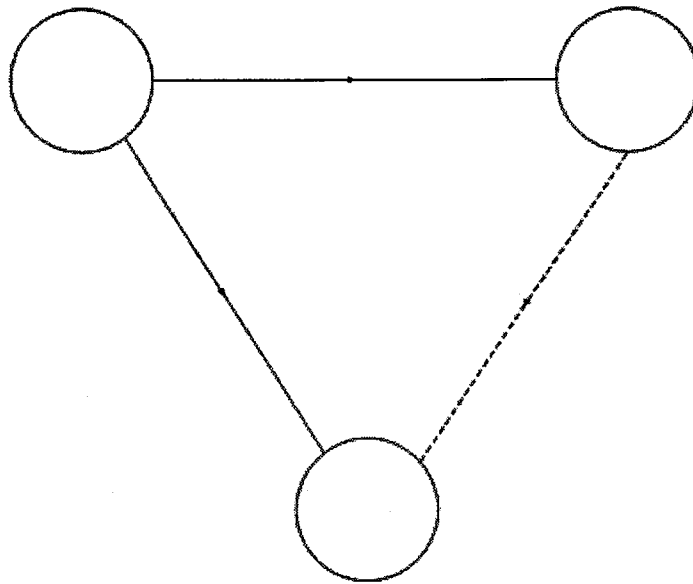


Figure 2: Information flow backwards causality
(the dotted line indicate the informational link).

There is an aspect of the informational link between S and R to be distinguished from a causal link: Nothing at S causes anything at R or vice versa; yet R contains information about S and S about R, even if R is so farther away from A than S that at R events occur later than S, hence, R has the information about what *will* happen at S. Certainly, no physical signal can travel backwards in time carrying information from S to R; this is an informational link.

Another case for non-causal, but informational relation is that two events are correlated through a common cause. For example, “if the windshield has become icy, then my car will not start up,” assuming that both are the consequences of the cold weather.

The information content of a signal is a function of the nomic (law-governed) relations it bears to other conditions.

Therefore, causal relationship between objects is neither sufficient nor necessary condition for the flow of information from one object to another.

That is why the *communication channel* may be thought of as the set of dependency relation between sender S and receiver R. If the statistical relations defining equivocation and noise between S and R are appropriate, then there is a channel between these points; information is communicated between them, even if there is no direct physical (causal) link connecting S and R.

It should be emphasized that the idea in making such a difference between informational relationship and causal relationship is not an attempt to separate a part of space-time continuum where causality does not rule, but to set the informational relationship on different level like mental states.

If two conditions are statistically independent, then the one event carries no information about the other. When there is a lawful regularity between two events, statistical or otherwise, then we can speak of one event is carrying information about the other. Even if there is a coincidence of seemingly statistical dependence, this does not carry information because one event's occurrence does not affect the *conditional probability* of the other's, thus they are statistically independent.

Communication theory has significance about the probability of certain specific conditions given certain specific signals, since the quantities of interest to communication theory are statistical functions of these probabilities. These conditional probabilities provide qualitative arguments through quantitative determinations for the information that a particular signal carries about a remote source. Then, according to Dretske, we can shape an understanding of semantic aspects of information that allows an account of distinction among pieces of information whether they have the same quantity or not.

On this purpose, the content of the signal, the information it carries, should be expressed by a sentence describing the condition at the source on which signal depends in some regular, lawful way.

If the following conditions are met, Dretske asserts, a signal that carries an informational content is said to exist:

1. The signal carries as much information about s as would be generated by s 's being F .
2. s is F .
3. *The quantity* of information the signal carries about s is (or includes) *that quantity* generated by s 's being F (that is, not s 's being else).

The theoretical definition of a signal's *informational content*, which satisfies the above conditions, is expressed as follows ([13] p.65):

A signal S carries the information that s is F = the conditional property of s 's being F , given S (and k), is 1 (but, given k alone, less than 1),

where s is an indexical element referring to some item at the source; the parameter k ($0 \leq k$) relativizes information to what the receiver already knows about the possibilities at the source.

This definition has several justifications. (1) is satisfied because if the conditional probability of s 's being F (given r) is 1, then the equivocation of this signal must be 0, thus $I_s(r) = I(s_F)$. Since information is considered to bear truth just as knowledge, any threshold other than 1 occurs to be arbitrary. Dretske gives the example of drawing randomly a ball from a bag with 94 red balls and 6 white balls. It cannot be said that color of the ball drawn is *known* just from the composition of the balls. Probability of being red is much higher (0.94), but all can be said that it is more probable that it is red.

The suitability that the conditional probability is 1 can be justified by the additivity of information. Let us consider the contrary case in which a signal carries the information that s was F with the conditional probability 0.9, then the signal could carry the information that s was G with probability of 0.9 as well. However, in the case that s was F and G , their *joint* occurrence might have probability less than 0.9, which is against our acceptance that addition of information should not change reliability.

Condition (2) is satisfied because if the conditional probability of s 's being F is 1, then s is F . Information is identified as a first-order epistemic commodity (in contrast to meaning); a structure can carry the information that s is F only if that s is F is factual. This follows directly from the veridicality of information.

Condition (3) is satisfied because this definition assures that whatever *other* quantities of information the signal may carry about s , the signal includes the correct quantity that provides a causal role for information.

4.3. Xerox Principle

If C carries the information that B , and B 's occurrence carries the information that A , then C carries the information that A . Information about the original (A) is not lost by perfectly reproduced copies (B of A and C of B). The *flow* of information would be impossible without the transitivity the principle describes. If we put the threshold of information at anything less than 1, the principle is violated.

The relativity of information is worth examining in itself, since it is asserted that information is objective as well.

4.4. Relativity and Objectivity of Information

The model assumes that the only connection between the source and the receiver is over the channel and one-way (simplex -in contrast to duplex) from the source to the receiver. Thus, the possibilities (and their probabilities) are determined and exist at the

T.C. YÖKSEKÖĞRETİM ENSTİTÜSÜ
DOKÜMAN YÖNETİMİ VE KÜTÜPHANECİLİK

source independently of the receiver. There may be surely another communication channel *from the receiver* to the source that provides changes in the state of affairs at the source, but such cases are regarded as iterations of the model's base case. This independency is the objectivity of information.

On the other hand, the amount of information a signal carries, and hence *what* information it carries, depends on the *background information* denoted by the parameter *k* of the above definition. Background information is what the potential receiver already possesses about the various possibilities that exist at the source.

The parameter *k* is reducible by recursive applications of the definition. For instance, if the information that, on the remote chessboard, the white king is not on E1 by some signal, this carries the information it is on G1, if the receiver has already the information of other possible positions to which the white king is allowed to be moved. However, to other receiver lacking such background information, the same signal does not bear this information, though it still carries the information that the king is not on E1. Thus, a receiver with background information gains more information *relative to* the one with less background information from the same incoming signal.

This gaining action can be identified as an inference, but the theory makes a distinction between the information gained by the present informational state of the receiver with its background information that eliminates a set of possibilities of the source, and the information gained by a logical process, which may not need any background information to reach the same final informational content.

A second relativity dimension comes from the *nested information*: The theory regards that anything carries the information that *s* is a square, also carries the information that it a rectangle, a parallelogram, *not* a circle, or square, and so on. Hence, in general, a signal's informational content is not unique; one piece of information is *nested* in the other. If a signal carries the information that *s* is F, it also carries all the information *nested in s's* being F.

There is analytically or nomically nested information within some piece of information. No single piece of information is entitled to the status of *the* informational content of a signal. The receiver of the signal may be more *interested* in one piece of information than in any other, and may succeed in *extracting* one piece of information without another, but these differences are irrelevant to the information the signal contains.

The linguistic meaning of an utterance may be unique (distinguishable, for instance, from what it implies), but not the information carried by that utterance. This feature of information and, more strongly, nomically nesting of information serve to distinguish it sharply from the concept of *meaning*.

It should be noted that the relativity of information does not conflict with its asserted objectivity; thus, it can be concluded that information is conserved to be objective, though it may not be absolute.

4.5. *De re* Report of Informational Content

Let us briefly touch upon the distinction between *de re* and *de dicto* statements. Consider a statement $\exists x A(\phi(x))$, in which A is a modal, indicates that there is an x and about *that* x (hence, *de re*), says that x is ϕ . Another statement $A(\exists x \phi(x))$ says that there is a proposition $\exists x \phi(x)$ and talks about *that proposition* (*de dicto*).

According to this distinction, the definition of informational content gives an account of the signal's *de re* (in contrast to *de dicto*) informational content; a content that might be expressed by saying that r carries the information *about* s that it is F . Because what is described is a relation between what is expressed by an open sentence (" $\dots$ is F ") and some individual s , which is merely an indexical and should not be associated with the *de dicto* propositional contents. Two factors determine a signal's *de re* informational content:

1. The individual s about which the signal carries information,
2. The information (determined by the open sentence " $\dots$ is F ") it carries about that individual.

The theory holds that *de re* beliefs and knowledge are more fundamental than their *de dicto* counterparts.

4.6. Intensionality of Information

The information a signal carries has the *intensionality* associated with such intentional attitudes as believing and knowing, which forms the basis of information's semantic character. For example, "she believes that s is F " is an intensional sentence, because even when F and G are *co-extensional* (they are true of exactly the same things), G cannot be substituted for F in this sentence without allowing the possibility of a truth-value change. In other words, if a signal carries the information that s is F , it does not necessarily carry the information that s is G despite the *extensional* equivalence of F and G . A state of carrying information about a source involves a certain intentional state relative to that source. If structure S carries the information that t is F , it does not necessarily carry the information

that t is G even though nothing is F that is not also G . The information contained in a structure defines a propositional content with intentional characteristics.

The intentionality of information leads to ordering of information-processing systems with respect to their intentional states; to qualify for *cognitive* attributes a system must be capable of occupying *higher-order intentional states*. Three orders of intentionality are identified with the conditions to be held jointly ([13] p.172) are as follows:

First order of intentionality:

- (a) All F s are G .
- (b) S has the content that t is F .
- (c) S does not have the content that t is G .

Second order of intentionality:

- (a) It is *nomically* necessary that F s are G .
- (b) S has the content that t is F .
- (c) S does *not* have the content that t is G .

Third order of intentionality (semantic content):

- (a) It is *analytically* necessary that F s be G .
- (b) S has the content that t is F .
- (c) S does *not* have the content that t is G .

All information-processing systems exhibit the first order of intentionality. Knowledge and belief have a high order of intentionality, distinguishing genuine cognitive systems from mere processors of information. Any propositional content exhibiting the third order of intentionality is called a *semantic content*. S carries the information that t is F as the *most specific* piece of information about t that S carries. Thus, if G is nested (nomically or analytically) in F , S has both that t is F and that t is G as its *informational contents*, but only the former, not the latter, as its *semantic content*. Information-processing systems are incapable of separating the information that t is F from the information that is nested in that t is F .

Therefore, a belief exhibits a higher-grade intentionality than does a structure with respect to its informational content; the belief has the informational content as its *exclusive content* while the information structure does not. The pieces of information embodied in a physical structure, although they qualify as propositional contents exhibiting (first-order) intentional characteristics, do not qualify as the semantic content which is characteristic of

belief. To occupy a belief state a system must discriminate among various pieces of information embodied in a physical structure and select *one* of these pieces for special treatment –as *the content* of that higher-order intentional state that is to be identified as the belief.

The source of the intentionality inherent in the transmission and receipt of information is rooted in the *nomic regularities*, on which the transmission of information depends, since lawful connection between two properties or magnitudes is an intentional phenomenon. Natural laws are essentially distinct by their *modal* quality (thus, they are indicative of what must be or what cannot be the case) which simple statements of exceptionless correlations do not have. The simple fact that all As are B (hence, exceptionless correlation) does not imply that any A *must* be B, and contrapositively, it does not imply that if this non-B were an A, it would also be a B, however, laws of nature has this property of *contrapositionality*.

The ordering of intentional states involves a production process of higher-order intentional structures from lower-order intentional states. The mechanism of this process is *digitalization*, which is grounded in the theory's analog and digital coding concepts of information.

The sort of information-theoretic construct that the third order of intentionality exhibits is that the semantic content of a signal can be identified with the information coded in digital form in contrast to analog form.

4.7. Analog and Digital Coding

Let us suppose that we have encountered a dog, which is a fact, the following are the two cases:

1. *We see the dog*; there is merely a sensory experience of an object (the dog). The information received through this experience is *extensional*; a concrete noun phrase stands as object of the verb.
2. *We see that it is a dog*; there is the recognition of that object as a dog (not a wolf or any other animal). The information is *intensional*; a factive nominal stands as complement of the verb.

The theory asserts that the distinction between (1) and (2) corresponds to the different ways *facts* can be represented. The passing from the sensory representation (the case (1)), to the cognitive representation (case (2)) is realized through *filtering out* such components of information that are superfluous for the purpose of selecting *one* component of this information. Dretske calls this process of filtering out irrelevant components of information

to have the one component as *digitalization*, in analogy with the analog-to-digital conversion of electrical signals. There is an information loss in this process, because we pass from a structure of richer informational content to one lesser information content.

Until information has been converted to the digital form, an information-processing system has failed to treat *different* things as essentially the *same*. It has failed to classify or categorize, failed to generalize, failed to “recognize” the input as being an instance (token) of a more general *type*.

A signal is said to carry the information that *s* is *F* in digital form if and only if the signal carries no additional (except for the nested) information about *s*. Otherwise, the signal is said to carry this information in analog form, which is a stage in the processing of the sensory information that is a preparation for the selective utilization by cognitive centers.

Let us consider a child's experience of the world and compare with an adult's. When a signal carries this information in analog form, the signal always carries more specific, more determinate, information about *s* than that it is *F*. Every signal carries information in both analog and digital form. The most specific piece of information it carries (about *s*) is the only piece of information the signal carries (about *s*) in digital form. All other information (about *s*) is coded in analog form.

The cognitive transformation of the perception in the analog form is associated with conceptualization (belief). The meanings as the conceptual contents of our internal states are developed out of informational contents.

4.8. Sensory versus Cognitive Processes

Information-theoretically, perception can be described as a process through which informational content, complexity of which depends the sensory capability of the receiver, is delivered in *analog* form to the cognitive centers for their selective use.

It is the successful conversion of information into appropriate digital form that constitutes the essence of cognitive activity. If the information that *s* is *F* is never converted from a sensory (analog) form to a cognitive (digital) form, the receiving system has not *seen that* it is *F* –does not identify that it is *F*- even if it has sensed an *s* being *F*. Therefore, the traditional idea that knowledge, belief, and thought involve *concepts* while sensory experience does not corresponds to this coding difference. Cognitive activity is the *conceptual* movement of incoming information, and this conceptual treatment is fundamentally a matter of ignoring differences as irrelevant to an underlying sameness, from the concrete to the abstract, from the particular to the general.

Then, we obtain a criterion to distinguish two types of information-processing system: Those that are capable of converting the information it receives into knowledge and those that are not.

4.9. Conception of Belief

The theory holds that a *concept* is an internal structure designed to carry or encode specific information in so-called completely digitalized form. The semantic content of a concept is identified with the information that the concept is designed to carry in such form. (*De re*) beliefs are tokens of such structures, deriving their semantic content from the concepts of which they are tokens. In other words, (primitive) *de re* beliefs are generated by instantiation of concepts that exercise control over the system output.

The presence of particular perceptual objects can trigger such tokens. When the object instantiates the relevant concept of which the triggered belief is a token, we have a true *de re* perceptual belief, otherwise, we obtain a false *de re* perceptual belief. False *de re* perceptual beliefs result when the mechanism which associates particular perceptual objects with concept token malfunction. Beliefs can *misrepresent* how things stand; the content of a belief, what we believe when we believe that something is so, can be either true or false.

Dretske makes a clear distinction between the concepts of meaning and information; there is no relationship of entailment between them, though information is a semantic concept.

For a difference between beliefs, a difference in meaning is clearly sufficient. On the other direction, for example as in the case of *de se* belief reports, such as “I am sick,” despite the fact that the sentence does not change meaning, the belief that someone expresses with this sentence is different from the belief that another one expresses with the same sentence, so a difference in meaning is not necessary.

Beliefs are internal representations that are associated with the intentional conduct of behavior, considered as the subjective component of knowledge.

4.10. Knowledge

Traditionally, knowledge is defined as justified true belief; what counts as knowledge is evaluated according to these three conditions, hence, belief, truth and justification. The information-theoretic analysis can be viewed as an attempt to overcome such problems as the Gettier paradox that this tripartite characterization of knowledge presents.

The analysis is restricted to perceptual (non-cognitive) knowledge of contingent states of affairs (conditions having an informational measure of something greater than 0) of *de re* form: Seeing (and knowing) that this (the perceptual object) is blue, moving, a dog, which is about an item *s* that is picked out or determined by factors other than what a cognitive agent *K* knows (or believes) about it. If *K* has a belief *about* this object, the belief that it is *F*, then this belief qualifies as knowledge if and only if that belief is caused (or causally sustained) by the information that it is *F*. The theory replaces the traditional view by the following ([13] p.86):

K knows that *s* is *F* = *K*'s belief that *s* is *F* is caused (or causally sustained) by the information that *s* is *F*.

It is important to notice that this is *not* an inferential account of perceptual knowledge. If *K* sees *s* moving and, as a result, comes to the belief that *s* is moving, we need not suppose that this belief about *s* was reached via some inference. The belief has no *discursive* derivation, and in this respect, it is acquired directly and without inference.

The analysis may appear circular: Knowledge is identified with information-sustained (or produced) belief, but the information a subject receives is relative to what it *already knows* about the possibilities at the source. Dretske holds that this objection overlooks the *recursive* character of the equation. Whether *K* can learn that *s* is *F* may depend on what *else* he knows about *s*, but not on its knowing that *s* is *F*.

K's belief that *s* is *F* is causally sustained by the information that *s* is *F* if and only if this piece of information affects the belief in such a way that it would, in the absence of other contributory causes, suffice for the existence of the belief. The idea of information that causes, in other words, causally sustains, a belief is intended for one's belief to qualify as knowledge, there must not only *be* evidence to support it, but the belief must be *based on* that evidence. Clearly, a belief that is caused by the information that *s* is *F* may not itself contain the information that *s* is *F*.

This definition does not deny that there are physical properties of the signal that cause the belief, however, which of these properties leads to the belief as its effect to be qualified as knowledge. If the belief that a friend has arrived is caused by the knocks on the door of which pattern is irrelevant, then he does not know that a friend has arrived; someone else could be knocking at the door, even if one might be caused to believe it by the knocks. Because those properties of the signal that carry the information (that a friend is at the door) are not the ones that are the cause of the resultant belief.

4.11. Against the Skeptical Thesis

Against the whole semantic theory of information, the skeptical thesis can be surely raised: If sensory experience depends not only on the character of the objects perceived, but also on the intervening conditions, such as, of sensory mechanisms, of the environment, then the sensory experience cannot be sufficiently reliable on the properties of the objects perceived. Hence, the sensory state can be equivocal by external conditions, even when the cognitive agent has the proper abilities.

Translation of the skeptical thesis into the an information-theoretic context is the claim that for purposes of determining the equivocation between source and receiver, amount of (and therefore what) information a signal carries about a source, no actual condition (unless known to obtain on independent grounds) can be taken as fixed, hence any possibility that can be consistently imaginable is a potential source of equivocation.

The skeptical view demands the approach that seeks the satisfaction of uncertainty in such cases as information-carrying signals were self-authenticating, which would lead to an infinite progression. Hence, in order to carry the information that *s* is *F*, a signal must also bear witness to the fact that what it carries is information, and that it (the signal) depends (in the requisite, unequivocal way) on conditions at the source.

The skeptic point is that an information-carrying signal typically depends, not only on the source, but also on a variety of other circumstances of which we are (normally) ignorant. The recommendation is that we treat this fact as showing that the signal does not *really* carry the kind of information about the source that it normally taken to carry. The discovery does not sustain the recommendation because some existing conditions (on which the signal depends) generate no information, or no *new* information, for the signal to carry. This is what constitutes the channel of communication ([13] p.115):

The channel of communication = that set of existing conditions (on which the signal depends) that either

1. generate no (relevant) information, or
2. generate only redundant information (from the point of view of the receiver).

This is why a signal, though depending on conditions both at the source and on the channel, carries information about *only* the *source*. Among the conditions on which the signal depends, *only the source* is a generator of new information.

The source, as a generator of information, is the locus of *relevant possibilities* that the signal is called upon to eliminate. The channel, being that set of conditions (on which the signal depends) that have no relevant alternative states –thus generating no information–

constitutes the fixed framework within which dependency relations between source and receiver are determined.

It is a fact that no signal can eliminate all the possibilities of what is consistently imaginable, such as the accidentally changing of voltage by an external factor in a transmission link, however, causing again meaningful data. We should make a distinction between what can be identified as possibilities and irrelevant contingencies that make *every* signal equivocal.

The communication theory avoids this result is through its statistical averages of events over a sufficiently long period as an index to what is possible. This method is fallible, since some possibilities that are relevant may fail to occur in the sampling for the averages, however, if every logical possibility would be a possibility, then everything would be noise; nothing could be communicated. The framework of fixed, stable, enduring conditions within which the degree of co-variation in other conditions can be evaluated (the flow of information) is *channel conditions*.

This relativity conforms to the alleged absoluteness of knowledge or the information on which it depends. To know, or to have received information, is to have eliminated *all relevant possibilities*. These concepts are absolute; what is not absolute is the way they are applied to concrete situations – to determine what will qualify as a relevant alternative.

This discussion of information presupposes a stable, permanent, regular, world that allows assessing the co-variation in some part with respect to fixed part. What may be taken as the fixed part to qualify as a channel condition is a pragmatic element raising some arbitrariness in the discussion, however, this is an element that resides in all our cognitive existence; it is not an objection to regarding the latter as fundamentally information-dependent notions.

The question of when a channel is reliable enough to transmit information of the sort required for knowledge (zero equivocation) is an empirical question.

Factual knowledge, the knowledge *that s is F*, is absolute in the sense that it does not admit comparisons; one either knows that *s is F*, or not. The same point is valid for justification, too.

Dretske asserts the view that the information requirement on knowledge faces the confusion between:

1. The information (about a source) a signal carries, and
2. The channel on which the delivery of this information depends.

If one supposes that a signal's dependence on a source is less than optimal insofar as it is conditioned by factors about which the signal carries no information, one will regard *all*

communication as impossible –hence, as impossible the knowledge that depends on the transmission of information.

The pragmatic aspects of knowledge are compatible with an information-theoretic analysis. Whether or not a signal carries some definite information depends on the channel conditions between source and receiver. The question of whether an existing condition is stable or permanent enough to qualify as part of the channel, as a condition which itself generates no new information, is a question of relativity that may not have an objective resolution. When a possibility becomes a *relevant* possibility is an issue that is to be decided with respect to the variable factors of evaluation that the participating elements in the communication process hold. The main constraint is that knowledge is possible, and so is information, which induces us away from the skeptical understanding that information requires 0 equivocation.

In the same manner, learning is described as a process of acquiring the ability to extract information from the sensory representation. Thus, we can have sensory data, but we cannot develop belief until learning is completed.

4.12. A Conflation in Dretske's Model

A fuzziness that occurs in Dretske's model of information flow is the conflation of particular events and types of events. Dretske makes a straightforward use of the concept of 'event' in probability theory. However, in probability theory, an event generally refers to a class of events. For example, flipping a coin and obtaining heads twice in consecution is a probability-theoretical event, which classifies any two flipping of coin, whereas the number of particular instances of flipping, in principle, may go to infinity.

Let us illustrate the consequence of this conflation by Wobcke's analysis of transitivity of information flow [31]. Wobcke sets Dretske's standard "doorbell" example into the following scheme:

- A: the doorbell is short-circuited
- B: the doorbell is ringing
- C: someone is at the door

Then, A carries the information that B, and B carries the information of C, but A does not carry the information that C.

It appears as if the transitivity stated as the Xerox principle is violated. However, as Wobcke remarks, the Xerox principle is held at the level of the particulars (signal tokens), and this is not a counterexample to the Xerox principle, since $A \Rightarrow B \Rightarrow C$ fails because of

the background condition on the channel that $\neg A$. But it is an example for the indiscrimination between a situation being of a type and a situation carrying information about another situation.

This point is the basis for Barwise's handling of Dretske's model in two distinct cases as one-event and multiple-event in terms of channels, which will be examined in the Barwise's information channels modeling.



CHAPTER 5

SITUATION THEORY

5.1. Framework of Situation Theory

Situation theory is a prominent attempt, first started by Jon Barwise in the late 1970s to provide semantics for perceptual reports. Barwise's idea, apparently inspired by model theory, was that the world is composed of situations in which individuals stand in relations together with relevant events, actions, locations, in general terms, *state of affairs*. In the early 1980s, by the joint efforts mainly of John Perry and John Etchemendy, beside Barwise, this idea developed into situation theory, providing rigorous basis that will be applicable in many areas of philosophy, mathematics, and computer science, semantics of natural languages, semantics of computer languages, the nature of informational content, the nature of computational processes.

One of the central themes of situation theory is that a theory of meaning and reference should be set within a general theory of information. Situation theory is an attempt in this context to develop a mathematical theory of meaning and information content. Application of situation theory to linguistic issues has given rise to *situation semantics*, which aims at the construction of a unified and mathematically rigorous theory of meaning, and providing a non-reductionist treatment of the various kinds of entities, such as, of states of affairs and facts, conditions, situations, propositions. It has an *information-based* approach to semantics in general; situated activities convey information.

Situation semantics is built upon simple facts. For example, it takes into account the observation that a statement can convey different information to a receiving subject when uttered in situations with regard to such parameters as time, location, utterer, and brings up the contextuality in its analyses.

Mathematical and logical issues that arise within situation theory and situation semantics have been explored widely. The theory has assembled its mathematical foundations based on intuitions basically coming from set theory and logic.

Situation theory leaves the idea of truth-values as the denotations of statements and aims at, in Keith Devlin's words, "a logic based not on truth but on information." Such logic can be an extension of first-order logic rather than being an alternative to it.

Information-based approach, which is essential in situation theory, is promising with its opening of wide-ranging views not only to traditional semantic issues, but also to formal systems. Actually, the same motives are used also in situation-theoretic view of mathematics, such as the role of proofs in extracting implicit information from explicit information, and logically equivalent statements can convey different information.

Then, a theory of situations would study and compare various types of situations or situation-like entities, such as facts, events and scenes with a non-holistic attitude.

In the following, the version of situation theory compiled and extended by Devlin will taken as the referential source since it is almost standard in this respect. [2], [6] and [7] are also used as the main sources for the subject.

5.2. Ontology of Situation Theory

In general, situation theory has a localist approach motivated by the determinations such as the context-dependence of our linguistic acts and partial data of perception. This approach is reflected in viewing the world as composed of *situations*. The notion of situation is closely associated with the ordinary meaning of the word; particulars as objects, sets of objects, events, activities, spatio-temporal locations, properties and relations compose situations and take place in situations, hence, they acquire meaning by being *situated*.

Situation theory develops a structured description of situations. There are *cognitive agents* (organisms, robots, computer programs, ...) that are capable of acquiring information from some source, process it, and act accordingly, for example, in order to modify its environment. Agents can be as simple as a thermostat, and as complex as human being; they are what we designate by *objective systems* in the general context.

Depending upon the agent and its *state*, different information is acquired and is made available for semantic processing. The agent creates new situations to arrive at new information and conveys to other agents the information that it already has.

A fundamental cognitive ability of a cognitive agent is *individuation*, the facility to discriminate *its* objects in a given environment. Metaphorically, the agent *carves out* situations out of the world on which applies a scheme of individuation. *Individuals* are

invariants; having properties and standing in relations, they persist in time and space. For example, words are invariants across statements.

One of the objectives for situation theory is modeling individuals by mathematical and logical methods. If a set-theoretic development is followed, there is a rich structure of abstract objects available; any collection of pure sets may be taken to represent the individuals. Alternatively, individuals may be taken as a collection of *ur-elements* or *atoms* that form part of the ground level of some set theory with ur-elements. But situations, in general, should be understood as extensionally defined objects.

Within individuation, the agent is able to recognize that various *properties*, in general, *relations* hold or fail to hold of some or all of those individuals.

Situation theory has a representation in terms of discrete items of information that depict a situation. These items, in a deliberate resemblance with such physical particles as electrons, photons, etc., are called *infons*. An infon has the representation

$$\langle\langle R, a_1, \dots, a_n, i \rangle\rangle$$

where R is an n -place relation on the objects $a_1, \dots, a_n$, and i is polarity of infon, which indicates whether that objects $a_1, \dots, a_n$ stand in the relation R (positive polarity) or not (negative polarity) in the situation s .

Therefore, a situation can be taken as a set of infons, which are *issues* with an assigned polarity (contrariety). An issue is some basic property of an object or relation between objects that can either hold or not hold at some situation, such that

$$\langle\langle R, a_1, \dots, a_n; 1 \rangle\rangle \text{ for the positive case, and}$$

$$\langle\langle R, a_1, \dots, a_n; 0 \rangle\rangle \text{ for the negative case.}$$

The arguments of $R, a_1, \dots, a_n$, is called a *constituent sequence*. Situations in which a sequence is assigned both truth-values are called *incoherent*. For instance, a situation s is incoherent if, for the definite individuals, a paramedic and a patient, at time τ

$$\langle\langle \text{injects, paramedic, insulin, patient, } \tau, 1 \rangle\rangle \in s \text{ and}$$

$$\langle\langle \text{injects, paramedic, insulin, patient, } \tau, 0 \rangle\rangle \in s.$$

It should be remarked that infons are mathematical constructs for semantic objects in situation theory, neither syntactic representations nor physically existing. An infon that corresponds to the *state of affairs*, that is, the way things actually are in the world is called a *fact*. Then, there are two groups of state of affairs with respect to a relation R and an appropriate constituent sequence, a situation s *supports* (or *makes factual*), that is, at most one of which is factual. The *supports* relation is denoted by ' $\models$ '. Thus, given a situation s and an infon σ such that

$$\sigma = \langle \langle R, a_1, \dots, a_n; i \rangle \rangle,$$

and that σ is made factual by s is shown as $s \models \sigma$.

Infons may involve parameters; such infons are called *parametric infons*. For example, $\langle \langle \text{injects}, X, \text{insulin}, Y, \tau, 0 \rangle \rangle$ is a parametric infon where X and Y are parameters over individuals on the first and the second arguments of the relation *injects*.

Parameters provide a method of abstraction, as they are generalizations over classes of all first-class entities of situation theory; that is, objects, situations, individuals, spatial locations, etc. The parametric objects actually define types of objects in that class.

Parametric infons can also be allowed to be undetermined with respect to relation and polarity, for example, $\langle \langle R, X, Y; i \rangle \rangle$ where R and i are parameters over relations and polarity, respectively. *Parameter-free infons* are the basic items of information about the world while parametric infons are the basic units that are utilized in the treatment of *abstract situations*.

Abstract situations are set-theoretic structures consisting of a set of parametric infons. Such situations can be viewed as models of concrete (real) situations, satisfying some of the properties of concrete situations, and which may or may not be supported by any concrete situation. For each concrete situation, there is an abstract situation that corresponds it. But there are also abstract situations that do not correspond to any concrete situations.

There may be also incoherent abstract situations (that is, situations s such that $s \models \sigma$ and $s \models \sigma'$ where σ and σ' differ only in that one has the polarity 1 and the other has the polarity 0). An abstract situation is said to be *actual* if the information supported by that situation conforms to the real world. Thus, any actual abstract situation satisfies coherence.

Information can be partitioned into situations by defining a hierarchy between situations. A situation can be larger, having other situations as its parts. Being in this larger situation gives the ability of having information about its subsituations, which can be used to build hierarchies among situations, conforming to the notion of nested information.

Since situations constitute a world, whether actual or not, and that world can have parts with different information content, the domain of situations is partially ordered (i.e., reflexive, anti-symmetric and transitive in the usual sense) by a *part-of* relation ' $\leq$ '.

A piece of information is *persistent* if and only if, whenever it holds at a situation s , it holds at all situations of which s is a part. Restricted to abstract situations, the assumption that all information is persistent allows treating $\leq$ as if it were the subset relation on abstract situations.

The recursive nature of the various definitions may lead a parameter to occur at one or more levels inside an infon or a type. Thus, the notion of occurrence of a parameter in an infon or a type satisfies *hereditariness*.

A method of referring used in situation semantics is *anchoring*. An anchor for a set A of basic parameters is a function defined on A , which assigns to each parameter a_n in A an object of type T . Thus, if f is an anchor for A , and a_n is a parameter in A , then the infon $\langle\langle$ of-type, $f(a_n)$, T , 1 $\rangle\rangle$ is a fact.

5.3. The Notion of Constraint

The acquisition of information from a situation depends upon those *constraints* to which the agent is *attuned*. The situation theoretical notion of constraints is in accordance with Dretske's theory of information. A constraint of the form "A *involves* B" asserts that a signal A carries the information B. The facts of a situation carry information. Moreover, the information carried by a fact is so carried by virtue of an *informational constraint*.

Constraints have a Humean character in that any intrinsic connections between events composing a situation are not required; information carried does not exist due to the intrinsic properties of the events.

The main forms of constraints can be identified as:

1. *Necessary constraints*, which are analytic, e.g., "every square is a rectangle."
2. *Nomic constraints* provide patterns in the way of natural laws.
3. *Conventional constraints* are those arising out of explicit or implicit conventions that hold within a population of agents, e.g., "the first day of a new year is the first of January." Since they are neither nomic nor necessary, they can be violated.
4. *Reflexive constraints*, which allow extraction of more information from the same situation.

Constraints enable one situation to provide information about another, thus can be used as inference rules. When viewed as a backward-chaining rule, a constraint can provide a channel for information flow between types of situations, from the antecedent to the consequent. This means that such a constraint behaves as some definition for its consequent part. Another way of viewing a constraint is as a forward-chaining rule, which enables an agent to alter its environment.

Attunement to a constraint, understood as being aware of and systematically acting in accordance with that constraint, is what enables an agent to acquire and use that information. Attunement is an essentially empirical phenomenon. Different constraints will give rise to information being picked up; given the possession of prior information and/or attunement to other constraints, the acquisition by an agent an item of information can also provide the agent with an additional item of information. For example, when it is said that a certain geometric figure is a square, then one also gains the information that the figure is a rectangle, all its internal angles are $\pi/2$, and so on. Each of these additional items of information is *nested* (*structured* information by way of certain constraints) in the information that the figure is a square.

5.4. Situation-Types

In order to achieve a structured view, higher-order uniformities, called *situation-types*, between situations are discriminated. These are relatively more general and static situations, actually making information flow possible; for example, two events, occurring at the same location but at different times, has the same *situation type s*, which persists in both situations. Type is what different parts of the world have in common when they resolve (decide) the same issues in the same way except for the difference in some common parameters. Like situations, situation types do not necessarily resolve every issue. Constraints serve as links between the types of situations. Basic situation types identified by Devlin are as follows:

1. TIM: the type of a temporal location,
2. LOC: the type of a spatial location,
3. IND: the type of an individual,
4. REL n : the type of an n -place relation,
5. SIT: the type of a situation,
6. INF: the type of an infon,
7. TYP: the type of a type,
8. PAR: the type of a parameter,
9. POL: the type of a polarity.

T.C. YÖKTEKİN KURULU
DENEYİM KURULU
DENEYİM KURULU

For each object x , there is at least one type such that x is of that type. For instance, if l is of type LOC, then the infon is $\langle\langle\text{of-type}, l, \text{LOC}; 1\rangle\rangle$. Situation types can be also regarded as partial functions from relations and objects to the polarities.

Being in a situation, one can derive information about other situations connected to it in some way. For example, from a statement situation it is possible to obtain information about the situation it describes. Accessing information through a structure of situations and explicit relationships among them requires a logical mechanism. This mechanism will put information about situation types related in some way into reach of the agent and can be made possible by a proper implementation of the *supports* relation.

Typing provides a *situational abstraction*; an infon ϕ partitions situations into types such that $[\acute{s} \mid s \models \phi]$ is a situation type, the notation $\acute{s}$ denoting parameters, which is the type of situations in which ϕ holds, denoted by $s : \phi$. Let us consider the following example:

The fact that the abnormal amount of leucocytes in the patient's blood indicates that has an infection.

The constraint that connects the fact that the number of leucocytes has such an amount and the information that the patient has an infection is the fact that, if a hematological test is done on the patient's blood and the leucocytes counted has that amount, then the patient has an infection. The laws of nature that govern the hematological tests, the nature of human blood, and so on provides this constraint.

Let us apply the discussion on constraints on the above infon in the form

$$\langle\langle\text{Involves}, \sigma, \psi, 1\rangle\rangle$$

where σ (increase in the number of leucocytes), ψ (occurrence of an infection) are situation types, and *Involves* is the constraint. The infon is interpreted as providing the information that “for every situation of type σ , there is one of type ψ .” Thus, if there is a situation in which the infon holds and there is a concrete situation of type σ , then there is also a concrete situation of type ψ .

5.5. Statements, Propositions and Infons

Situation theory makes a distinction between statements and propositions: Statements are declarative sentences by which a speaker asserts something, whereas propositions are nonlinguistic abstract objects that have claims about the world. Truth is fundamentally a property of propositions and derivatively a property of statements.

As to the nature of propositions, situation theory holds the Austinian view with the distinction between *demonstrative* and *descriptive conventions* of language, rather than the Russellian (for an extensive discussion of this point see [6]). In situation-theoretic terms, demonstrative conventions roughly correspond to situations, for example, being in such a situation that snow is encountered for the first time and stating that “snow is white,” “snow melts when held,” etc., and descriptive conventions to situation-types, in which case “snow is white” is stated in contrast to “snow is black,” “snow is purple.”

A situation and a type determine an Austinian proposition, whereas a type alone determines a Russellian proposition. Thus, an Austinian proposition is true if *the* situation is of the type; a Russellian proposition is true if *some* situation is of that type.

For example, the statement that refers to a card game being played at a specified location “*A* has the card *x*” is true in Russellian view, even if *A* is not at that location, but somehow, has the card *x*, for instance, playing at another location at the same time, whereas it bears no relevancy in the Austinian view.

A proposition is the whole of $s \models \sigma$, where σ is an infon.

5.6. Partiality

A crucial assertion of situation theory is partiality of situations and situation types; situation theory is not committed to a largest total situation of which the rest are parts. This realistic assertion provides a notion of *relevance* in the logical sense.

Partiality should be understood in two distinct senses:

1. A situation does not necessarily decide every fact.
2. Situation types may contain objects and event types by abstraction.

Classical (holistic) semantic approaches assert that the statements that are true in the same models has the same information content disregarding *subject matter* and *partiality of information*. On the other side, situation semantics holds that logically equivalent sentences may have different subject matter, describing different situations involving different object and properties.

The notion of partial models incorporates an improved notion of information content and an enhanced relevance notion of logical consequence.

A consequence of partiality in situation theory is the forking of the negative case, unlike classical logic, as *negation* and *denial*:

- *Negation*: Situation s is of type where the proposition is *not* true.
- *Denial*: Situation s is *not* of type where the proposition is true.

Therefore, given a state of affairs and its negative case, the following alternatives are admissible for situation theory:

- Some situation will make one of them factual,
- No situation will reject the other one,
- Some situations will leave the issue unresolved, making neither of them factual.

The conceptual distinction negation/denial together with statement/proposition play an essential role at many issues where classical semantics considers paradoxical.

5.7. Situation Semantics

Situation semantics is supposed to be able to represent, for example, the information that is available to one in perception. such information is partial, we do not perceive the whole world, but only a part of it. Situation semantics takes this as a principle in its analyses, not a weakness to overcome.

Situation theory provides a semantic theory as an appropriate framework for a realistic model-theoretic semantics of language, considering natural languages as a special case. Situation semantics is capable of handling various linguistic issues to deal with the multiplicity of linguistic elements, such as indexicals, demonstratives, referential uses of definite descriptions, deictic uses of pronouns, tense markers, and names.

The fundamental status given situation semantics to contextual interpretation makes the theory remarkably compatible with speech act theory and discourse pragmatics. The essential assertions of situation semantics can be identified as follows:

- *Context-dependence*.

As a situation-theoretic consequence, the linguistic elements depend heavily on context for their interpretation and are fundamental to the way language conveys information.

The statement of an expression E in a context u is regarded as a flow of information generated by both the linguistic form encoded in E and the contextual factors in the

statement situation u . These are combined to form a set of constraints on the described situation s . Basically, situation semantics tries to look at the meaning of a sentence as providing a constraint between two proper parts of the world: the statement and the described situation.

Context supports not only facts about the discourse-participants, such as speakers, addressees, etc., but also facts about their relations to other contextually relevant situations. Meanings of expressions are thought to reside in systematic relations between different types of situations.

The speaker refers to an individual using a name or a pronoun. Such a situation s , called an *anchoring situation*, can be seen as a partial function from the referring words to their referent, hence the individual. This function is the speaker's connections for a particular statement.

- *Efficiency of language.*

A sentence can be used in many different situations to state different things. Therefore, its interpretation, which is the class of situations described by the statement, is subordinate on the situation in which the statement is used. Since agent (source of information) are always in different situations, having different causal connections to the world and different information, the information conveyed by a statement will be relative to the agent's and the receivers' situation.

In contrast to classical semantics that disregards ambiguity and contextual elements of language as anomalies, situation semantics seeks to handle them in a formal way and to explore how they can contribute to achieving a comprehensive theory of linguistic meaning.

- *Productivity of language.*

The mechanism of *compositionality* on given a finite vocabulary provides a potentially endless list of meaningful expressions that we can form and understand even if never before encountered.

- *Partiality of information.*

This feature of situation semantics follows directly from the situation-theoretical view of the world.

- *External significance of language.*

To each situation, some informational content conveyed linguistically corresponds; the situations are related to each other by the information flow.

- *Mental significance of language.*

Situation semantics is intended to give an account of mental states, not directly by referring to mental contents, such as ideas, senses, thoughts, but indirectly by referring to situations that are extra-mental.

5.8. Situation-Theoretic View of Conditionals

A situation-theoretic analysis of conditional statements is given in [5]. According to this analysis, conditional statements are in analogy with the information flow and can be modeled by *conditional* constraints, which demands situations to satisfy some condition (in contrast to *unconditional* which are applicable to all situations). A conditional constraint is formulized as $S \Rightarrow S'|B$, where S and S' are types of situations, $\Rightarrow$ is the “involves” relation and B is the *background conditions* on the situations.

The situations must be of type B if the constraint $S \Rightarrow S'$ is to hold. For example, the constraint “matches struck will ignite” holds relative to the background condition that there is oxygen in the environment.

If all of the arising situations in a given environment are of type B , then B turns out to be *environmental constants* for which there is no reason to be aware of. Therefore, whether background conditions are environmental constants implicit in the constraints or not is a situated issue to be resolved. An account of information content conveyed by a statement needs comprehension of the situations it can describe and the background conditions on this description.

By this analysis of conditionals, situation theory departs from the classical truth conditions (as in material implication) to the information conditions, which comes out of considering subject matter of a statement. This enforcement is stronger than the truth conditions on the conditionals and has affinity with the relevant logics. Background conditions are represented by *channel conditions* in formal modeling of information flow.

5.9. Comparison of Possible Worlds Semantics and Situation Semantics

Possible worlds semantics and situation semantics share the view that understanding the assertive use of declarative sentences is basic to understanding their meaning. The basic assumptions embodied in possible worlds semantics as semantics for modal operators and for natural language can be identified as:

- The primitives of the theory are individual particular objects and possible worlds. Properties and relations of these objects are represented through set theory.
- There is an actual world w_0 , which determines the basic facts, settling basic issues as to what properties various objects (sequences of objects) actually have or do not have in a space of possibilities.

- There is a set W of all possible worlds, with $w_0 \in W$. The worlds $w \in W$ other than w_0 correspond to other ways these basic issues in the space of possibilities might have been.
- The fundamental semantic value to be associated with a declarative sentence S is proposition p_S . Propositions can be identified with functions from possible worlds to truth-values; a proposition p is true just in case $p(w_0) = \text{true}$.

In contrast, situation semantics presuppose:

- The primitives of the theory are individual particulars, properties and relations. Thus properties and relations are not represented set-theoretically, but are first-order constituents.
- There is a single world w_0 , which is the real world that determines the basic facts as to what particulars stand in what relations.
- Valid assertions are about the real world or its parts. These parts are situations, not worlds which the total ways the real world might be. Each particular situation s corresponds to a particular set of facts that hold in s . Hence, a situation may leave some issues unresolved, whereas every proposition must be assigned a truth value in every possible world.
- There are properties and relations among all the first-order constituents including situations. Since the types of situations are not extensional, two distinct types of situations may not be types of distinct real situations; also every type of situation may not be the type of some real situation.
- The fundamental semantic value to be associated with declarative sentence S is its meaning, a relation $U_S : D_S$ between the type U_S of situation where S is assertively stated and the type D_S of situation by which it is described. A valid assertive statement u of type U_S is true if the situation s_u that u is about is of type D_S . The information content of the statement u is that s_u is of type D_S .

Let us consider a mathematical case in which a partial function from A to B will have properties not possessed by any of its total extensions. For example, consider the real function f defined by

$$f(x) = x(x+1)/x$$

for $x \neq 0$, $f(0)$ undefined. There is a single total function g extending f so that g is a continuous function, in which $g(0) = 1$. In analogy with this mathematical case, situation

semantics claims that the regularities in the space of possible ways the world could be, do not entail that a part of the world is a plain extension of the totality. In other words, situation semantics supports a *localist* view of the world in contrast to a *globalist* one.

Indeed, possible worlds semantics and situation semantics are not comparable, but rather, they are two different attitudes toward doing semantics.

5.10. The Issue of Well-Foundedness

A consequence of the non-reductionist ontology of situation theory is the view that situations, thus reality, are *not* well-founded. Situation theory defines the following foundational schemes:

A binary relation R is *well-founded* if whenever a property P holds of some x in the range R , there is something x that has P but such that no y with yRx has property P . That reality is well-founded means that the constituent relation $\in$ on situations well-founded.

A situation s is *circular* if there is a finite sequence of length $n \geq 2$ of situations, starting and ending with s , each of which is a constituent of the next:

$$s \in \dots \in s.$$

A situation s is *ungroundable* if there is an infinite sequence of distinct situations, starting with s , each a constituent of the next:

$$\dots \in s''' \in s'' \in s' \in s.$$

Situation s is called well-founded if it is neither circular nor ungroundable. Then reality is well-founded if and only if every situation is well-founded. Nevertheless, when someone makes a reflexive statement with respect to the situation, such as “this is a complex situation,” the statement itself becomes a part of the situation. Situation theory does not qualify such situations as paradoxical, rather accepts it a part of its view.

Upon this understanding, non-well-founded sets are widely used in situation theory as an alternative to Zermelo-Fraenkel set theory with axiom of choice (ZFC), which presupposes:

- A set is any family that can be comprehended as a totality.
- Sets are built up in a cumulative hierarchy. It is this assumption that justifies the axiom of foundation in ZFC, the axiom asserts that, if $P(a)$ holds for some set a , then there is a set b such that $P(b)$ holds, but there is no set $b' \in b$ such that $P(b')$.

The idea is that choose one b that the primordial in the construction of all sets holding property P . Then none of the members of b can be sets with the property P .

The main idea in the non-well-founded sets is that set construction is a process, in which the set itself can be an element of itself, a feature well suited to model such phenomena as self-referential. As to the purposes of this study, an intuitive understanding of sets is sufficient.



CHAPTER 6

A SITUATION-THEORETIC MODELING OF INFORMATION FLOW

In the following part, an exposition of formal model of information flow derived from the combination of Jerry Seligman's model based on his concept of *perspectives* given in ([28]) and Jon Barwise's channel model ([4]). Both of them are construe systematic relations among situations in the framework of situation theory backed up by Dretske's conceptual apparatus. The formal analyses given by the mentioned authors are quite straightforward, so we will not need to modify them for some reason.

6.1. Situations in a Perspective

Situations are typed with respect to the information they support. Types provide us a scheme of collecting situations that enables us to make explicit the relations between them. This is particularly suitable to model the flow of information from the source to the receiver, since both of them must support the same information, therefore, must be of the same type. A concept to collect situations to construct the relations between them is *perspectives*. A perspective is expected to exhibit, particularly,

- which situations are of the same type, and
- what law-like relations exist between those types.

If a situation s supports the information σ , then it is of type $[x \mid x \models \sigma]$. Logical compounds of various kinds, both at the level of information and at the level of types, provide a classification of situations ultimately based on the scheme of individuation of a system.

In accordance with a localist attitude to law-like dependencies, perspectives should be thought of as idealizations of particular parts of the world rather than theories of the whole world. Associating a set of situations with each perspective, called its domain, captures this.

The situations in a perspective's domain should be understood as unstructured particulars, rather than sets of infons or some similar structure; any structure they may be understood to have must arise from a perspective on them.

6.2. Information Flow

The model uses two constraints represent information flow:

- the positive $\Rightarrow$: 'involves' relation,
- the negative $\perp$: 'precludes' relation.

The negative $\perp$ constraint is needed because types have no internal structure (in contrast to "proposition") and so negative informational content, which provides information growth as well as its positive counterpart, must be externally supplied.

Information flows from source to receiver iff the signal to the receiver carries information about the source's state, in virtue of certain probabilistic dependencies between their state descriptions. Both the source and the receiver are specified by mutually exclusive and exhaustive sets of states. Associated with each state is probability of the source (or receiver) being in that state. A particular signal is specified by the receiver being in a particular state. The quantity of information conveyed by a signal is measured by the extent to which it constrains which state the source is in.

Let us recall that, according to Dretske, a signal is said to carry the information that p iff the conditional probability of p given the signal is 1 and less than 1 otherwise. $\phi \rightarrow \psi$ should mean that the conditional probability of ψ , given ϕ , is 1; $P(\psi | \phi) = 1$, where $P(\psi) \neq 1$.

Dretske's study is confined to the propositions of the form " s is F ", whereas in the present model, arbitrary propositions p about the source is considered. Also Dretske's relativization of the content of a signal to prior knowledge is placed into the specification of the system (as a part of the *channel conditions*).

6.3. Principles of Information Flow

Upon the discussion on the nature of information and the situation theory, the model is expected to satisfy a group of principles.

Principle 1 (Facticity). Situations are first-order objects, if there is a situation type, then there must be (at least) one actually existing situation that induces that type, hence, “involves” relation $\Rightarrow$ is fact preserving:

$$\text{If } s:t \text{ and } t \Rightarrow t' \text{ then } s':t'$$

where s' is any situation. There is also *strong facticity* in which s' is necessarily s :

$$\text{If } s:t \text{ and } t \Rightarrow t' \text{ then } s:t'.$$

Both versions of facticity will be used in the modeling, however, it should be noted that weaker version is more in line with the characteristics of situation theory, whereas stronger version is better at modeling logic as information flow, in which case situations represent all possibilities.

Principle 2 (Xerox Principle). The “involves” relation $\Rightarrow$ satisfies the Xerox principle: If $s_1 : \phi$ carries the information that $s_2 : \psi$ and $s_2 : \psi$ carries the information that $s_3 : \theta$ then $s_1 : \phi$ carries the information the information that $s_3 : \theta$. In symbols: if $s_1 : \phi \Rightarrow s_2 : \psi$ and $s_2 : \psi \Rightarrow s_3 : \theta$ then $s_1 : \phi \Rightarrow s_3 : \theta$.

Principle 3 (Addition of Information). Assuming that the particular types ϕ and ϕ' can be conjoined:

If $s_1 : \phi$ carries the information that $s_2 : \psi$ and $s_1 : \phi'$ carries the information that $s_2 : \psi'$ then $s_1 : (\phi \wedge \phi')$ carries the information that $s_2 : (\psi \wedge \psi')$. In symbols: if $s_1 : \phi \Rightarrow s_2 : \psi$ and $s_1 : \phi' \Rightarrow s_2 : \psi'$ then $s_1 : (\phi \wedge \phi') \Rightarrow s_2 : (\psi \wedge \psi')$.

Principle 4 (Exhaustive Cases). Suppose that

$$s_1 : \phi \Rightarrow s_2 : (\psi \wedge \psi'),$$

$$s_2 : \psi \Rightarrow s_3 : \theta, \text{ and}$$

$$s_2 : \psi' \Rightarrow s_3 : \theta, \text{ then}$$

$$s_1 : \phi \Rightarrow s_3 : \theta.$$

The fifth principle, *contraposition*, is to be stated in the subjunctive, not in the indicative; since a counterfactual case is taken under consideration. In order to distinguish

this difference, the symbol $\Rightarrow$ is used to denote the indicative, while ' $\rightarrow$ ' is used to denote subjunctive. Then, the following definitions can be given:

- $s_1: \phi$ would carry the information s_2 was of type ψ relative to the channel c :

$$s_1: \phi \rightarrow_c s_2: \psi \text{ iff } c: \phi \rightarrow \psi \text{ and } s_1 \rightarrow_c s_2.$$

- $s_1: \phi$ carries the information that $s_2: \psi$ relative to the channel c :

$$s_1: \phi \Rightarrow_c s_2: \psi \text{ iff } s_1: \phi \rightarrow_c s_2: \psi \text{ and } s_1: \phi.$$

Principle 5 (Contraposition). If $s_1: \phi \rightarrow s_2: \psi$ then $s_2: \neg\psi \rightarrow s_1: \neg\phi$.

Thus, $s_1: \phi \Rightarrow s_2: \psi$ is equivalent to $s_1: \phi \rightarrow s_2: \psi$ and $s_1: \phi$.

Principle 6 (Local Preclusion). The “precludes” relation $\perp$ is (locally) fact inverting. If a situation s is of type t , and $t \perp t'$ then $\neg(s:t')$. This is a minimal assumption based on which $\perp$ can be a carrier of negative information. If in some situation a coin is flipped and lands heads, then it is precluded that the coin lands tails in the same situation. It is presumed that one of the two possibilities might be actualized in the situation, that is, the process of *refinement* has been completed *locally*.

On the other side, the *global preclusion* (if $s: t$ and $t \perp t'$ then there is no s' of type t') is too stronger condition for situation theory, since we can assume that types are fine-grained classifiers of situations, it is not possible to exhaust every situation, or possibilities in a situation.

Principle 7 (Mutual Preclusion). The “precludes” relation $\perp$ is symmetric. In combination with local preclusion, this is motivated by the idea that if t precludes t' , then t, t' are mutually exclusive ways of classifying situations.

6.4. The Source/Receiver Perspective

First, let us note Seligman’s formal definition of perspective:

Definition. A *perspective* is a structure $\langle S, T, :, \Rightarrow, \perp \rangle$ where $\langle S, T, : \rangle$ is a classification, ‘ $:$ ’, of a collection S (of situations) by a collection T (of types), and $\Rightarrow$ (*involves*) and $\perp$ (*precludes*) are binary relations on T , such that for all $s \in S, t \in T$, satisfying the above defined principles:

1. Facticity,
2. Xerox,

3. Local preclusion,
4. Mutual preclusion.

Given a particular perspective $P = \langle S, T, :, \Rightarrow, \perp \rangle$, there will be a situation domain S of P, S^P , and a type set T of P, T^P .

The model of communication is restricted to a single source, S , and receiver, R . The behavior of S and R is taken to be entirely specified by the states they can adopt, together with the probabilities of their being in each state. S_{st} and R_{st} are distinct, exhaustive and mutually exclusive sets of states for S and R respectively. At any given moment, the state of the system as a whole is represented by a pair, $\langle \sigma_1, \sigma_2 \rangle$, where $\sigma_1 \in S_{st}$ is the state of the source and $\sigma_2 \in R_{st}$ is the state of the receiver. Therefore, behavior of the system follows a trajectory in $S_{st} \times R_{st}$.

Although a temporal dimension is included in the model, only the present states that the systems S and R adopt will be considered. The behavior of the system as the future or past states depend on the character of the systems, whether they are with memory or memoryless, deterministic or not.

A probability measure P is defined over the state space of the system that satisfies the usual axioms of probability theory. The space of possible events, E , consists of those subsets of $S_{st} \times R_{st}$ with non-zero probability. The model considers the particular events of

1. “ S is in state σ_1 ,” given by $t_{\sigma_1} = \{ \langle \sigma_1, x \rangle \mid x \in R_{st} \}$,
2. “ R is in state σ_2 ,” given by $t_{\sigma_2} = \{ \langle \sigma_2, x \rangle \mid x \in S_{st} \}$.

for each $\sigma_1 \in S_{st}$ and $\sigma_2 \in R_{st}$, since these classify the states of S and R independently. Each t_{σ} has a non-zero probability and so is in E . The events t_{σ_1} and t_{σ_2} are mutually exclusive ($t_{\sigma_1} \cap t_{\sigma_2} = \emptyset$) when $\sigma_1 \neq \sigma_2$ and both states are of the same situation ($\sigma_1, \sigma_2 \in S_{st}$ or $\sigma_1, \sigma_2 \in R_{st}$), so the type sets are singletons, $\{t_{\sigma} \mid \sigma \in S_{st}\} = \{t_{\sigma S}\}$ and $\{t_{\sigma} \mid \sigma \in R_{st}\} = \{t_{\sigma R}\}$, that is, there is only a single event in each of them. Note that $\cup \{t_{\sigma} \mid \sigma \in S_{st} \times R_{st}\} = S_{st} \times R_{st}$

Let us assume a perspective on the system that is divided into two epistemic contexts at a time, τ . The situation at the source, s , and the situation at the receiver, r . They are classified independently by the relation ‘:’, according to the respective states of the source and the receiver at time τ , taken as the situations, s, r , and the events t_{σ} for $\sigma \in S_{st} \cup R_{st}$, defined by

1. $s : t_{\sigma}$ iff S is in state $\sigma \in S_{st}$
2. $r : t_{\sigma}$ iff R is in state $\sigma \in R_{st}$

The set of types in the classification is the set $T = \{t_\sigma \mid \sigma \in S_{st} \cup R_{st}\}$. The ordered pair $\langle \sigma_S, \sigma_R \rangle$ is the state of the system at time τ . Because of the facticity, *actuality property* is enforced on the model, that is:

$$P(\{\langle \sigma_S, \sigma_R \rangle\}) > 0$$

In this simplified model, the only constraints taken into account are positive and negative dependencies between types:

1. $t_{\sigma_1} \Rightarrow t_{\sigma_2}$ iff $P(t_{\sigma_2} \mid t_{\sigma_1}) = 1$ and $P(t_{\sigma_2}) < 1$
2. $t_{\sigma_1} \perp t_{\sigma_2}$ iff $P(t_{\sigma_2} \mid t_{\sigma_1}) = 1$ and $P(t_{\sigma_2}) > 0$

Both source and receiver are assumed to be non-constant, hence have more than one possible state. Although there is a notional distinction between source and receiver, the model does not represent this asymmetry. Therefore, when generalizing to a system with many sources and receivers, they can be referred to as “units.” The proofs can be extended to the n units case, D_n .

Proposition. For any states σ_1, σ_2

1. $t_{\sigma_1} \Rightarrow t_{\sigma_2}$ iff $P(t_{\sigma_2} \cap t_{\sigma_1}) = P(t_{\sigma_1})$
2. $t_{\sigma_1} \perp t_{\sigma_2}$ iff $P(t_{\sigma_2} \cap t_{\sigma_1}) = 0$

Proof. Each t_σ is in E , so $P(t_\sigma) > 0$. Suppose $P(t_\sigma) = 1$ for some $\sigma \in S_{st}$. Then, assuming the source is non-constant, there is a $\sigma' \in S_{st}$ such that $\sigma \neq \sigma'$. But then $t_\sigma \cap t_{\sigma'} = \emptyset$, so

$$1 \geq P(t_\sigma \cup t_{\sigma'}) = P(t_\sigma) + P(t_{\sigma'}) = 1 + P(t_{\sigma'}),$$

so $P(t_{\sigma'}) = 0$, contradicting the fact that $t_{\sigma'} \in E$. So for all $\sigma \in S_{st}$, $0 < P(t_\sigma) < 1$. The argument for $\sigma \in R_{st}$ is identical and the rest follows directly from the definition of conditional probability.

Taken together the above structure $\langle \{s, r\}, T, :, \Rightarrow, \perp \rangle$, called D_2 , constitutes a model of the flow of information of the form “ U is in state σ ” (where $U = S$ or R) between a single source and a single receiver.

Proposition. D_2 is a perspective.

1. $r : t_{\sigma R}, s : t_{\sigma S}$ and $t_{\sigma R}, t_{\sigma S}$ are the only grounded types. So the only cases to consider for *facticity* are when $t_{\sigma R} \Rightarrow t_\sigma$ or $t_{\sigma S} \Rightarrow t_\sigma$ for some arbitrary state σ . The argument is the same for both cases, so the first one will be examined. If $t_{\sigma R} \Rightarrow t_\sigma$ then $P(t_\sigma \cap t_{\sigma R}) = P(t_{\sigma R})$, and a situation of type t_σ must be showed to exist, which amounts to showing that either $\sigma = \sigma_S$ or $\sigma = \sigma_R$.

If $\sigma \in R_{st}$ then either $\sigma = \sigma_R$, as required, or $t_\sigma \cap t_{\sigma R} = \emptyset$, so that $P(t_{\sigma R}) = P(t_\sigma \cap t_{\sigma R}) = 0$, contradicting the fact that $t_{\sigma R} \in E$. If $\sigma \in S_{st}$ then $t_\sigma \cap t_{\sigma R} = \{\langle \sigma, \sigma_R \rangle\}$ so that $P(t_{\sigma R}) = P(\{\langle \sigma, \sigma_R \rangle\})$.

Now either $\sigma = \sigma_R$, as required, or $\{\langle \sigma, \sigma_R \rangle\}$ and $\{\langle \sigma_S, \sigma_R \rangle\}$ are mutually exclusive events, so that

$$P(\{\langle \sigma, \sigma_R \rangle\}) + P(\{\langle \sigma_S, \sigma_R \rangle\}) = P(\{\langle \sigma, \sigma_R \rangle, \langle \sigma_S, \sigma_R \rangle\})$$

$$\{\langle \sigma, \sigma_R \rangle, \langle \sigma_S, \sigma_R \rangle\} \subseteq t_{\sigma R}, \text{ we have } P(\{\langle \sigma, \sigma_R \rangle, \langle \sigma_S, \sigma_R \rangle\}) \leq P(t_{\sigma R}).$$

Hence,

$$P(\{\langle \sigma, \sigma_R \rangle\}) + P(\{\langle \sigma_S, \sigma_R \rangle\}) \leq P(\{\langle \sigma, \sigma_R \rangle, \langle \sigma_S, \sigma_R \rangle\})$$

therefore, $P(\{\langle \sigma_S, \sigma_R \rangle\}) = 0$, which contradicts the actuality property. Thus, the possibilities for σ are exhausted.

2. *Xerox Principle* is straightforward.

3. *Local Preclusion*. For each situation x , if $x:t_\sigma$ and $t_\sigma \perp t_{\sigma'}$ then $P(t_\sigma \cap t_{\sigma'}) = 0$. But $t_\sigma \in E$, so $P(t_\sigma) \neq 0$, so $t_\sigma \neq t_\sigma \cap t_{\sigma'}$, and so $\sigma \neq \sigma'$. But then $\neg(x:t_{\sigma'})$, since the type set of x is a singleton.

4. *Mutual Preclusion*. Follows from the symmetry property of $\perp$.

6.5. Information Channels

Barwise's information channel model is a formalization of information flow and inference. A heuristic example for a channel is an indicator of a physical quantity, such as

temperature, speed, or pressure, so let us proceed on the example of a thermometer as an indicator.

There are the height of the thermometer liquid situations h and the temperature situations t . Since the height indicates temperature (or from another point of view, “height is caused by temperature”), there is connection between these situation, which will be denoted by $h \mapsto t$ in the model. These situations turn out to be *tokens* that are classified under types, such as $h = x$ or $t = y$, which will be denoted by ϕ or ψ .

Also, there are a systematic relations between the height of the thermometer liquid situations h and the temperature situations t it indicates that ground many constraints in the conditional form of “if $h = x$, then $t = y$.” Note that each pair (x, y) should be understood as a different constraint. The constraints are relations between types of situations denoted by $\phi \rightarrow \psi$.

In situation-theoretic terms of the Austinian origin, the types and constraints provide the descriptive content, while the situations and channels provide the demonstrative content for the conditional statements (or, for any structure that can be represented in the conditional form).

Given these concepts, the model asserts this fundamental inference, where s_1 and s_2 are situations and ϕ and ψ are types:

$$\text{If } \phi \rightarrow \psi, s_1 \mapsto s_2 \text{ and } s_1 \models \phi, \text{ then } s_2 \models \psi.$$

Based on the framework summarized in the preceding expression, the channel model aims at a description of the information flow. We will examine the model under the topics of definitions, axioms, and propositions, then look at its application to Dretske’s theory.

First, let us note the following definitions:

Definition. There is a three-place relation $s_1 \mapsto_c s_2$ that holds between a channel c and situations s_1, s_2 . Channels may be among the situations or types.

Definition. For every pair of types, ϕ and ψ , there is an object $\phi \rightarrow \psi$, called a *constraint*. The constraints may or may not be among the types, but the *of-type* relation ‘:’ is also a relation between channels and constraints. For example, a channel c that is of type $\phi \rightarrow \psi$ is written as $c: \phi \rightarrow \psi$.

Definition. Given a channel c , if $s_1 \mapsto_c s_2$, s_1 is called a *signal* for s_2 , and that s_2 is called a *target* of s_1 , *relative* to the channel c . The relation $\mapsto_c$ that holds between signal and target is called the *signaling relation* of c . An ordered pair $\langle s_1, s_2 \rangle$ consisting of a signal and one of its targets is called a *signal/target pair* for c .

Definition. If $c: \phi \rightarrow \psi$ then ϕ is said to *indicate* ψ *relative to* c .

Definition. If $c: \phi \rightarrow \psi$ and $s_1: \phi$, but there is no s_2 such that $s_1 \mapsto_c s_2$, then s_1 is said to be a *pseudo-signal* for $c: \phi \rightarrow \psi$.

Definition. Suppose $c: \phi \rightarrow \psi$. The constraint is said to be *absolute* on c (denoted by $c: \phi \rightarrow_! \psi$) iff there are not pseudo-signals for $c: \phi \rightarrow \psi$. Otherwise, the constraint is said to be *conditional* on c .

Definition. Channel c is a *sequential composition* of c_1 followed by c_2 (denoted by $c = c_1; c_2$), provided for all situations s_1 and s_2 , $s_1 \mapsto_c s_2$ is the case iff there is a (intermediary) situation s such that $s_1 \mapsto_{c_1} s$ and $s \mapsto_{c_2} s_2$.

Hence, channels can be composed by using the targets of c_1 as the signals of c_2 . Clearly, the composition of channels is associative: For all channels c_1, c_2, c_3 , $c_1; (c_2; c_3) = (c_1; c_2); c_3$.

Definition. Given types ϕ, ψ , ϕ entails ψ (or is a subtype of ψ , denoted by $\phi \subseteq \psi$) iff every situation of type ϕ is also of type ψ .

Definition. The identity relation is modeled as an information channel as well; it is called “channel 1.”

A channel c is said to be a *parallel composition* of channels c_1 and c_2 provided for all situations s_1, s_2 :

$$s_1 \mapsto_c s_2 \text{ iff } s_1 \mapsto_{c_1} s_2 \text{ and } s_1 \mapsto_{c_2} s_2$$

Definition. Channel c_1 is said to be a *refinement* of channel c_2 , written $c_1 \preccurlyeq c_2$ or $c_2 \succcurlyeq c_1$, iff $c_1 = c_1 \parallel c_2$.

If c_1 is a refinement of c_2 , then any information that could be obtained by using c_1 and c_2 in parallel could be obtained by using c_1 alone. If this is the case, then there is no actual difference between using one of them and using the both in parallel, however, refinement shows that c_1 is more discriminative, therefore, supports more constraints than c_2 .

Consequently, if c_1 is a refinement of c_2 , then for all situations s_1 and s_2 , if $s_1 \mapsto_{c_1} s_2$ then $s_1 \mapsto_{c_2} s_2$.

The following are the axioms for the model:

Axiom of Soundness. The channel c is of type $\phi \rightarrow \psi$ iff for all situations s_1, s_2 , if $s_1 \mapsto_c s_2$ and $s_1 : \phi$, then $s_2 : \psi$.

The direction from left to right is justified by the facticity of information: If $s_1 : \phi$ and $s_1 : \phi$ carries the information $s_2 : \psi$ then $s_2 : \psi$.

The converse derives from the classical universal quantifiers, ranging over situations with the standard Tarskian account of its truth conditions, hence, we need to think of the collection of situations as containing not just actual situations, but also all possible situations, since $c : \phi \rightarrow \psi$ is claimed to express a law-like connection between c and $\phi \rightarrow \psi$, not just a Humean constant conjunction of real events.

Sequential Composition Axiom. There is a *total* binary operation ‘;’ of channel composition. Sequential composition of channels is associative.

Axiom of Logic Channel. There is a special *logic* channel, which is channel 1, that relates each object a only to itself: $a \rightarrow_1 b$ iff $a = b$. For each channel c , $c;1 = 1;c = c$.

Parallel Composition Axiom. There is a (*partial*) binary operation $\parallel$ of *parallel composition* of channels. Parallel composition is commutative, associative, and idempotent ($c \parallel c = c$).

In what follows, we will use the *total* version of the parallel composition in which it is defined for each instance.

Inversion Axiom. For each channel c , there is a channel c^{-1} with the following properties:

1. $s_1 \mapsto_{c^{-1}} s_2$ iff $s_2 \mapsto_c s_1$.
2. $1^{-1} = 1$.
3. $(c_1; c_2)^{-1} = c_2^{-1}; c_1^{-1}$.
4. $(c_1 \parallel c_2)^{-1} = c_1^{-1} \parallel c_2^{-1}$.

Note that $c; c^{-1}$ is not necessarily equal to the channel 1.

Modularity Axiom. Provided that the channels d has an inverse, for all channels c, d and e , $(c; d) \parallel e \ll ((c \parallel (e; d^{-1})); d)$ holds.

Modularity Axiom provides a basis, in a sense, for a geometric interpretation of information; if there is any information that can be carried from s_1 to s_3 in parallel with $c; d$ and e , then it can be carried there by first doing c and $e; d^{-1}$ in parallel, and then following that by d :

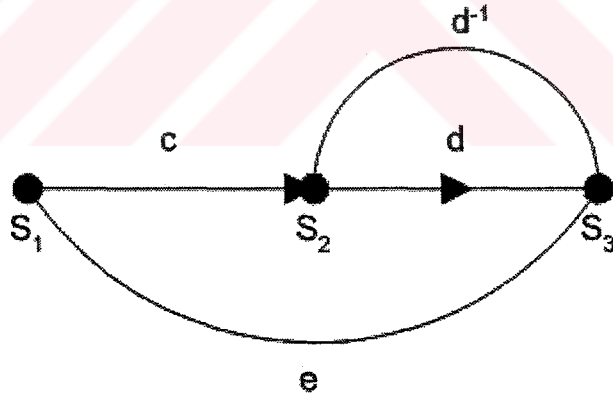


Figure 3: The modular law for channels.

Then we can continue with the propositions to get an insight of the model:

Proposition. Suppose $c = c_1; c_2$.

If $c_1: \phi \rightarrow \psi$ and $c_2: \psi \rightarrow \theta$,

then $c: \phi \rightarrow \theta$.

Proof. Follows from the definitions of sequential composition and constraint support.

Proposition (Xerox Principle).

If $s_1: \phi \rightarrow_{c_1} s_2: \psi$ and $s_2: \psi \rightarrow_{c_2} s_3: \theta$,
then $s_1: \phi \rightarrow_{c_1; c_2} s_3: \theta$.

Proof. Note that the Xerox principle needs not the preceding proposition, but the sequential composition axiom, since the axiom asserts that all channels can be sequentially composed.

Corollary (of Axiom of Logic Channel). Given types ϕ, ψ , the following are equivalent:

1. $\phi \subseteq \psi$.
2. $1: \phi \rightarrow \psi$.
3. $1: \phi \rightarrow! \psi$.

Given the assumption that channels have non-vacuous types and parallel composition is *total* (defined for each instance), the propositions for the addition of information and the exhaustive cases are obtained:

Proposition (Addition of Information). Assume that $c = c \parallel c$ exists.

If $s_1: \phi \rightarrow_{c_1} s_2: \psi$ and $s_1: \phi' \rightarrow_{c_2} s_2: \psi'$
then $s_1: (\phi \wedge \phi') \rightarrow_c s_2: (\psi \wedge \psi')$.

Proof. Follows from the parallel composition axiom. Note that the addition principle asserts that if $s_1: \phi$ carries the information that $s_2: \psi$ relative to a channel c_1 and that $s_1: \phi'$ carries the information that $s_2: \psi$ relative to another (parallel to c_1) channel c_2 , then there must be a channel relative to which $s_1: \phi \wedge \phi'$ carries the information that $s_2: \psi \wedge \psi'$.

Proposition (Exhaustive Cases). Assume the following:

$c_2 \parallel c_3$ exists,
 $s_1: (\phi \wedge \phi') \rightarrow_{c_1} s_2: (\psi \wedge \psi')$,

$$s_2 : \psi \rightarrow_{c_2} s_3 : \theta,$$

$$s_2 : \psi' \rightarrow_{c_3} s_3 : \theta.$$

Then $s_1 : \phi \rightarrow_c s_3 : \theta$, where $c = c_1 ; (c_2 \parallel c_3)$.

Proposition. If $c_2 : \phi \rightarrow \psi$ and $c_2 \succcurlyeq c_1$ then $c_1 : \phi \rightarrow \psi$.

Proposition. The ordering $\preccurlyeq$ is a partial ordering of channels.

Proof. Suppose $c_1 \preccurlyeq c_2$ and $c_2 \preccurlyeq c_3$, i.e., $c_1 = c_1 \parallel c_2$ and $c_2 = c_2 \parallel c_3$. We want to prove that $c_1 \preccurlyeq c_3$, i.e., $c_1 = c_1 \parallel c_3$. This is shown as follows:

$$\begin{aligned} c_1 &= c_1 \parallel c_2 \\ &= c_1 \parallel (c_2 \parallel c_3) \\ &= (c_1 \parallel c_2) \parallel c_3 \\ &= c_1 \parallel c_3. \end{aligned}$$

Proposition. Parallel composition preserves refinement:

If $c_1 \preccurlyeq c_2$ then

- $(c_1 \parallel d) \preccurlyeq (c_2 \parallel d)$
- $(d \parallel c_1) \preccurlyeq (d \parallel c_2)$

Proof. Assume that $c_1 \preccurlyeq c_2$, i.e., $c_1 = c_1 \parallel c_2$. Then we have

$$\begin{aligned} c_1 \parallel d &= c_1 \parallel (d \parallel d) \\ &= (c_1 \parallel c_2) \parallel (d \parallel d) \\ &= c_1 \parallel (c_2 \parallel d) \parallel d \\ &= c_1 \parallel (d \parallel c_2) \parallel d \\ &= (c_1 \parallel d) \parallel (c_2 \parallel d). \end{aligned}$$

Thus, $c_1 \parallel d \preccurlyeq c_2 \parallel d$, as desired. The second proof is entirely parallel.

Proposition (*Monotonicity of Sequential Composition*). Sequential composition preserves refinement. That is, if $c_1 \preceq c_2$ then

- $(c_1; d) \preceq (c_2; d)$
- $(d; c_1) \preceq (d; c_2)$.

Proposition. Given the channels c_1, c_2, d_1, d_2 :

$$d_1; (c_1 \parallel c_2); d_2 \preceq (d_1; c_1; d_2) \parallel (d_1; c_2; d_2).$$

Proposition (*Contraposition –subjunctive case*).

$$\text{If } s_1: \phi \rightarrow_c s_2: \psi \text{ then } s_2: \neg \psi \rightarrow_{c^{-1}} s_1: \neg \phi.$$

Proof. Follows from the inversion axiom.

Note that invertibility of channels does not imply the same property for constraints, that is, just because $c: \phi \rightarrow \psi$ it does *not* follow that $c^{-1}: \psi \rightarrow \phi$.

Proposition (*Contraposition –indicative case*).

$$\text{If } s_1: \phi \Rightarrow_c s_2: \psi \text{ then it is not the case that } s_2: \neg \psi \Rightarrow_{c^{-1}} s_1: \neg \phi.$$

The contrast with the subjunctive analogue derives from the observation that $s_1: \phi \Rightarrow_c s_2: \psi$ asserts that s_1 is of type ϕ , however, $s_2: \neg \psi \Rightarrow_{c^{-1}} s_1: \neg \phi$ would entail the contradictory assertion that s_1 is not of type ϕ .

As the basis for the modeling, it is in place to examine Dretske's conceptual model in the context of channels. As noted earlier, Dretske's model conflates the event and the event-types in probability theory. Thus, in conjunction with his definition of information content, we arrive at the following statements:

- $P(\phi | s) = 1$ iff $s: \phi$.
- $P(\psi | \phi) = 1$ iff $\phi \rightarrow \psi$.

We can see from the above statements that a situation's being of a type and information flow among situations coincide in Dretske's model.

Against this fundamental loss of particulars, Barwise chooses to examine two cases of one-event and multiple-event in Dretske's conceptual framework, but in two separate models; an approach that we will follow.

The One-Event Dretske Model. In this model, there is only one event, ω , then the only reasonable channel is the identity channel: $\omega \mapsto_1 \omega'$ iff $\omega = \omega'$. The identity channel satisfies the principles of information flow given above in a straightforward manner, however, one point should be noted: 1: $\phi \rightarrow \psi$ requires that $P(\psi | \phi) = 1$, however, it does not follow from $P(\psi | \phi) = 1$ and $\omega \in \phi$ that $\omega' \in \psi$, since it does not exclude the possibility that the points ω can be members of some event-type probability 0.

Against this possibility, the condition that $P(\{\omega\}) > 0$ for all points ω in the probability space must be required. This requirement converts the model to the possible-worlds model of information containment:

$$s_1: \phi \Rightarrow_1 s_2: \psi \text{ iff } s_1 = s_2, \psi \subseteq \phi, \text{ and } s_1 \in \phi.$$

The Multiple-Event Dretske Model. In this model, there are two kinds of channels:

- A channel c_p determined by the probability function P :

$$s_1: \phi \mapsto_{c_p} s_2: \psi \text{ iff } P(s_1 | s_2) = 1 \text{ and } P(\psi | \phi) = 1.$$

- The identity channel as in the one-event model:

$$s_1: \phi \mapsto_1 s_2: \psi \text{ iff } s_1 = s_2, \text{ and } P(\psi | \phi) = 1.$$

Note that the order of s_1 and s_2 in the conditional probability is the reverse of the identity channel. An intuitive justification for this reversal can be given by referring to the example of indicators, as $s_1 \mapsto_{c_p} s_2$ means "height is caused by temperature." This channel can be taken to model the information flow in a different direction from causality.

Because of the event and event-type conflation, sequential and parallel composition becomes the same operation.

In this model, each type has a set-theoretic compliment as its negation type. Thus, channels have inverses in general. The inverse channel c_p^{-1} is defined as $s_1 \mapsto_{c_p^{-1}} s_2$ iff $P(s_2 | s_1) = 1$.

CHAPTER 7

RELEVANT LOGICS

Logic, as the science of reasoning, needs languages that are sufficiently expressive to encode the entailment relation between the premises and the conclusion of a valid argument, discriminating the valid arguments from the invalid ones. These languages should be capable of responding the normative and the descriptive constraints of logic; while judging the validity, they should reflect the common intuitions of reasoning. The interaction between the normative and the descriptive comes to the fore as a main motivation for development in logic.

A development in the vein of this motivation is the family of logical systems called relevant logics (called *relevance logics* in North America because of historical reasons. Since the mainstream of works originate from Britain and Australasia and in order to keep coherence with the naming of other kinds of logic, we will prefer the other usage), of which objective is to warrant truth of an argument together with the condition that its conclusion is, in some way, relevant to its premises, can be viewed as a product of this motivation.

Though each one of these systems is entitled to be called a logic, we will refer to them singularly as relevant logic, since we will be concerned the main themes and will follow the canonical accounts, leaving out the variety of approaches, except for the mentioning briefly for the sake of completeness.

The following exposition is based mainly on [3], [23], [25], and [27].

7.1. Paradoxes of Implication

It has been recognized for long that the notion of implication in the classical logic, which is called material implication, results in stating some arguments valid contrary to the sound reasoning principles. These arguments are of the schemes that are known as the *paradoxes of material implication*. The underlying source of these paradoxes can be

identified as the commitment of the classical logic to *ex falso quodlibet* (EFQ): A false proposition implies any other, thus, allowing inferences in which premises and conclusion has no reasonably considerable connection. To cite a few implicational schemes, using ‘ $\neg$ ’ for negation in the classical logic and ‘ $\supset$ ’ for material implication:

$$\begin{aligned} A &\supset (B \supset A) \\ \neg A &\supset (A \supset B) \\ (A \supset B) &\vee (B \supset A) \end{aligned}$$

Consider, for example, an application of the last scheme that has the value true in any case:

Either $2 + 2 = 5$ implies that snow is white or snow is white implies that $2 + 2 = 5$.

Upon the observation of such untenable results of material implication, a prominent attempt, which is due to C.I. Lewis, has been the addition of a new connective to the classical logic, called *strict implication*. The idea is that A *strictly* implies B just in case that A is inconsistent with the negation of B :

$$A \prec B \triangleq \neg(A \supset \neg B)$$

Though an improvement over the material implication, the strict implication, as well, fails to avoid *fallacies of relevance*, such as:

$$\begin{aligned} (A \wedge \neg A) &\prec B \\ A &\prec (B \prec B) \\ A &\prec (B \vee \neg B) \end{aligned}$$

It can be observed on such instances as the above that the implication still confers validity on formulae in spite of the irrelevance of premises and conclusion, or antecedent and consequent.

Introducing an implicational connective is not only defining a new connective by its different properties, but also asserting a different theory of implication: A point of view that

calls for an analysis of the grounds on which the implication and validity notions of classical logic stand.

7.2. Implication Theory of Classical Logic

In order to clarify the position of the classical logic with respect to the relevance issues, both sides of the logical consequence, denoted by ' $\vdash$ ', and the validity of movement from the premises side onto the conclusion side should be examined.

The notion of logical consequence is developed as equivalence to deducibility, hence for single premise:

$A \vdash B$ iff it is a theorem of logic that B is a logical consequence of A .

In case that B is deducible from multiple premises, $A_1, \dots, A_n, A$, provided that $A_1, \dots, A_n$ are true, since if A is true, then B must be true, and by the same line of argument in the converse direction, we obtain

$A_1, \dots, A_n, A \vdash B$ iff $A_1, \dots, A_n \vdash A \supset B$

which is the general form of the *deduction (or deduction equivalence) theorem*. The comma that indicates the combination of premises is taken in classical logic to be the *and* connective, ' $\wedge$ '. A critical point to remark is that the combination of premises must be taken into account when the behavior of implication is examined, since both sides of $\vdash$ is intrinsically correlated.

The validity view of the classical $\vdash$, which is usually called the *classical account of validity*, excludes any formal constraints on the linkage between premises and conclusion except for the plain preservation of truth such that the impossibility of the premises to be true and the conclusion false as the necessary and sufficient condition of the validity of an argument.

The classical account of validity and the classical form of the deduction theorem lead us to the entirely truth-functional character of the classical logic; truth-values of all its propositional connectives are determined by the values of their component formulae alone. Hence, the truth-value of "if A then B " is determined by the truth-values of A and B . Accordingly, the logical consequence is identified with the truth-function of material

(*extensional*) implication, rejecting any intensional dimension that would capture the relation of entailment or that of implication between propositions. On the contrary, relevant logic has both extensional and intensional (non-truth-functional) connectives. We will use the unary connective ' $\sim$ ' (relevant negation), the binary connectives ' $\wedge$ ' (extensional conjunction) and ' $\rightarrow$ ' (relevant implication or entailment).

7.3. Main Themes of Relevant Logic

The attempts to diagnose the fallacies of relevance inherent in the classical logic and to develop modifications in order to give an account of logical validity that respects our intuitions about what can and cannot be inferred from what date back to I.E. Orlov, who had an important, though not known widely, work of 1928 over the subject. In the 1950s, W. Ackermann, A. Church and M. Shaw-Kwei and later A. R. Anderson, N. D. Belnap, M. J. Dunn and R. K. Meyer, to cite a few among the many workers of the field, have resulted in the relevant logic.

Relevant logic belongs to the family of the formal systems called *substructural logics*, which are non-classical logics weaker than classical logic, focusing on the behavior of *structural rules* present in classical logic. For relevant logic:

If $X \vdash A$ holds, then X must be *relevant* to A .

A complete formal theory, comprising both a deductive system and a sound and complete semantics, can be provided for the notion of relevant implication (if the conditional $\rightarrow$ is *relevant*, hence, if to say that $A \rightarrow B$ is true is to say, at least, that B truly *depends* on A). The requirement that relevance is a necessary condition of entailment may appear trivial, relevant entailment is nothing but circularity: Two propositions are logically relevant if one of them entails the other. However, the point of relevant logic is not to give formal criteria of relevance that enables us to pick out true entailments *a priori* among the others, but to give a relevant account of entailment. Hence, relevant logic asserts itself to be relevant in two respects:

- Exclusion of irrelevances, avoiding the paradoxes of implication while providing non-trivial and improved formalisms of our pre-formal notions of implication and entailment.
- A proof theory based on a relevant notion of proof, such as requiring variable sharing.

7.4. Model-Theoretical Semantics for Relevant Logic

We need a semantics within a relevance metalanguage such that a relevancy *discriminant* can provide valuation with respect to the truth-conditions that map propositions onto their translations in the metalanguage. This is achieved by indexing truth-value, each index encoding a possible world. Thus, truth of formulae comes out to be relative to worlds; a wff A is not simply true or false, but *at* some index a of a possible world.

Non-actual worlds can be constituted in a nominalist manner as the sets of statements that are descriptive of those possible situations, that is, a wff A is true at a world a is defined with the condition that $A \in a$, and that A is one of the propositions constituting a .

Traditionally, two notions have been associated with the conception of a possible world:

- A possible world is maximal in model-theoretic sense; that is, intuitively it contains as much information as possible to specify its constituents.
- It is consistent, satisfying all logical laws; otherwise, it would be an *impossible* world.

In the context of relevant logic, manifestation of the underlying thoughts must be conceived differently. Worlds are maximal, but in the sense that the truth-valuation function will be total; every atomic wff will be either true or false. However, a world that is maximal in this sense, is not necessarily negation-complete. Indeed, the truth-definition based on this atomic valuation entails that not all logical laws will hold at some worlds –those where they hold are called *normal* worlds, constituting the subset N of possible worlds K .

Routley-Meyer (Relational) Model. Relevant logic has *Routley-Meyer model* as its formal semantics due to Richard Routley and Robert K. Meyer, which comprises frames in the modal-logical sense. Hence, there is a set of worlds (equivalently, indices), K , of which members can be interpreted as states of affairs of information for our context and there is an accessibility relation R on K , K and R constituting a frame. A wff A is valid in a frame iff A is true on the set of normal worlds N ($N \subseteq K$) on all interpretations. R is a ternary relation (that is, $R \subseteq K^3$) between worlds providing truth-condition for the *relevant implication*, ' $\rightarrow$ ', which is a binary relation.

Logical consequence can be defined with appropriate constraints on the set K . The task of the formal work is to construct a recursive specification of indexed truth-values,

leading through to the definition of logical consequence. Since relevant logic contains intensional connectives as well as extensional, and for example, the truth-value of “ $A \rightarrow B$ ” at a in which the connective is intensional, cannot depend on the values of A and B at a alone, the recursive definition of truth from the atomic formulae will not be straightforward as in the extensional semantics. However, relevant logic preserves the idea of logical consequence in extensional semantics, which can be expressed in the model-theoretical manner:

A is a logical consequence of a set of wffs X (that is, $X \models A$)
iff
 A is true whenever every member of X is true.

An interpretation in extensional logic consists simply in an assignment I of truth-values to the atomic wffs. Let $I(A) = \text{true}$ indicates “wff A is true under I ” (= false if A is false under I). Then the condition becomes

$X \models A$ iff for all I , $I(A) = \text{true}$ iff $I(B) = \text{true}$ for all $B \in X$.

In the context of frames, the truth-condition for the relevant implication becomes:

$I(A \rightarrow B, a) = \text{true}$
iff
for all b, c , $I(B, c) = \text{true}$ whenever $I(A, b) = \text{true}$ and $Rabc$

which formalizes the essentially relational conception of relevant implication $\rightarrow$.

An argument is *valid* in a model just when in any point at which the premises are true, so is the conclusion. This truth-condition of a conditional $A \rightarrow B$ evaluated at a world allows avoiding paradoxes. For example, we can have $B \rightarrow B$ fail to be true at a by having $Rabc$ where B is true at b , but not at c . Thus there can be worlds at which $B \rightarrow B$ fails, therefore, the paradoxical $A \rightarrow (B \rightarrow B)$ fails. Intuitively, if $Rabc$ holds, then wffs true at a and b are composite at c , in which case a and b are said to be *compatible* relative to c .

The relation R is adaptive to the behavior of the mode of premise combination in the proof theory through different conditions that can be placed on R . Thus, ternary relational

semantics is apt to model the behavior of various logics. For example, if premise combination is commutative, a *symmetry* condition on R reflects this:

$$Rabc \text{ iff } Rbac$$

The truth conditions for extensional conjunction and disjunction are, if conjunction and disjunction satisfy the distribution axiom, they can be modeled straightforwardly: A conjunction is true at a point just when both conjuncts are true at that point, and a disjunction is true at a point just when at least one disjunct is true there. Hence,

$$I(A \wedge B, a) = \text{true} \text{ iff } I(A, a) = \text{true} \text{ and } I(B, a) = \text{true}.$$

$$I(A \vee B, a) = \text{true} \text{ iff } I(A, a) = \text{true} \text{ or } I(B, a) = \text{true}.$$

The relationship between non-normal worlds and normal worlds is remarkable for the purposes of information-theoretical analysis. Let us define a partial ordering relation on worlds $\leq$ such that

$$a \leq b \text{ iff there is some world } n \text{ in } N \text{ such that } Rnab$$

and given I , let the *hereditariness* property holds:

$$\text{If } I(A, a) = \text{true} \text{ and } a \leq b, \text{ then } I(A, b) = \text{true}.$$

Thus, we obtain the *semantical entailment fact*:

$$A \text{ entails } B \text{ iff, for all worlds } a, \text{ if } I(A, a) = \text{true} \text{ then } I(B, a) = \text{true}.$$

The true $\rightarrow$ statements corresponding to entailment, which preserves truth, are those that are true at all normal worlds n in N :

$$A \text{ normally implies } B \text{ iff, for all } n \text{ in } N, I(A \rightarrow B, n) = \text{true}.$$

Negation. The ternary accessibility relation R is not sufficient to avoid all the paradoxes of implication, since it, by itself, makes such paradoxes as $(A \wedge \sim A) \rightarrow B$ and $A \rightarrow (B \vee \sim B)$ to be valid.

These paradoxes are avoided by the inclusion of inconsistent and non-bivalent worlds in the semantics. If there were no worlds at which $A \wedge \sim A$ holds, then $(A \wedge \sim A) \rightarrow B$ would also hold everywhere according to our truth condition. Likewise, if $B \vee \sim B$ held at every world, then $A \rightarrow (B \vee \sim B)$ would be universally true.

Avoiding $(A \wedge \sim A) \rightarrow B$ makes relevant logic a paraconsistent logic, tolerating contradiction, since contradictions are everywhere. This is not to suggest that the contradictions in all (or even any) of these cases is ineluctable. Great efforts have been made to resolve (or at worst live with) the associated problems. Rather, people reason around any inconsistency in their present beliefs.

$\sim A \wedge A$ is to be true at some index and B is to be false at the same index. Therefore, since $\wedge$ is truth-functional, $\sim A$ and A must be true at the same index. However, A cannot be both true and not true, otherwise there cannot be any reasonable interpretation of A .

This leads to the semantics for negation. The use of non-bivalent and inconsistent worlds requires a non-classical truth condition for negation, which is handled by the *Routley star operator*. This is a unary operator on worlds (theories, set-ups or situations according to the context), which can be defined as a function $*$: $K \rightarrow K$ such that for each world a , there is a companion world a^* that satisfies,

$$I(\sim A, a) = \text{true iff } I(A, a^*) = \text{false}$$

where I is an interpretation.

Intuitively, $*$ -operator sets up a framework in which both $\sim A$ and A may be true at some world a . Therefore, a is an inconsistent world. In accordance with the formal truth-condition for $\sim$, both $\sim A$ and A will be false at a^* . Therefore, a^* is negation-incomplete. Hence, if what one of a and a^* asserts about A is incomplete, then the other one will be inconsistent, and vice versa.

The Routley star operator provides a semantic refutation of $(A \wedge \sim A) \rightarrow B$ within this formalism, which is known as Australian plan of relevant semantical analysis: Setting $I(A, a) = \text{true}$ and $I(A, a^*) = I(B, a) = \text{false}$ at a fixed a , and applying the postulate makes the premise

true at a but its conclusion false at a , thus, failing to preserve truth, which is essential to any conception of logical consequence.

An interpretation of the Routley star that is due to Dunn uses a binary relation, C , on worlds. ‘ Cab ’ means that b is *compatible* with a . Then, a^* is the maximal world that is compatible with a .

There are other semantics for negation and to refute $(A \wedge \sim A) \rightarrow B$, such as a four-valued semantics, which is known as American plan.

7.5. Proof Theory of Relevant Logic

A formal proof is expected to show the syntactic transformation of statements from the premises of an argument to its conclusion through the use of rules, hence, intuitively saying, a proof should enable us to understand which statement *depends on* which other one. In accordance with its objective, relevant logic asserts a proof theory that forces the dependence between the statements to satisfy a notion of relevance. For example, the traditional deducibility allows that if a proposition is a premise then it can also be a conclusion, that is

$$A, B \vdash A$$

is a valid deduction for classical logic, but not for relevant logic. The idea of relevant logic is that if a statement has no role in deriving another statement, then it bears no relevance to the derived statement. Therefore, the rule of *weakening*, which has the general form:

$$\frac{X \vdash A}{X, Y \vdash A}$$

is not a valid inference; from a stronger statement, that A entails X to a weaker one is derived, in which Y is an arbitrary premise from the relevancy point of view. The point can be seen in another way; setting X to A , Y to B and applying the deduction theorem to the general form of weakening twice, we obtain the theorem

$$\vdash A \rightarrow (B \rightarrow A)$$

Y.G. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

which is a paradox of implication. Note that, if $\rightarrow$ is a material conditional, $B \rightarrow A$ follows from A .

Relevant logic has another conception of deducibility that requires all premises to be used in the deduction without arbitrariness. A large variety of approaches have been developed for the proof theory of relevant logic. Here, the natural deduction system, due to Anderson and Belnap, will be dealt, which traces the usage of the premises $A_1, \dots, A_n$ in the deduction $A_1, \dots, A_n \vdash B$ by the method of *relevance indices*.

In this method, hypotheses in a proof are tagged by indices, denoted by the numbers in set brackets, which indicate the assumptions used to prove the formula and are used to keep track of their usage in the deduction. The other steps are indexed by the indices on the hypotheses that are used to obtain them. For an hypothesis to be qualified as contributing to deduce the conclusion, an index denoting the hypothesis must be appear in the deduction and be discharged at some later step. For an argument to be valid in this system, all hypotheses stated must be used.

Discharging a hypothesis A in a conditional sub-proof is performed, if the conclusion C of that sub-proof is tagged by the index on the hypothesis A . If it is, $A \rightarrow C$ is inferred by discharging A , which marks the end of the sub-proof. After this application of the $\rightarrow$ I rule, the index on the discharged hypothesis A is removed from the indices on $A \rightarrow C$. If an index on a hypothesis does not appear in the steps on which the conditional proof is based, then the hypothesis cannot be discharged in that proof, thus it is ensured that only hypotheses that effectively have a role can be discharged.

This natural deduction system gives an intuitive understanding of relevance in proofs. Let us examine the following example:

1. $A_{\{1\}}$	hyp
2. $(A \rightarrow B)_{\{2\}}$	hyp
3. $A \rightarrow (B \rightarrow C)_{\{3\}}$	hyp
4. $B_{\{1,2\}}$	1, 2, $\rightarrow$ E
5. $(B \rightarrow C)_{\{1,3\}}$	1, 3, $\rightarrow$ E
6. $C_{\{1,2,3\}}$	4, 5, $\rightarrow$ E
7. $A \rightarrow C_{\{2,3\}}$	1, 6, $\rightarrow$ I

- | | |
|--|-----------------------|
| 8. $((A \rightarrow B) \rightarrow (A \rightarrow C))_{\{3\}}$ | 2, 7, $\rightarrow$ I |
| 9. $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$ | 3, 8, $\rightarrow$ I |

But the system of indices can allow irrelevant premises into the argument. One way is using a rule of conjunction introduction by adding in an irrelevant premise. For example:

- | | |
|--------------------------------------|-----------------------|
| 1. $A_{\{1\}}$ | hyp |
| 2. $B_{\{2\}}$ | hyp |
| 3. $(A \wedge B)_{\{1,2\}}$ | 1, 2, $\wedge$ I |
| 4. $A_{\{1,2\}}$ | 3, $\wedge$ E |
| 5. $(B \rightarrow A)_{\{1\}}$ | 2, 4, $\rightarrow$ I |
| 6. $A \rightarrow (B \rightarrow A)$ | 1, 5, $\rightarrow$ I |

which is a paradox of implication due to the first premise that is irrelevant to the deduction. In order to avoid such situations, Anderson and Belnap give the following conjunction introduction rule:

$$A_i, B_i \vdash (A \wedge B)_i$$

which stipulates that two formulae must have the same index before the rule of conjunction introduction can be used.

The theory of relevance depends, in a sense, on real use of a premise and the rules are allowed to access premises in the natural deduction system. In relevant logic, the real use requirement is enforced by treating premises and their relations to conclusions in an intensional manner. On the other side, a sequent calculus treats the structural ‘,’ in a sequent of the form $A_1, \dots, A_n \vdash B$ as extensional conjunction $\wedge$, thus, $A \wedge B \vdash C$.

In relevant logic, the inference $A_1, \dots, A_n \vdash B$ is equivalent to $\vdash A_1 \rightarrow (\dots (A_n \rightarrow B) \dots)$ and, the latter is not equivalent to $\vdash (A_1 \wedge \dots \wedge A_n) \rightarrow B$, since the premises are not bound with the classical (extensional) conjunction, whose truth conditions are determined by a truth-table), but with an intensional conjunction called *fusion*, ‘ $\circ$ ’.

Fusion. The traditional truth-functional conjunction, expressed by “ A and B ,” entails each of A and B , therefore the falsity of either conjunction suffices for the falsity, and the

truth of both for the truth. However, there is also a non-truth-functional conjunction sense of ' A and B ' whose falsity supports the inference from A to 'not- B '. There is a distinction between these two senses:

- If asserting “not- $(A$ and $B)$ ” is intended simply to state that the simple (possibly, arbitrary) union of A and B is false, then truth of A does not entail “not- B ”; B might be true. In this sense, “not- $(A$ and $B)$ ” is weaker than both “not- A ” and “not- B ”.
- If asserting “not- $(A$ and $B)$ ” is intended to bind two propositions together in such a way that assures their mutual relevance in deriving any consequence from the fused conjunction.

Hence, the fusion (that is, intensional conjunction) characterizes a logical relation between two propositions with correct formal properties and respects the intuition that a contradiction is not relevant to every proposition. Whatever “ $A \circ B$ ” entails requires both A and B in its derivation. Since logic is about *logical consequence*, the *conditional* is a central notion because of its intimate connection with logical consequence. This connection is expressed in *residuation condition*: The central requirement of fusion is that it obeys *residuation*:

$$A \circ B \vdash C \text{ iff } A \vdash B \rightarrow C$$

It says that C follows from A together with B just when $B \rightarrow C$ follows from A alone. The validity of the transition from B to C (given A) is recorded by the conditional $B \rightarrow C$. (It is called residuation by analogy with residuation in mathematics. Consider the connection between addition and subtraction. $a + b = c$ if and only if $a = c - b$. The resulting a (which is $c - b$) is the *residual*, what is left of c when b is taken away.) Thus, fusion needs to satisfy the deduction theorem; the other properties of fusion vary among relevant logic systems.

Some relevant logic systems do not allow fusion to commute, that is, $A \circ B$ is not equivalent to $B \circ A$, while there are systems like **R** (to be examined below), where $A \circ B \rightarrow B \circ A$ is a theorem. In **R**, $A \circ B$ is equivalent to “ $\sim(A \rightarrow \sim B)$ ”, which leads to is circularity, unlike the systems without commutativity.

Fusion is not idempotent in many relevant logic systems, such as in **R** and **E**, that is, $A \circ A$ is not equivalent to A .

The semantic definition for fusion can be given as follows:

$$I(A \circ B, a) = \text{true}$$

iff

there are some worlds b, c , such that $Rbca$ and $I(A, b) = \text{true}$ and $I(B, c) = \text{true}$.

An intuition can be gained by construing fusion as a type of binary possibility connective. $I(A \circ B, a) = \text{true}$ says that, A and B are jointly possible at a in the sense that a recognizes the combination of worlds in which A and B obtain.

Fission. In one sense of 'or' is the truth of ' A or B ' follows from that of A or of B , and it is false if both are false, therefore, the values of A and B completely determine the value of " A or B ".

There is another sense in which " A or B " supports the inference from " $\text{not-}A$ " to B (equivalently, from ' $\text{not-}B$ ' to A). In this non-truth-functional sense, " A or B " is equivalent to " $\text{if not-}A$ then B ". This sense of disjunction is referred to as *fission*, denoted by " $A + B$ ".

Consequently, disjunctive syllogism is considered valid for fission but for $\vee$, invalid. $+$ has the same algebraic relationships to $\circ$ that $\vee$ has to $\wedge$.

Variable Sharing. Relevant logic's thesis that logically true implications should not have antecedents that are completely irrelevant to their consequents leads to enforcing a connection between the content of the antecedent and the content of the consequent. Content is a semantic notion, however, what is to be specified either in terms of a set of valid proofs or a set of theorems is mainly syntactic.

A minimal connection between the semantic and the syntactic can be provided by the variable sharing constraint. When possible (that is, for propositional relevant logic systems without Ackermann constant t and f or Church constant T and F), the variable sharing constraint forces the antecedent and consequent to share some content through one or more common propositional variables. Thus, they cannot be absolutely semantically irrelevant to one another.

Variable sharing also has the formal properties that the relation of meaning-connection is plausibly considered to have: It is reflexive (A is certainly connected in meaning with itself), symmetric (if A is connected in meaning with B , then B with A) and non-transitive (that A is meaning-connected with B and B with C does not warrant the conclusion that A is meaning-connected with C).

The variable sharing constraint by itself, however, does not yield an analysis of relevance. It is a necessary but not sufficient condition for relevant entailment. Note that the condition cannot be sufficient, since one merely accepted all of the theorems of classical logic that satisfied this constraint still be left with the paradoxes like

$$A \vdash A \wedge B$$

$$A \vee B, \sim A \vdash B$$

$$A \rightarrow (B \rightarrow A)$$

$$(A \rightarrow B) \vee (B \rightarrow C)$$

satisfy the variable sharing condition, but are invalid.

7.6. Overview of Relevant Logic Systems

Like in the modal logics, there is a family of relevant logics, some of which can be classified as modal and some as non-modal. We can schematize the family setting logic **B** (or its positive fragment **B+**) into the root and define the others as branches of the tree in regard to **B** (indeed, the central systems of relevant logic historically are the logic **E** and the system **R**). The classical extensional connectives are either primitive or defined as the classical, for example, $\vee$ and $\leftrightarrow$ defined as:

$$A \vee B \triangleq \sim(\sim A \wedge \sim B)$$

$$A \leftrightarrow B \triangleq (A \rightarrow B) \wedge (B \rightarrow A)$$

The axiom schemes and rules for **B** are as follows:

$$A1. \quad A \rightarrow A$$

$$A2. \quad (A \wedge B) \rightarrow A$$

$$A3. \quad (A \wedge B) \rightarrow B$$

$$A4. \quad ((A \rightarrow B) \wedge (A \rightarrow C)) \rightarrow (A \rightarrow (B \wedge C))$$

$$A5. \quad ((A \rightarrow C) \wedge (B \rightarrow C)) \rightarrow ((A \vee B) \rightarrow C)$$

$$A6. \quad A \rightarrow (A \vee B)$$

$$A7. B \rightarrow (A \vee B)$$

$$A8. (A \wedge (B \vee C)) \rightarrow ((A \wedge B) \vee (A \wedge C))$$

$$A9. \sim \sim A \rightarrow A$$

Modus Ponens

$$\vdash A \rightarrow B$$

$$\vdash A$$

$$\vdash B$$

Adjunction

$$\vdash A$$

$$\vdash B$$

$$\vdash A \wedge B$$

Affixing

$$\vdash B \rightarrow C$$

$$\vdash D \rightarrow A$$

$$\vdash (A \rightarrow B) \rightarrow (D \rightarrow C)$$

Contraposition

$$\vdash A \rightarrow \sim B$$

$$\vdash B \rightarrow \sim A$$

B+ is formed by omitting A9 and the contraposition rule from **B**. Various axiom schemes are added to these systems. For example, the logic **R** results from adding to **B** the schemes:

$$A10. (A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C)) \quad (\text{Suffixing})$$

$$A11. A \rightarrow (A \rightarrow B) \rightarrow (A \rightarrow B) \quad (\text{Contraction})$$

$$A12. (A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C)) \quad (\text{Permutation})$$

$$A13. (A \rightarrow \sim B) \rightarrow (B \rightarrow \sim A) \quad (\text{Contraposition})$$

$$A14. (A \rightarrow \sim A) \rightarrow \sim A \quad (\text{Reductio})$$

With the addition of these schemes, there is no need for the affixing and contraposition rules; they can be derived. Adding these axiom schemes to **B+** yields the positive fragment **R+** of **R**. **E** results from adding to **B** suffixing, contraction, contraposition, reductio and another axiom, for example:

$$A15. (((A \rightarrow A) \wedge (B \rightarrow B)) \rightarrow C) \rightarrow C$$

E is supposed to formalize the notion of entailment (that is, strict relevant implication), while **R** is supposed to be the *demodalized* version of **E**.

The system **T** of *ticket entailment* results from adding to **B** suffixing, contraction, contraposition (A13) and reductio (A14) and

$$A16. (B \rightarrow C) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C)) \quad (\text{prefixing})$$

The idea underlying this logic is to distinguish between logical rules, such as implications, and *inference tickets*. An inference ticket allows one to infer directly from the premise to the conclusion as a lawlike or causal (that is, nonlogical, insofar as logical principles can be distinguished from the nonlogical) relation in contrast to logical inference, for example

Today is Thursday. Therefore, tomorrow is Friday
or
This material is copper, so it conducts electricity

Hence, in **T**, an implication formula, $A \rightarrow B$, is taken as an inference ticket and cannot be used for another purpose of information carrying.

There are various arguments against and for one of type system or the other. For example, some favor logics stronger than **R** on semantic grounds, while some support systems weaker than **R** or **E** on the grounds such as that the weaker systems allow decidability, nevertheless, neither **R** nor **E**.

An extreme system is the logic **S**, which contains no theorems of the form $A \rightarrow A$. The system **S** gains the advantage that it excludes any *circular* arguments, since any argument in which a proposition implies itself is invalid.

CHAPTER 8

SITUATION SEMANTICS FOR RELEVANT LOGICS

8.1. Overview

Formal semantics provides a mapping of the syntax into another formalism; thus, the meaning of the terms still remains in need of interpretation. Such an interpretation states the conditions under which formulae of formal theory can be asserted; thus, it connects the proof-theoretic analysis with their use. In this connection, situation semantics can be used to give an intuitive interpretation to the semantics for relevant logic.

Possible worlds semantics characterizes logics mainly by grounding them on frames $\langle W, R \rangle$, where W is a non-empty set of worlds, and R is a binary (accessibility) relation on W , and truth conditions for the connectives. Logics vary with conditions on the accessibility relation between worlds or, with the truth conditions of definite connectives at definite worlds. On the other side, the perspective of situation semantics is directed to the *internal* structures of situations, so that the relations between infons are expounded. For example, when $s \models (\sigma \wedge \psi)$ is stated to hold if and only if $s \models \sigma$ and $s \models \psi$. Along with truth conditions for connectives, this statement possesses the potential to reflect the internal structure of abstract situations is comprised.

The Routley-Meyer semantics, the formal device to model relevant logic, has the possible worlds semantics approach and uses a theory of frames together with a set of truth conditions. It has features that enable frame theory to merge with situation semantics to acquire an intuitive view of the internal structure of indices (situations in the present context).

There are two main attempts to apply Routley-Meyer semantics to situation semantics as an interpretative connection for relevant logic. One of them is due to Edwin Mares [22], which is based on the theory of information developed by David Israel and John Perry [19] within situation semantics. The other one is due to Greg Restall [26], using Jon Barwise's

channel theory [4]. Both of them are based mainly on information-theoretical interpretation of Routley and Meyer's ternary accessibility relation in terms of in situation semantics. In the other direction, these interpretations can be viewed as a proof theory for situation semantics, hence, in Mares's words, a "situation syntax."

A world a contains informational links of nomic character, such as laws of nature and conventions, beside other information making up the world. There are other worlds, b and c , accessible from a within the relation ternary accessibility $Rabc$. The information in the world a is applied to the information in b is to be true (contained) in c . For example, a world of electrical field will contain the information that all charged particles of the same polarity repel each other and attract each other if they are of the different polarity. If b contains the information that a group of particles have positive polarity, then, c contains the information that they repel each other in accordance with the informational links in a . On these interpretations, the terms world, situation, and with some simplification, frame can be used interchangeably.

It is worth noting that by an independent argument in the context of biological information, a similar ternary relation is obtained in [20]. Its basis is C.S. Peirce's triadic sign relation (primary sign – interpretant – signified object). For example, DNA is the primary sign, the zygote "reads" the DNA as the interpretant, and transforms into the signified objects, which is the organism (for a detailed discussion, see [20]).

A common feature of both interpretations is the use of a special world (denoted by L in Mares and g in Restall) for the logical construct of information that provides non-empty truth sets for the relation R , since there is nothing true everywhere in all frames. This world records tautologies that are preserved across all worlds (on all frames in the present context). In other words, these are *logically informed* situations where semantic theorems must be valid.

Both interpretations focus on modeling information flow, which is an intrinsic behavior of information. There may be other interpretations for relevant logic as well as for information-theoretic approaches such as situation theory. A semantics with a ternary accessibility relation is not sufficient for a logic to be a relevant logic; many logics can have such a semantics.

The interpretations start from a common ground on a common purpose, however, they diverge in some points. The main one is that Israel-Perry information theory on which Mares's interpretation resides considers two situations that one involves the other, whereas Barwise's theory brings a channel construct between two situations which function as sites that contains information. Israel-Perry theory asserts a ternary relation, too, however, it is

again in the same manner of involvement relation. This difference is reflected in the notation; Restall uses $Rrsu$ instead of the conventional $Rsru$ where s is the linking element

8.2. Mares's Interpretation

Mares provides an interpretation of the Routley-Meyer semantics for a weak negation-free relevant logic, which he calls *Infologic*. Toward the interpretation, several conditions derived from an analysis of Israel-Perry's theory of information are stipulated such that the models of Infologic should satisfy. Infologic turns out to be the minimal logic that can be interpreted in this perspective, providing a starting point for stronger relevant logics.

The *involvement* relation between types of situations gives the constraints and thus, it is the basis of information flow. Let $\Sigma = [s : s \models \sigma]$, and $\Psi = [s : s \models \psi]$ be two types. Situation semanticists also use an abstraction operation. The only sort of abstraction that will be used here is *situational abstraction*. Suppose that σ is an infon. Then $[s] s \models \sigma$ is a situation type; it is the type of situations in which σ holds. We will use $s : \Sigma$ to mean that situation s is of type Σ .

If type Σ involves type Ψ , then every situation of type Σ , there is one of type Ψ , and since this is an infon, it is written as $\langle\langle \text{Involves}, \Sigma, \Psi; 1 \rangle\rangle$. Taking into account the closure of infons under the logical operations, conjunctions, disjunctions, and negation in situation semantics, let us denote this infon as the connection of two infons given in the definitions of Σ and Ψ . This connection is shown as $\sigma \rightarrow \psi$ which means " σ carries the information that ψ ." ' $\Rightarrow$ ' denotes the conditional in the standard logical sense.

In order to represent the relation *Involves* in Routley-Meyer semantics for relevant logic, the connective ' $\rightarrow$ ' ("carries the information that") is used which replaces situation types in the canonical model, and makes explicit the connection between the *Involves* relation and the theory of relevant implication.

A situation s supports the infon $\sigma \rightarrow \psi$, the truth condition for $\rightarrow$ is obtained as:

$$s \models \sigma \rightarrow \psi \quad \text{iff} \quad \langle\langle \text{Involves}, \Sigma, \Psi, 1 \rangle\rangle \in s$$

Every connective is interpreted with an accessibility relation (between situations) in the Routley-Meyer semantics.

The *Involves* relation for relevant implication, ternary accessibility relation R between situations are defined as

$$Rsru \quad \text{iff} \quad \forall \sigma \forall \psi (s \models \sigma \rightarrow \psi \Rightarrow (r \models \sigma \Rightarrow u \models \psi)).$$

On this interpretation, the situations are abstract; they consist in sets of infons. Therefore, an ordering relation, $\leq$, on situations, is also a part-whole relation. Some information is *persistent* (*hereditary* in relevant logic terms) if and only if, whenever it holds at a situation s , it holds at all situations of which s is a part. Assuming that no information is deteriorated by being including within larger situations, hence, that all information is persistent, $\leq$ turns out to be the subset relation on abstract situations:

Condition 1. $s \leq t$ iff $s \subseteq t$.

Since $\leq$ is a set inclusion in essence, it is transitive, reflexive, and antisymmetric. The second condition follows from the Routley-Meyer truth condition for implication, the Routley-Meyer ternary relation is interpreted as representing informational constraints:

$$\text{Condition 2. } s \models \sigma \rightarrow \psi \quad \text{iff} \quad \forall r \forall u (Rsru \Rightarrow (r \models \sigma \Rightarrow u \models \psi)).$$

$Rsru$ holds if s represents the world as being such that, if r is an accurate representation of the world, then u *might* also be an accurate representation of the world. Israel-Perry's claim that situation type Σ involves situation type Ψ if, according to s , if s represents the world as being such that “for every situation of type Σ , there is one of type Ψ ” within frame theory. Suppose that $\Sigma = [\dot{r} : \dot{r} \models \sigma]$ and $\Psi = [\dot{r} : \dot{r} \models \psi]$, if $s \models \sigma \rightarrow \psi$, then for all $r : \Sigma$, there is a $r' : \Psi$ such that $Rsrr'$.

The interpretation asserts that a ternary relation reflects the idea of situation semantics better than a binary relation does. A binary relation would treat “ $s \models \sigma \rightarrow \psi$ ” as “every situation that satisfies σ also satisfies ψ ,” therefore every situation of type Σ is stated to be a situation of type Ψ . Nevertheless, this interpretation conflicts with the aim of providing a theory of partial information. However, what we seek for is that $\langle\langle \text{Involves}, \Sigma, \Psi, 1 \rangle\rangle$ holds if and only if for every situation of type Σ there is a situation of type Ψ . This result induces that

all situations would contain the information about the causal history of all the infons in them, a case against partiality of information in a situation.

By the conception up to this point, it can be defined that an infon σ *semantically entails* ψ iff $\{s : s \models \sigma\} \subseteq \{s : s \models \psi\}$. This definition allows us to obtain the following are consequences, which leads to the assumption that informational constraints are closed under semantic entailment:

1. If $s \models \sigma \rightarrow \psi$ and ρ semantically entails σ , then $s \models \rho \rightarrow \psi$.
2. If $s \models \sigma \rightarrow \psi$ and ψ semantically entails ω , then $s \models \sigma \rightarrow \omega$.

If a situation s' has less information than s , then it contains fewer constraints. Thus, the ability of s' to determine which situations are true given particular situation r is less fine-grained the ability of s . Let $U = \{u : Rsr u\}$ and $U' = \{u' : Rs'ru'\}$. Then, $U \subseteq U'$. In ternary relation notation, which is one of Routley –Meyer's postulates:

Condition 3. If $s' \preceq s$ and $Rsr u$, then $Rs'ru$.

As for Dretske's Xerox Principle, it can be formulated in Israel-Perry's theory as:

If $\langle\langle \text{Involves}, \Sigma, \Psi; 1 \rangle\rangle \in s$ and $\langle\langle \text{Involves}, \Psi, \Pi; 1 \rangle\rangle \in s$,
then $\langle\langle \text{Involves}, \Sigma, \Pi; 1 \rangle\rangle \in s$.

Reformulating in frame semantics:

Condition 4. For all situations s , s is closed under the following rule:

$$s \models \sigma \rightarrow \psi$$

$$s \models \psi \rightarrow \pi$$

$$s \models \sigma \rightarrow \pi$$

In terms of ternary-relation:

If $Rsr u$, then there is a situation x such that $Rsr x$ and $Rsx u$.

The Xerox principle requires the *Involves* relation only to treat *full information* (in contrast to partial information): If $s \models \sigma \rightarrow \psi$, then given the information in s , whenever σ obtains, so must ψ . As a remedy, Mares proposes development of a semantics based on a probability measure on the set of worlds. However, as shown in [31], transitivity property holds at the level of signal tokens, but does not hold in general at the level of types. Therefore, the delineation of this principle should be reconsidered.

Factuality is adopted as the strong facticity condition: If a situation s is of type Σ and the infon $\langle\langle \text{Involves}, \Sigma, \Psi \rangle\rangle$ is in s , then s is also of type Ψ . Thus, any situation s is such that, if $s \models \sigma \rightarrow \psi$ and $s \models \sigma$, then $s \models \psi$. Therefore, situations are closed under modus ponens:

Condition 5. For all situations s , s is closed under the informational links:

$$s \models \sigma \rightarrow \psi$$

$$s \models \sigma$$

$$s \models \psi$$

The correlate of this condition in the ternary relation semantics: $Rsss$ for all situations s .

Condition 6. $s \models \sigma \wedge \psi$ iff $s \models \sigma$ and $s \models \psi$.

Hence, the truth conditions adopted for conjunction and disjunction in situation theory are the usual ones (however, this treatment of disjunction is controversial).

Condition 7. $s \models \sigma \vee \psi$ iff $s \models \sigma$ or $s \models \psi$.

The Logic of Information. The logic with the conditions on the model theory together with some postulates, which is complete over the class of models satisfying these postulates, is a weak relevant logic. Construction of a canonical model for Infologic will be used as an existence proof that models that satisfy all these conditions.

An *infoframe* is a quadruple $F = \langle \text{Sit}, L, R, \leq \rangle$ such that Sit is a non-empty set of situations, L is the truth set which is a subset of Sit , R is a ternary relation on Sit and $\leq$ is a partial ordering on Sit . The situations in L are abstract situations that are supposed to reflect

correctly a part of the actual world and where semantic entailment is conceived as implication; hence, they link to the information about the logical structure of situations in the actual world. Every infoframe satisfies the following postulates:

1. $\leq$ is transitive, reflexive, and antisymmetric.
2. If $s \leq s'$ and $Rs'ru$, then $Rsr u$.
3. If $Rsr u$, then there is some situation x such that $Rsr x$ and $Rsx u$.
4. $Rsss$.
5. $s \leq r$ iff there is a $u \in L$ such that $Rusr$.

The language consists of a set of primitive infons $i_1, i_2, \dots, i_n, \dots$ and three infon-forming connectives, $\wedge, \rightarrow, \vee$ and parentheses. Outer parentheses are dropped in accordance with the standard convention when no ambiguity results. The formation rules are the usual ones: If σ and ψ are infons, then so are $(\sigma \wedge \psi)$, $(\sigma \vee \psi)$, and $(\sigma \rightarrow \psi)$.

An *infomodel* is a structure $M = \langle F, \models \rangle$ in the same manner as any frame semantics, where F is an infoframe and $\models$ is a binary relation of a 'valuation' between situations and infons such that it assigns not truth values, but primitive infons to situations. For an infomodel, the following conditions hold:

1. $(s \leq r \wedge s \models p) \Rightarrow r \models p$.
2. $s \models \sigma \wedge \psi$ iff $s \models \sigma$ and $s \models \psi$.
3. $s \models \sigma \vee \psi$ iff $s \models \sigma$ or $s \models \psi$.
4. $s \models \sigma \rightarrow \psi$ iff $\forall x \forall y (Rsxy \Rightarrow (x \models \sigma \Rightarrow y \models \psi))$.

An infon σ is *info-valid* iff for all $s \in L$, $s \models \sigma$ for all infomodels.

In Hilbert-style presentation of the Infologic system, Mares gives the nuclear axioms and rules of relevant logic **B+** replacing affixing with prefixing and suffixing and adding the axioms:

1. $((\sigma \rightarrow \psi) \wedge (\psi \rightarrow \phi)) \rightarrow (\sigma \rightarrow \phi)$ (transitivity)
2. $((\sigma \rightarrow \psi) \wedge \sigma) \rightarrow \psi$
3. $((\sigma \rightarrow \phi) \rightarrow ((\sigma \wedge \psi) \rightarrow \phi))$ (strengthening the antecedent)

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MENKUZİ

As it will be examined below, the additional axioms of transitivity, and strengthening the antecedent are controversial for relevant logics.

8.3. Restall's Interpretation

The accessibility relation R is interpreted as channels denoted by ' $\rightarrow$ ' that models information flow between situations. Channels themselves are also situations.

Definition 1. A *Routley-Meyer frame* is a quadruple $\langle g, W, R, \sqsubseteq \rangle$ satisfying the following conditions:

1. $\sqsubseteq$ is a partial ordering on W .
2. If $Rwuv$ and if $w' \sqsubseteq w$, $u' \sqsubseteq u$ and $v \sqsubseteq v'$ then $Rw'u'v'$ also.
3. $g \in W$. Rwg iff $w \sqsubseteq v$.

where g is the truth set for R . Intuitively interpreting, $\sqsubseteq$ is informational inclusion.

Definition 2. An *evaluation* $\models$ on a Routley-Meyer frame $\langle g, W, R, \sqsubseteq \rangle$ is a relation between worlds and formulae such that

- E1. Hereditariness: If $w \sqsubseteq v$ then if $w \models p$ then $v \models p$ for atomic propositions p .
- E2. $w \models \phi \wedge \psi$ iff $w \models \phi$ and $w \models \psi$.
- E3. $w \models \phi \vee \psi$ iff $w \models \phi$ or $w \models \psi$.
- E4. $w \models \phi \rightarrow \psi$ iff for each u, v where $Ruwv$, if $w \models \phi$ then $w \models \psi$.

Example 1. Let us show that $\phi \rightarrow \phi \vee \psi$ is a theorem. By E4, we have

$$g \models \phi \rightarrow \phi \vee \psi \text{ iff for each } u, v \text{ where } Rugv, \text{ if } u \models \phi \text{ then } v \models \phi \vee \psi$$

By E3, we obtain $v \models \phi \vee \psi$ iff $v \models \phi$ or $v \models \psi$. Since $Rugv$ iff $u \sqsubseteq v$, and $u \models \phi$, then $v \models \phi$, $v \models \phi$ or $v \models \psi$ holds.

Example 2. Let us show that $\phi \rightarrow (\psi \rightarrow \psi)$, which is a paradox of implication, is *not* a theorem. As in Example 1, we obtain:

$$\begin{aligned} g \models \phi \rightarrow (\psi \rightarrow \psi) & \text{ iff for each } u, v \text{ where } Rugv, \text{ if } u \models \phi \text{ then } v \models \psi \rightarrow \psi \\ v \models \psi \rightarrow \psi & \text{ iff for each } w, z \text{ where } Rwvz, \text{ if } w \models \psi \text{ then } z \models \psi \end{aligned}$$

Taking $w = u$, and $z = u^*$, that is Routley-star companion of u . Then, $v \models \psi \rightarrow \psi$ fails, and so does the whole expression.

Note that $\psi \rightarrow \psi$ alone is valid, since this means $g \models \psi \rightarrow \psi$ and, if $u \models \psi$ then $v \models \psi$ for such u and v that $u \sqsubseteq v$ (that is, what information is contained in u is contained also in v) by the accessibility relation R on u, v , and g .

Definition 3. An *information structure* $\langle S, T, C, \models, \mapsto, \parallel, ; \rangle$ is a structure made up of a set S of situations, T of types, C of channels, a binary relation $\models$ which relates both pairs of situations and types and pairs of channels and constraints and a ternary relation $\mapsto$ channels to pairs of situations, ' $\parallel$ ' is the parallel composition and ';' is sequential composition of channels. " $a \mapsto_c b$ " is to be read as " c is channel from a to b ."

Definition 4. A *bare frame* $\langle g, S, \mapsto, \sqsubseteq \rangle$ is a Routley-Meyer frame in which W corresponds to a set of situations S , $Rwuv$ to $w \mapsto_u v$, and g is the *logic situation* with the additional condition of *associativity*:

$$(\exists v)(x \mapsto_y v \text{ and } v \mapsto_z u) \text{ iff } (\exists w)(x \mapsto_w u \text{ and } y \mapsto_z w).$$

The associativity condition derives from the requirement for a channel to support a conditional type and its formulation can be seen as the result of the following consecution, noting the point that any composition $(x; y)$ forms another channel:

1. Associativity holds in case that $(x; y); z \sqsubseteq u$ iff $x; (y; z) \sqsubseteq u$ for each u .
2. $(x; y); z \sqsubseteq u$ iff $x; y \sqsubseteq v$ and $v; z \sqsubseteq u$ for some v .
3. Conversely, $x; (y; z) \sqsubseteq u$ iff $x; w \sqsubseteq u$ and $y; z \sqsubseteq w$ for some w .

Restall takes situations as *prime*, hence, if a situation x is of type $A \vee B$ iff either x is of type A or x is of type B . There may be case in which two situations that disagree with regard to a disjunction $A \vee B$ can agree with regard to all conditionals. Because situations are prime, a minimal situation cannot be determined in such a case. This is particularly problematic in the serial composition of two situations, because $a \mapsto_b c$ is identified with $a; b \sqsubseteq c$, where c is the minimal situation satisfying the composition.

Preserving distribution of conjunction over disjunction, there is a two-tiered collection of points (situations) under consideration; the whole class S and its special subset P of prime points (that is, for points in P , a disjunction is true iff a disjunct is). For each arbitrary point $x \in S$, there is at least one $y \in P$ where $x \sqsubseteq y$. Although a disjunct may not be true given a body of information, it is possible for the information to be filled out, so that one or other disjunct is obtained. Then, a disjunction is evaluated in this model as follows:

$$x \models A \vee B \text{ iff for each } y \in S \text{ where } x \sqsubseteq y, y \models A \text{ or } y \models B.$$

Upon this discussion, it appears to be more suitable to take a channel as a set of situations $\{x : a \mapsto_b x\}$. This is formalized as the *cone* of situations defined as:

Definition 5. Given a bare frame $\langle g, S, \mapsto, \sqsubseteq \rangle$

1. $X \subseteq S$ is a *cone* iff for each $x \in X$, if $x \sqsubseteq y$ then $y \in X$.
2. If X, Y and Z are cones, $X \mapsto_Y Z$ iff for every $z \in Z$, there are $x \in X$ and $y \in Y$ where $x \mapsto_y z$.
3. If X and Y are cones, $X \sqsubseteq Y$ iff $Y \subseteq X$. In addition, $X; Y = \{z : X \mapsto_Y z\}$, $X \parallel Y = \{z : X \sqsubseteq Y \text{ and } Y \sqsubseteq z\}$.
4. For each situation x , $\uparrow x$ is the principle cone on x . In other words, $\uparrow x = \{x' : x \sqsubseteq x'\}$.

Lemma 1. Given a bare frame $\langle g, S, \mapsto, \sqsubseteq \rangle$ with an evaluation $\models$, the following holds:

1. $X \models \phi \rightarrow \psi$ iff for each pair of cones Y, Z , where $Y \mapsto_X Z$, if $Y \models \phi$ then $Z \models \psi$.
2. $X \models \phi \rightarrow \psi$ iff for each pair of situations y, z , where $\uparrow y \mapsto_X \uparrow z$, if $y \models \phi$ then $z \models \psi$.

3. $X \models \phi \wedge \psi$ iff $X \models \phi$ and $X \models \psi$.
4. $X \models \phi \vee \psi$ iff for each $x \in X$, either $x \models \phi$ or $x \models \psi$.
5. $\uparrow x \sqsubseteq Y$ iff for each $y \in Y$, $x \sqsubseteq y$.
6. $X \sqsubseteq \uparrow y$ iff $y \in X$.
7. $(\exists v)(X \mapsto_Y v \text{ and } v \mapsto_Z U)$ iff $(\exists w)(X \mapsto_w U \text{ and } Y \mapsto_Z w)$.
8. $\uparrow x \sqsubseteq \uparrow y$ iff $x \sqsubseteq y$.
9. $\uparrow x \mapsto_{\uparrow y} \uparrow z$ iff $x \mapsto_y z$.
10. $\uparrow x \models \phi$ iff $x \models \phi$.

Lemma 2. For all cones X and Y , and for all situations a and c , $a \mapsto_{X;Y} c$ iff there is a situation b such that $a \mapsto_X b$ and $b \mapsto_Y c$.

Proof. Suppose that $a \mapsto_{X;Y} c$. Then for some $d \in X;Y$, $a \mapsto_d c$. However, if $d \in X;Y$ we must have an $x \in X$ and a $y \in Y$ where $x \mapsto_y d$. So, $x \mapsto_y d$ and $a \mapsto_d c$. This means...

Lemma 3. In any bare frame, for any cones X , Y and Z , $X;(Y;Z) = (X;Y);Z$.

Proof. If $w \in X;(Y;Z)$ then $X \mapsto_{Y;Z} w$, which means that for some $x \in X$ and $v \in Y;Z$, $x \mapsto_v w$. But this means that for some u , $x \mapsto_y u$ and $u \mapsto_Z w$, which gives $u \in X;Y$ and so, $w \in (X;Y);Z$ as desired. The converse proof is the proof run backwards.

Definition 5. A bare frame $\langle g, S, \mapsto, \sqsubseteq \rangle$ is an *information frame* iff for each a, b, x, y where $a \mapsto_x b$ and $a \mapsto_y b$, there is a z where $x \sqsubseteq z$ and $y \sqsubseteq z$ and $a \mapsto_z b$. *Parallel composition* in an information frame is defined as $X \parallel Y = \{z: X \sqsubseteq z \text{ and } Y \sqsubseteq z\}$.

Lemma 4. In any information frame, for all cones X and Y , and for all situations a and b : $a \mapsto_X b$ and $a \mapsto_Y b$ iff $a \mapsto_{X \parallel Y} b$. Note that $X \parallel Y = X \cap Y$.

It is trivial to show that $X \parallel X = X$, $X \parallel Y = Y \parallel X$, and $X \parallel (Y \parallel Z) = (X \parallel Y) \parallel Z$ holds.

The relation $\preceq$ of refinement [4] is simply $\supseteq$:

$$X \preceq Y \text{ iff } X \parallel Y = X, \text{ iff } X \cap Y = X \text{ iff } X \subseteq Y \text{ iff } X \supseteq Y.$$

Lemma 5. In any information frame, if $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$, then $X_1; Y_1 \subseteq X_2; Y_2$.

Proof. By the hereditary conditions on $\rightarrow$:

8.4. Complementary Arguments to the Models

As we have seen, a crucial point for situation semantics is the adoption of the Austinian descriptive and demonstrative contents. In Barwise's information channels model, this point for a conditional statement is met by a constraint $\phi \rightarrow \psi$ as the descriptive content of the conditional, and a channel c as its demonstrative content. Since situations can be demonstrative contents of any statement, a situation s can be a channel between situations. Frame semantics satisfies this requirement of information flow model.

Relevant logics eliminate the paradoxes of implication, however, the models based on relevant logics examined are still prone to the problematic cases as transitivity and strengthening the antecedent rules. Let us consider the following statements and the unqualified application of transitivity:

If the doorbell is short-circuited, then the doorbell will ring.

If the doorbell rings, then someone is at the door.

Therefore,

If the doorbell is short-circuited, then someone is at the door,

which is an untenable result.

However, taking into account the channel conditions, the issue is resolved. In order to apply the transitivity, channels must be constructed such that one's target is the other's signal. In the situation that the doorbell is short-circuited, the constraint in the second antecedent is, in Restall's words, "filtered out," ringing does not indicate someone's being at the door any further, hence the target of the former has become a pseudo-signal for the latter.

This discussion shows that the channel modeling concepts are complementary to the arguments of the relevant logical models.

CHAPTER 9

CONCLUSION

Relevant logics have brought forward a fruitful perspective to study the intensional aspects of logic, and are improvement over the classical logic in this respect. When integrated with a proper semantics, relevant logics can provide sound solutions to the problems that demand incorporation of a coherent view of the subject matter to which logic is applied formally.

As the main parameters that a semantics is expected to satisfy, we can identify the following:

- *Explanation of the connectives*: Semantics should provide an intuition for the comprehension of the system beyond the proof-theoretic rules.
- *Conformity*: Proof-theoretic and semantic behavior of the connectives should conform to each other, while syntactically distinct connectives or logical constants should be described in sufficient granularity.
- *Intuitive insight*: So far as possible, whether an assertion holds or not in the model should be seen rather than seeking a proof.

A semantic system composed of frame semantics and situation semantics on an information-theoretic ground can be said to satisfy these requirements, while keeping accordance with the relevantistic approach. Thus, it can be claimed that a picture responding to the various issues, particularly in information flow and logic, occurs.

The constituents support each other making the picture satisfactory in various respects. However, some problems from the constituent subjects are still inherited. We may collect the problems under two headings: Those that are related to the definitional features of the information notion, and those that arise from the higher-level behavioral description of information.

The definition of information depends on concepts from several topics, each of which is a field of study by itself. The controversial points of these concepts remain controversial in the information notion as well. For example, the probabilistic basis of information is sound so long as the possible alternatives are definite, but this leaves out the problematic cases of probability theory (consider, for instance, non-repeatable single- event cases). On another side, causality is already a hot topic of philosophy, and so its connection to the informational relation that we have examined in the context of Dretske's model.

The higher-level problems come mainly from the tensions and assumption in situation theory. For example, it is hardly avoidable to employ such a concept as *full information* (in contrast to partial information) as in the case of Xerox principle, since we have to limit the transitive progression from an infon to another at some point in order to satisfy partiality. Nevertheless, that point is left entirely to pragmatical determination.

Another problematic case is due to the assumption of the prior existence of a channel in the model of information flow. Under what conditions a channel exists is in need of an account. The distinction between a simple connection from an information source to the receiver and a connection grounded on a constraint (hence, a law-like connection) demands explanation in situation-theoretic channel model.

There are open questions for frame semantics and relevant logics as well. For instance, the disjunctive syllogism is still a controversial issue.

However, observing that the studies on such difficulties bring improvement and innovation rather than being locked to a loop, we can say that there are sufficient prospects in order to respond positively to the arising questions and the constituents of the "picture" function as a test for each other. On the other side, such fields as semiotics provide considerable insights for those questions.

In general, the relation between logic and information is promising. Information not only provides a sound semantic basis, but also is apt to bear formal constraints. As the investigations in logic and information theory go deeper, we can expect that relation will become more satisfactory and stronger.

REFERENCES

- [1] Aisbett, Janet, and Gibbon, Greg, "A Practical Measure of the Information in a Logical Theory," *Journal of Experimental and Theoretical Artificial Intelligence*, #11 (1999), pp. 201-217.
- [2] Akman, Varol and Tin, Erkan, "Computational Situation Theory," (on line). Accessible in 2002: <http://www.cs.bilkent.edu.tr/~akman/jour-papers/sigart/sig.html>.
- [3] Anderson, Alan R., and Nuel, Belnap D., *Entailment: The Logic of Relevance and Necessity* (vol. 1). Princeton: Princeton University Press, 1975.
- [4] Barwise, Jon, "Constraints, Channels, and the Flow of Information," in *Situation Theory and Its Applications*, CSLI Publications, Stanford, 1993.
- [5] Barwise, Jon, "Conditionals and Conditional Information" in *The Situation in Logic*. Stanford: CSLI Publications, 1989.
- [6] Barwise, Jon, and Etchemendy, John, *The Liar: An Essay on Truth and Circularity*. New York: Oxford University Press, 1987.
- [7] Barwise, Jon, and Perry, John, *Situations and Attitudes*. Cambridge, Massachusetts: MIT Press, 1983.
- [8] Collier, John D., "Intrinsic Information," in *Information, Language and Cognition*, Hanson, Philip P. (ed). Vancouver: University of British Columbia Press, 1990.
- [9] Chaitin, Gregory J., "Gödel's Theorem and Information," (on line). Accessible in 2002: <http://www.umcs.maine.edu/~chaitin/georgia.html>.
- [10] Devlin, Keith, *Logic and Information*. New York: Cambridge University Press, 1991
- [11] Devlin, Keith, "Infons as Mathematical Objects," *Minds and Machines*, #2, 1992, pp. 185-201.

- [12] Devlin, Keith, "The Logical Structure of Computer-Aided Mathematical Reasoning," *American Mathematical Monthly*, August-September, pp. 632-646, 1997.
- [13] Dretske, Fred I., *Knowledge and the Flow of Information*. Massachussts: MIT Press, 1981.
- [14] Dretske, Fred I., "Précis of Knowledge and the Flow of Information" in *Naturalizing Epistemology*, Kornblith, Hilary (ed). Massachusetts: MIT Press, 1985.
- [15] Dretske, Fred I., "Putting Information to Work" Dretske, Fred I., in *Information, Language and Cognition*, Hanson, Philip P. (ed). Vancouver: University of British Columbia Press, 1990.
- [16] Floridi, Luciano L., "Is Information Meaningful Data?" (on line). Accessible in 2002: <http://www.wolfson.ox.ac.uk/~floridi/papers.htm>.
- [17] Floridi, Luciano L., "Outline of a Theory of Strongly Semantic Information," (on line). Accessible in 2002: <http://www.wolfson.ox.ac.uk/~floridi/papers.htm>.
- [18] Floridi, Luciano L., *Philosophy and Computing: An Introduction*. London: Routledge, 1999.
- [19] Israel, David, and Perry, John, "What is Information?" in Hanson, Philip P. (ed), *Information, Language and Cognition*, University of British Columbia Press, Vancouver, 1990.
- [20] Hoffmeyer, Jesper and Emmeche, Claus, "Code-duality and the Semiotics of Nature" in Myrdene, Anderson, and Floyd, Merrell (eds.), *On Semiotic Modeling*, Mouton de Gruyter, Berlin, 1991.
- [21] Li, Ming, and Vitányi, Paul, *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed. New York: Springer, 1997.
- [22] Mares, Edwin D., "Relevant Logic and the Theory of Information," *Synthese* 109 (1997), pp. 345-360.
- [23] Mares, Edwin D., and Meyer, Robert K., "Relevant Logics" in Goble, Lou (ed), *Guide to Philosophical Logic*, Oxford: Blackwell Publishers, 2001.
- [24] Papoulis, Athanasios, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

- [25] Restall, Greg, *Introduction to Substructural Logics*. London: Routledge, 2000.
- [26] Restall, Greg, "Information Flow and Relevant Logics," in J. Seligman and D. Westerstahl (eds), *Logic, Language and Computation*, Volume 1, Stanford: CSLI, pp. 463-478.
- [27] Read, Stephen, *Relevant Logic*. Oxford: Basil Blackwell, 1988.
- [28] Seligman, Jerry, "Perspectives in Situation Theory" in Cooper, R., Mukai K., and Perry J. (eds), *Situation Theory and Its Applications* (Vol. 1), Stanford, CSLI Publications, 1990.
- [29] Shannon, Claude, and Weaver, Warren, *The Mathematical theory of Communication*. Urbana: University of Illinois, 1964.
- [30] Stonier, Tom, *Information and the Internal Structure of the Universe: An Exploration into Information Physics*. London: Springer-Verlag, 1990.
- [31] Wobcke, Wayne, "An Information-Based Theory of Conditionals," *Notre Dame Journal of Formal Logic*, Vol. 41, # 2, 2000, pp. 95-141.

APPENDIX

REVIEW OF SOME CONCEPTS IN PROBABILITY THEORY

Probability is a widely utilized theoretical notion that sustains strong linkages with its manifold pre-theoretical senses, thus, leading to a variety of its definitions. Statements about probabilities can be analyzed broadly under two kinds, which are related to *objective* (also called *ontological*) and *subjective* (also called *epistemic*) probabilities.

Objective probabilities designate claims about the *world*, whereas subjective probabilities designate claims about strengths of beliefs that individuals have about the world, in contrast to the world itself. In what follows, objective probabilities will be dealt with, since it is this kind that information theory primarily utilizes, while having extensive applications taking subjective probabilities into account. The following exposition is mainly based on [24].

Probability theory is concerned with phenomena that, explicitly or implicitly, can be modeled by an experiment with an outcome that is subject to uncertain circumstances. Hence, as the experiment is repeated under identical conditions, the output can differ in accordance with an underlying random phenomenon. When a random experiment is performed, the various outcomes are likely to arise. It is usually convenient to think of an experiment and its possible outcomes as defining a space and its points. Thus, with the k -th outcome of the experiment, a point that is called the sample point is associated, which is denoted by s_k . The set of all sample points corresponding to the aggregate of all possible outcomes of such a random experiment is called the *sample space*. An *event* is a subset of the sample space; it corresponds to either a single sample point or a set of sample points. Particularly, the entire sample space S is called the *certain event*; the null set $\emptyset$ is called the *impossible event*; and a single sample point is called an *elementary event*.

For example, the experiment may be the observation of the result of tossing a fair pair of coins. In this experiment, the possible outcome of a trial is from the sample space

$\{(H, H), (H, T), (T, H), (T, H)\}$, where each pair denotes the outcomes of the first and the second coins as 'H' for *heads* and 'T' for *tails* of a coin. By assigning a sample point to each of these possible outcomes, we have a one-dimensional sample space that consists of four sample points. Then, the elementary event that one of the coins shows heads and the other one tails is, $\{(H, T), (T, H)\}$ while the event that both coins show heads is $\{(H, H)\}$.

Definitions of Probability. The classical and relative-frequency definitions of probability are acceptable for many technological applications and such activities as games of chance. They respond to intuitive ideas about probability. However, they allow conceptual ambiguities and lack deductive character of the theory, in which the axiomatic definition is superior to meet the requirements of rigorous studies.

Classical Definition. According to the classical definition, the probability $P(A)$ of an event A is determined *a priori* without actual experimentation, given by the ratio

$$P(A) = n_A/n$$

where n is the number of possible outcomes and n_A is the number of outcomes *favorable* to the event A .

The main drawback of this definition is the ambiguities on the determination of the numbers n_A and n . It can be said in situation-theoretic terms that the *individuation* of the possible and favorable outcomes excessively varies with each cognitive agent.

Relative Frequency Definition. The probability of an event is intended to represent how frequently it is expected to occur in a sequence of trials of the experiment. For example, in the coin-tossing experiment, let event A represent the outcome of showing heads (H) and suppose that in n trials of this experiment, event A occurs n_A times. Then, the ratio n_A/n can be assigned to the event A , which is called the *relative frequency* of the event A . The relative frequency is a nonnegative real number less than or equal to one: $0 \leq n_A/n \leq 1$.

Thus, in case that event A occurs in none of the trials, $n_A/n = 0$. In case that event A occurs in all the n trials, $n_A/n = 1$.

The experiment is said to exhibit statistical regularity if for any sequence of n trials the relative frequency n_A/n converges to the same limit as n becomes large. Expecting naturally a statistical regularity, the probability of event A can be defined as

$$P(A) = \lim_{n \rightarrow \infty} (n_A/n)$$

The limit in the above equation should be regarded as a statement of the probability that a particular event A occurs in a long sequence of trials, rather than the mathematical limit operation. For example, in the tossing of a fair coin experiment, we may expect that as the number of tosses grows, approximately one half of them will show up heads.

By the relative frequency definition, the problems associated with *a priori* definitions are eliminated, and the theory is founded on observation. However, its use as the basis of a deductive theory is inadequate. If the definition is used to define $P(A)$, the limit must be accepted as a hypothesis, not as a number that can be determined experimentally. For example, in such cases as the statistical analysis of the stock market, it is impossible to guarantee that an experiment occurs under the identical, even similar, conditions.

Axiomatic Definition. A more improved approach is to state the requirements that any measure of probability is to have as axioms, while applying relative-frequency interpretations to justify them.

The prominent one of such an axiomatic systems, due to Kolmogorov, involves a sample space S of elementary events (outcomes), a class E of events, which is the power set of S , and a probability measure $P(\cdot)$ assigned to each event A in the class E . The axioms that are sufficient to deal with experiments in finite sample spaces are:

A1. $P(S) = 1$

A2. $0 \leq P(A) \leq 1$

A3. If $A + B$ is the union of two *mutually exclusive* events (hence, it is logically impossible for both to occur) in the class E , then $P(A + B) = P(A) + P(B)$.

A1 and A2 reflect the properties of probabilities discussed above with clear relative frequency interpretations.

A3 states that the probability of the disjunction (in set theoretical terms, union) of two mutually exclusive events is the sum of the probabilities of the individual events. For the relative frequency interpretation of A3, it can be observed that if event A occurs n_A times in n trials and event B occurs n_B times, then the union event " A or B " occurs in $n_A + n_B$ trials, since A and B can never occur on the same trial because of mutual exclusion:

$$n_{A+B} = n_A + n_B$$

$$n_{A+B}/n = n_A/n + n_B/n$$

which has a mathematical form of A3, which is also called *restricted disjunction rule*.

The *general disjunction rule* includes the case in which events A and B are *not* mutually exclusive, that is, they can occur together. Then the probability of the union event " A or B " equals

$$P(A+B) = P(A) + P(B) - P(AB)$$

where $P(AB)$ is the probability of the joint event " A and B " with the following relative frequency interpretation:

$$P(AB) = \lim_{n \rightarrow \infty} (n_{AB}/n)$$

A3 is a special case of the general disjunction rule: When A and B are mutually exclusive, $P(AB)$ is zero, and it reduces to A3.

Let us review two important propositions to depict a better picture of the subject:

Proposition 1. $P(\bar{A}) = 1 - P(A)$ where $\bar{A}$ (denoting "*not-A*") is the complement of event A , hence, the nonoccurrence of the event A . Note that the sample space S can be expressed as the union of an event and its complement:

$$S = A + \bar{A}$$

Then, by A1 and A3 (since A and $\bar{A}$ are mutually exclusive events)

$$1 = P(A) + P(\bar{A})$$

from which the theorem follows.

Proposition 2. If mutually exclusive events $A_1, A_2, \dots, A_n$ constitute a partition of a sample space:

$$A_1 + A_2 + \dots + A_n = S$$

then

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1$$

This is a generalization of Proposition 1:

$$P(A_1 + A_2 + \dots + A_n) = 1$$

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

When the n events are equiprobable, then the expression simplifies as

$$P(A_i) = 1/n, \quad i = 1, 2, \dots, n$$

Conditional Probability and Conjunction Rules. Suppose we perform an experiment that involves a pair of events A and B . Let $P(B|A)$ denote the probability of event B under the condition that event A has occurred. The probability $P(B|A)$ is defined by

$$P(B|A) = P(AB)/P(A)$$

where $P(AB)$ is the joint probability of A and B .

We justify the definition of conditional probability given in the above equation by presenting a relative frequency interpretation of it. Suppose that we perform an experiment and examine the occurrence of a pair of events A and B . Let n_{AB} denote the number of times the joint event AB occurs in n trials. Suppose that in the same n trials, the event A occurs n_A times. Since the joint event AB corresponds to both A and B occurring, it follows that n_A must include n_{AB} . In other words,

$$n_{AB}/n_A \leq 1$$

The ratio n_{AB}/n_A represents the relative frequency of B given that A has occurred. As n grows, this ratio equals the conditional probability $P(B|A)$:

$$\begin{aligned} P(B|A) &= \lim_{n \rightarrow \infty} (n_{AB}/n_A) \\ &= \lim_{n \rightarrow \infty} ([n_{AB}/n]/[n_A/n]) \end{aligned}$$

$$= P(AB) / P(A)$$

The conditional relation can be rearranged as the *general conjunction rule*:

$$P(AB) = P(B|A) P(A) = P(A|B) P(B)$$

Accordingly, we may state that the joint probability of two events may be expressed as the product of the conditional probability of one event given the other and the other's the elementary probability.

If the conditional probability $P(B|A)$ is simply equal to the elementary probability of event B ,

$$P(B|A) = P(B)$$

then the probability of occurrence of the joint event AB is equal to the product of the elementary probabilities of the events A and B :

$$P(AB) = P(A) P(B)$$

which gives the *restricted conjunction rule*. Note that

$$P(A|B) = P(A)$$

that is, the conditional probability of event A , assuming the occurrence of event B , is simply equal to the elementary probability of event A . Hence, the two events are such that the occurrence of one has no effect on the occurrence of the other, from another view, the occurrence of one event provides no more information about the probability of occurrence of the other event. Such events are said to be (*statistically*) *independent*.

Bayes's Theorem. If mutually exclusive events $A_1, A_2, \dots, A_n$ constitute a partition of a sample space, hence,

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1$$

and B is an arbitrary event, then we can derive the following expression upon the above discussion

$$P(A_i|B) = [P(B|A_i) P(A_i)] / [P(A_1|B) P(A_1) + \dots + P(A_n|B)P(A_n)]$$

which is known as *Bayes's theorem*. Here, the terms $P(A_i)$ and $P(B|A_i)$ are referred to as *a priori* and *a posteriori* probabilities, respectively.

