Aus dem Departement für Informatik
Universität Freiburg (Schweiz)

Logic and Information

A Unifying Approach to Semantic Information Theory

Inaugural–Dissertation

zur Erlangung der Würde eines *Doctor scientiarum informaticarum*
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Freiburg in der Schweiz

vorgelegt von

Jutta Langel

aus

Deutschland

Nr. 1657
UniPrint, Freiburg
2009

Von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Freiburg in der Schweiz angenommen, auf Antrag von

- Prof. Dr. Jürg Kohlas, Universität Freiburg, Schweiz
- Prof. Dr. Jérôme Mengin, Université Paul Sabatier, Toulouse, Frankreich
- Prof. Dr. Ulrich Ultes-Nitsche, Universität Freiburg, Schweiz
- Prof. Dr. Beat Hirsbrunner, Universität Freiburg, Schweiz (Jurypräsident)

Freiburg, 10.12.2009

Der Dissertationsleiter:

Prof. Dr. Jürg Kohlas

Der Dekan:

Prof. Dr. Titus Jenny

© 2009 by Jutta Langel
All Rights Reserved

Pragmatically speaking, a question sets up a choice-situation between a set of propositions, namely, those propositions which count as answers to it.

Charles Leonard Hamblin (1922 - 1985)
Questions in Montague English

Acknowledgements

During the last five years and especially while writing this thesis, I was supported by many people. They have all contributed to the successful completion of this work.

First and foremost, I am grateful to Prof. Jürg Kohlas, the director of my thesis, who put confidence in me and motivated me to start this work. He supported me in finding the subject of this thesis and had time for many discussions to improve my work with ideas and constructive feedback. Many thanks go further to the (former) tcs and run research group members of the Universities of Fribourg and Berne for their various support and the friendly and stimulating working ambiance they created: Bernhard Anrig, Christian Eichenberger, Rolf Haenni, Jacek Jonczy, Reto Kohlas, Norbert Lehmann, Marc Pouly, Cesar Schneuwly and Michael Wachter. I am also thankful to Prof. Jérome Mengin and to Prof. Ulrich Ultes-Nitsche who accepted to act as referees for my thesis.

Also, I am greatly indebted to Christian Eichenberger who proof-read important parts of this thesis and who helped to improve it by pointing out lots of details. Furthermore, I thank Andrea Stengel and Heidemarie Langel for linguistic advice. Chocolat Villars also merits recognition for confecting chocolate and coffee supporting me constantly during the last five years.

My friends Ale, Andrea, Anita, Christian, Christian, David, Lionel, Manuel, Markus, Moritz, Nadine, Nathalie, Nicole, Sarah, Susa, Ursula and Valérie made my life beyond the thesis enjoyable and I am very thankful for that. Last but not least, my cordial thanks are addressed to my family. Alexis, Heidemarie and Ulrich, Helma, Laurent and Lena were always there for me and backed me up in every situation.

Abstract

The commonly used information theory, going back to Shannon, is almost exclusively concerned with the measure of information. However, by measuring information, one does not get very much to know about its nature, about what information actually is. Therefore, this thesis has two main goals:

- The first is to provide an adequate definition of the concept of information, by a semantic interpretation of the abstract, axiomatic information algebra framework. This leads to the formulation of an algebraic theory of semantic information, which is exemplified by logics.
- The second is to validate this theory, by comparison with already established semantic information theories (which also apply to logics) of other disciplines.

The Algebraic Theory of Semantic Information

An information algebra is a two-sorted algebra, consisting of a set of possible pieces of information, of a lattice of questions and of two operations (combination and focusing), that satisfy a set of five axioms. The following properties of information and questions are formalized by the information algebra framework: pieces of information refer to related questions; pieces of information can be focused, in order to extract information relative to some specific question of interest; pieces of information may be combined (aggregated).

Information can be perceived in two ways. One can look at how information is represented, or one can examine what information expresses. Information representation involves a (formal) language. In order to understand the meaning of the information, semantics is needed. In this thesis, we have chosen the latter approach. A semantic interpretation of the information algebra framework allows to draw conclusions about the nature of information and questions: A semantic piece of information is a set of possibilities, which may be interpreted in different ways. A piece of information is perceived as an answer to a question. A question, in turn, is semantically given by its possible answers. These results constitute the algebraic theory of semantic information, which applies to many instances, including logics. Propositional logic and predicate logic are shown to be information algebra instances. Since the meaning of information matters, the proofs are given on the semantic level.

A Unifying Approach

In disciplines which are also dealing with logics, like philosophy or linguistics, semantic information theories can be found, too. Three of them are presented in this thesis.

1. Carnap and Bar-Hillel's theory of semantic information from the early 1950s provides a very basic framework for semantic information and its measure. Information is perceived as a set of excluded possibilities. The theory is instantiated by a restricted monadic predicate logic language.

2. Groenendijk and Stokhof's theory of the semantics of questions and the pragmatics of answers provides a framework for questions. This theory has been developed in the early 1980s and has been extended by van Rooij in the first decade of the 21st century. As questions are identified with their possible answers, a detailed description of the nature of answers is also given. Two instances, propositional and predicate logic, are considered.

3. Barwise and Seligman's theory of information flow, which came up in the late 1990s, provides a framework for the representation of information and the computation with it in distributed systems. As to the representation, information is seen from a dual perspective, taking into account syntax and semantics. This dual representation is identified in this thesis with the approach of formal concept analysis, which was introduced in the 1980s. Barwise and Seligman mainly exemplify their theory by predicate logic.

All these three theories fit into the information algebra framework. Therefore, the algebraic theory of semantic information encompasses these theories.

Zusammenfassung

Die renommierte Informationstheorie der Informatik, die auf Shannon zurück geht, beschäftigt sich fast ausschliesslich mit dem Messen des Informationsgehalts von Nachrichten. Das blosse Messen sagt jedoch wenig über die Beschaffenheit von Information aus. Die Frage, was Information wirklich ist, bleibt offen. Deshalb verfolgt diese Dissertation zwei Hauptziele:

- Zum einen soll eine adäquate Definition von Information auf konzeptueller Ebene gegeben werden. Diese Definition beruht auf einem abstrakten, axiomatischen Framework, genannt Informationsalgebra, das aus einem semantischen Blickwinkel betrachtet wird. Daraus resultiert eine algebraische Theorie semantischer Information, die von Logik veranschaulicht wird.

- Zum anderen soll die algebraische Theorie semantischer Information validiert werden. Dazu wurden drei semantische Informationstheorien aus anderen Disziplinen ausgewählt, dargestellt und mit der algebraischen Theorie semantischer Information verglichen. Logik dient als Beispiel für alle vier Theorien.

Die algebraische Theorie semantischer Information

Eine Informationsalgebra ist eine zweisortige Struktur, die aus einer Menge von möglichen Informationen und einem Verband von Fragen besteht sowie aus zwei Operationen (Kombination und Fokussierung), die fünf Axiome erfüllen. Information hat Eigenschaften, die mit Hilfe des Informationsalgebra-Frameworks formalisiert werden können: Information bezieht sich auf Fragen, Fragen wiederum stehen untereinander in einer gewissen Beziehung. Um Information zu extrahieren, wird sie auf eine bestimmten Frage fokussiert. Da Information meist aus verschiedenen Quellen stammt und teilweise unvollständig ist, muss sie kombiniert werden um einen Gesamteindruck zu vermitteln.

Information kann auf zwei verschiedene Arten betrachtet werden. Das Interesse kann entweder auf die Darstellung oder die Bedeutung der Information gerichtet sein. Zur Darstellung der Information wird eine (formale) Sprache eingesetzt. Die Auseinandersetzung mit der Bedeutung der Information bedarf jedoch der Semantik, die Gegenstand dieser Dissertation ist. Die semantische Auslegung des Informationsalgebra-Framework erlaubt folgende Rückschlüsse über die Beschaffenheit von Information und Fragen: Semantische Information ist durch eine Menge von verschiedenen Möglichkeiten gegeben. Information wird als Antwort auf eine (implizite) Frage wahrgenommen. Eine Frage wird semantisch durch mögliche Antworten beschrieben. Diese Erkenntnisse führen zur algebraischen Theorie semantischer Information, die auf viele Formalismen, u. a. auf Logik, angewandt werden kann. Es wird gezeigt, dass Aussagenlogik und Prädikatenlogik Informationsalgebren bilden. Da das Gewicht auf der Bedeutung von Information liegt, werden die dazugehörigen Beweise auf semantischer Ebene ausgeführt.

Ein allumfassender Ansatz

Auch in anderen Disziplinen, die Logik einsetzen, wie z. B. in der Philosophie oder der Sprachwissenschaft, finden sich semantische Informationstheorien. Drei dieser Theorien werden in dieser Dissertation vorgestellt.

1. In den frühen 50er Jahren des 20. Jahrhunderts entwickelten Carnap und Bar-Hillel eine semantische Informationstheorie, die sich sehr grundlegend mit semantischer Information und ihrer Messung beschäftigt. Als Beispiel dient ihnen eine beschränkte, monadische prädikatenlogische Sprache.

2. In den frühen 80er Jahren des 20. Jahrhunderts haben Groenendijk und Stokhof eine Theorie über die Semantik von Fragen und die Pragmatik von Antworten entwickelt. Diese Theorie wurde zu Beginn des 21. Jahrhunderts von van Rooij erweitert. Sie setzt eine Frage mit ihren möglichen Antworten gleich und bietet somit auch eine detaillierte Beschreibung der Beschaffenheit von Antworten. Aussagenlogik und Prädikatenlogik werden im Rahmen von Beispielen betrachtet.

3. In den späten 90er Jahren des 20. Jahrhunderts entstand die Theorie des Informationsflusses von Barwise und Seligman, die hauptsächlich auf Prädikatenlogik angewandt wird. Die Theorie des Informationsflusses dient der Darstellung von Information und ihres Transports innerhalb von verteilten Systemen. Barwise und Seligman haben eine duale Sichtweise von Information, die sowohl Syntax als auch Semantik berücksichtigt. Ihre Vorgehensweise wird in dieser Dissertation mit dem Ansatz der formalen Konzeptanalyse aus den 80er Jahren in Verbindung gebracht.

Das Informationsalgebra-Framework ist eine Abstraktion dieser drei Theorien. Aus diesem Grund stellt die algebraische Theorie semantischer Information einen allumfassenden Ansatz für die drei Theorien dar.

Contents

1
Introduction

The important thing is not to stop questioning; curiosity has its own reason for existing. One cannot help but be in awe when contemplating the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of the mystery every day. The important thing is not to stop questioning; never lose a holy curiosity.

Albert Einstein (1879 – 1955)
Statement to William Miller, as quoted in LIFE magazine, 2 May 1955

The main tasks in computer science are information processing, storage, transmission and extraction. However, it is not known what information actually is in the discipline of computer science. No adequate definition of the concept of information has been given so far, information theory is restricted to the measure of the information content of a message. Therefore it is interesting to investigate the concept of information to clarify the scientific foundations of computer science.

1.1 Motivation & Purpose

1.1.1 The Origins of Information Theory

In the year 1948 Claude Shannon published his landmark paper "A Mathematical Theory of Communication". Shannon was working for the Bell Telephone Laboratories as an electronic engineer and mathematician, and intended to solve some practical communication problems with his paper (Shannon, 1948). But in addition, he laid the foundation of the information theory in computer science and is nowadays known as the father of information theory. The first, fundamental theorem of information theory, stating how much information can be transmitted over a channel, is due to Shannon. It was his idea to measure the information content of a message by the reduction of uncertainty, which results from learning the message at the receiver site. Shannon proposes to measure the uncertainty about the outcome of a message by an entity which he calls entropy. In 1971 he is cited in the Scientific American magazine (volume 225, page 180) with the following words:

> *My greatest concern was what to call it. I thought of calling it "information", but the word was overly used, so I decided to call it "uncertainty". When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage."*

These days, Shannon's information theory is widely applied in information and communication technology, where information is coded for the storage or transmission of data and for this purpose compressed. His theory has also been further developed, so there is actually a theory of information in computer science. But this information theory is almost exclusively concerned with measuring information. This is already a striking achievement. However, by measuring information one does not get to know very much about either information processing or the meaning of information. This is the motivation for and the raison d'être of this thesis.

1.1.2 The Origins of the Algebraic Theory of Information

So, what is information, and how is it processed? These questions are answered by the observation and description of certain facts, which lead to a formal definition of the nature of information. It all started in the field of probability theory. In the year 1988 Steffen Lauritzen and David Spiegelhalter showed in a pioneering work how to compute marginals of a multidimensional probability density by means of a specific computation scheme called "local computation" (Lauritzen & Spiegelhalter, 1988). Based on this fundamental paper, Prakash Shenoy and Glenn Shafer proposed an abstract, axiomatic framework, which is sufficient to enable local computation schemes (Shenoy & Shafer, 1990a). They discovered that many other formalism also satisfy the axioms needed for local computation. In the literature of inference in artificial intelligence and elsewhere the framework defined by Shenoy and Shafer is very often implicitly used, but without explicitly referring to it. Motivated by this fact, Jürg Kohlas proposed a slightly changed and further developed version of the axiomatic formulation of Shenoy and Shafer in (Kohlas, 2003) and called it "information algebra". In particular, Kohlas pointed out that information is always idempotent, which leads to a mathematical structure describing very basic properties of information from an algebraic point of view, as well as from a computational one. Many instances of Kohlas' algebraic framework are important in computer science, like relational data bases, constraint systems or various logics. Uncertainty formalisms, including Bayesian networks, possibility theory, or belief functions, are also covered. The information algebra framework can therefore justifiably be seen as an extension of the existing information theory in computer science, as it describes the nature of most of the common formalisms for representing information and thereby provides a definition of information and its processing.

1.1.3 Semantic Information

When a computer scientist is asked what information is the answer will be most of the time something like: "A sequence of zeros and ones". But this is only the representation of information in a form that allows it to be processed by computers. However, when we, as human beings, deal with information, we always do it in a semantic way. We reproduce a text that we have read, or something we have heard, not word-for-word, but we transmit its meaning. In the same way, when we want to share with another person something we have seen or that has happened, we will describe it by words. What really matters is the meaning of the information, not its representation. Therefore, we develop an algebraic theory of semantic information in this thesis, which results from a semantic interpretation of the information algebra framework.

1.1.4 Is Such a Theory Really Needed?

Information is the main matter of computer science. This discipline is often also called "informatics", which shows that information is what it is all about. One may object that so far remarkable results have been produced in computer science, without knowing what information actually is. But the question "What is information?" is not uninteresting, it rather has never been asked. The research on information was simply continued in the direction set by Shannon. The algebraic theory of semantic information, however, is not so much interested in measuring information, but aims at identifying those properties, which are important for providing a concept of information and its processing. The ambitious idea is that understanding the algebraic properties of information helps to improve the research in the other fields of computer science, so that even better, more stable, more efficient and more useful systems can be conceived one day.

1.1.5 Link to Other Disciplines

The algebraic theory of semantic information makes it possible to link computer science to other disciplines. In this thesis, we exemplify the theory by propositional and predicate logic, both being information algebra instances. In disciplines, which are also dealing with logics, semantic information theories can be found, too. We have chosen three of them and present them in this thesis. Our algebraic theory of semantic information turns out to be a unifying approach to these theories, as our theory could be shown to be a generalization of them. This establishes a strong link between computer science and disciplines, which are not at first sight thought to be related to it, like philosophy or linguistics.

1.1.6 Logic and Information

Propositional and predicate logic, two formalisms, which are covered by the algebraic theory of semantic information, are considered in this thesis on the semantic level. At first glance, this approach might not be very familiar. Logic is often regarded in the literature from a proof theoretic point of view only, which requires working on the syntactic level. But we are interested in the meaning of the information, which is expressed by the logical sentences. The semantic approach allows to determine where the information is localized, when making use of a formal language like propositional or predicate logic.

1.2 Thesis Outline

We now give a detailed outline of the four main parts of this thesis.

Part I: The Algebraic Theory of Semantic Information

In the first part, the algebraic theory of semantic information is set up, based on the information algebra framework.

In Chapter 2 it is shown what we mean by semantic information. An informal introduction to the concepts which are behind the following Chapters 3 to 7 is provided. An example guides through this chapter, in order to illustrate the natural algebraic structure of semantic information.

In Chapter 3, questions are introduced. They make up a basic element of the algebraic theory of semantic information. This chapter provides a mathematical structure that models questions. The fundamental properties of questions, which lead to a lattice structure, are explained. Refinements between questions are linked to partitions. It is stressed that a question is given by its possible answers.

In Chapter 4, the information algebra framework is introduced in its labeled version, where each piece of information pertains to some specific question. Different axiomatics – in terms of marginalization or variable elimination – are discussed. The transport operation is introduced. Particular cases of information algebras, namely Boolean information algebras and atomic information algebras, are defined.

In Chapter 5, the information algebra framework is introduced in its domain-free version, postulating that every piece of information tells something about every question, so pieces of information are looked at in a global way. Domain-free information algebras are best used for understanding theoretical concerns. It is sketched how a labeled information algebra can be transformed into a domain-free one and vice versa.

In Chapter 6, further properties of semantic information are given: relativity of information, partial order of information, contradictory and atomic information. These properties are very natural in the sense that everybody agrees with them, even people without a mathematical background. It is shown how these properties are anchored in the information algebra framework.

In Chapter 7, the disjunctive and the conjunctive interpretation of information is explained. For both points of view, a qualitative measure of information is given, which is based on the partial order of pieces of information. A quantitative measure, which applies the measures proposed by Hartley and Shannon, is also provided for both ways of interpreting a piece of information.

Part II: Two Information Algebra Instances

In the second part, two formalisms, which exemplify the information algebra framework, are shown.

In Chapter 8, the syntax and the semantics of propositional logic are provided. It is shown that propositional logic forms an information algebra. The proof is carried through on the semantic level and it is shown by isomorphism that the same follows for the syntactic level. The measures proposed in Chapter 7 are applied.

In Chapter 9, the syntax and the semantics of predicate logic are provided. Quantifier algebras, as an algebraization of predicate logic, are introduced, covering the syntactic and the semantic level of predicate logic. Predicate logic is shown to form an information algebra via the quantifier algebra formalism. The measures proposed in Chapter 7 are applied.

Part III: Semantic Information Theories

In the third part of this thesis, three semantic theories of information, and questions, respectively, are presented and compared to the algebraic theory of semantic information.

In Chapter 10, Carnap and Bar-Hillel's theory of semantic information is discussed. Their concept of semantic information is looked at in detail. A monadic predicate logic is introduced, which is the instance of their theory. The framework that Bar-Hillel and Carnap propose for measuring semantic information is presented, as well as two concrete measure functions. The algebraic theory of semantic information and Carnap and Bar-Hillel's theory of semantic information are compared at some length, first regarding the concept of information and afterwards relative to the measure of information.

In Chapter 11, Groenendijk and Stokhof's theory of the semantics of questions is considered. Their idea of identifying a question with its possible answers, and thereby inducing a partition, is explained and applied to the cases of propositional and predicate logic. Answers are also looked at from a propositional and predicate logical point of view. The order of answers, as proposed by Groenendijk and Stokhof, is presented. It is followed by a measure of answers, coming from van Rooij, who extended Groenendijk and Stokhof's theory also by a measure of questions. This chapter closes with a detailed comparison between the presented theory of questions and answers and our algebraic theory of semantic information. The comparison covers both fields, questions and pieces of informations (answers).

In Chapter 12, an overview of Barwise and Seligman's theory of information flow is given. The fundamental notions of classification and infomorphism are introduced, followed by the concept of a channel, which is a core concept of their theory. In order to establish a link between Barwise and Seligman's theory of information flow and our algebraic theory of semantic information, a slightly modified formulation of classifications is considered, namely contexts, which stem from formal concept analysis. It is shown that an information algebra can be associated with a context. Contexts are more convenient for information representation than Barwise and Seligman's classifications. Finally, a link between channels and the information algebra transport operation is established.

Part IV: Conclusion

In the fourth part, this thesis is concluded by showing that the algebraic theory of semantic information is a unifying approach to logic and information.

In Chapter 13, the main points of our algebraic theory of semantic information are summarized. It is shown to be a unifying approach to the three foregoing semantical information theories, which all apply to logics. Furthermore, it is shown, for each theory separately, in which sense the information algebra framework is a generalization of the respective theory. To conclude we compare, relate and confront the theories with each other.

In Chapter 14, a synopsis and a discussion of this thesis are given, including information theoretical aspects and the results related to logics. Open questions for future research are briefly summarized.

Part I

The Algebraic Theory of Semantic Information

2 Semantic Information

All our work, our whole life is a matter of semantics, because words are the tools with which we work, the material out of which laws are made, out of which the Constitution was written. Everything depends on our understanding of them.

Felix Frankfurter (1882 – 1965)
Associate Justice of the United States Supreme Court

When dealing with information, there are always two aspects: One is to look at its representation, the other one to consider its meaning. We need a formalism to express or describe what we are looking at. In computer science, data bases are probably the most standard way of information representation, but there are many other formalisms, such as classical logic formulae, linear equations etc. The representation formalism is the syntactic part of information. The other aspect of information is its meaning. Probably all current formalisms in computer science make use of the concept of variables and information is then given in a formal way by values of variables. Looking at the semantic part of information means being interested in what information tells about the values of the variables, especially of those values which are so far unknown. The semantic side of information thus corresponds to the values of variables and semantic information is provided using sets of tuples or sets of sequenes of values, as illustrated in Section 2.1. When information is uttered or received, it is in most of the cases the answer to some question. So questions are an inherent part of the algebraic theory of semantic information, see Section 2.2. The algebraic part of the theory is explained in Section 2.3 by pointing out the operations it relies on. This chapter provides an informal introduction to the concepts which are behind the following Chapters 3 to 7. An example will guide us through this chapter in order to illustrate the natural algebraic structure of semantic information. In the end in Section 2.4, the two main formalisms for capturing semantic information which were step by step introduced in this chapter are again shown in a condensed form, as they will be important in this whole thesis.

2.1 Information

Information may arise from different sources and is therefore considered to come piecewise. In the example considered in this chapter, the pieces of information may tell something about the values of a given set Vbl of variables which are denoted by v_1, v_2 and so on.[1] Each variable v_j, $j \in \mathbb{N}$, may take its values out of a certain set $\mathfrak{D}_{v_j}$, also called frame. There are two possible ways of considering semantic information, namely

1. in the local way, where we only deal with the values of the variables occurring in the actual situation, and
2. in the global way, where the values of all the variables of the whole system, i. e. all variables in Vbl, are looked at.

The first approach will be called labeled in the sequel, the second one is the domain-free point of view. Both methods are interchangeable. The first one is more appropriate for practical, computational purposes, the second one fits better for theoretical, abstract purposes.

Labeled semantic information is best captured by sets of tuples. A tuple is always relative to a finite set x of variables v_j, referred to as an x-tuple. An x-tuple attributes a value to every $v_j \in x$. So, if $x = \{v_1, \ldots, v_n\} \subseteq Vbl$, the semantic information arising from an x-tuple may be displayed as a sequence $\omega = \langle \omega_1, \ldots, \omega_n \rangle$, where $\omega_j \in \mathfrak{D}_{v_j}$, for $1 \leq j \leq n$. The set of all x-tuples is given by the Cartesian product of the frames associated with the variables in x. It is denoted by $\mathfrak{D}_x$ and defined as $\mathfrak{D}_x := \bigtimes_{v_j \in x} \mathfrak{D}_{v_j}$. In (Kohlas, 2003), the elements of such a frame $\mathfrak{D}_x$ are called configuration.

Domain-free semantic information takes all variables of Vbl into account. In contrast to the labeled approach, where only finite sets x are considered, Vbl may be countable. A sequence of values for all variables will be called valuation in the sequel and will be denoted by $\omega = \langle \omega_1, \omega_2, \ldots \rangle$, where $\omega_j \in \mathfrak{D}_{v_j}$. All possible valuations are given by the Cartesian product $\mathfrak{D}_{Vbl} = \mathfrak{D}_{v_1} \times \mathfrak{D}_{v_2} \times \ldots$. So a domain-free piece of information is a subset of $\mathfrak{D}_{Vbl}$.

Example 2.1.1 (Semantic Information) Consider a company looking for a new employee. There are four candidates invited to the interview, as shown by the data base of Figure 2.1.

We will attribute values to three variables v_1, v_2, v_3 and make use of two relations R_1 and R_2, in order to describe the above information semantically. The variable v_1 is a placeholder for the applicants' names, so $\mathfrak{D}_{v_1} = \{\text{Alice}, \text{Bob}, \text{Carol}, \text{Dave}\}$. Variable v_2 designates the candidates' age; they are all between 30 and 33 years old:

[1] Note, however, that this is a special case which has been chosen for an intuitive introduction to the subject. It will be generalized in the following chapters.

name	age	affiliation	marital status	# children
Alice	30	external	married	0
Bob	30	external	single	2
Carol	33	internal	single	1
Dave	31	internal	single	0

Figure 2.1: Information given in a relational data base

$\mathfrak{D}_{v_2} = \{30, 31, 32, 33\}$. Finally, the number of children is given by variable v_3 with $\mathfrak{D}_{v_3} = \{0, 1, 2\}$. So $Vbl = \{v_1, v_2, v_3\}$.

In order to get an overall picture of the whole system, let us first look at the global way of considering semantic information, using the second, domain-free approach. Figure 2.2 below illustrates the information stored in the data base from Figure 2.1: v_1, v_2 and v_3 are situated on the x-, y- and z-axis and their respective possible values are indicated. Every possible valuation is given by one node. All 48 nodes together form $\mathfrak{D}_{Vbl}$, the set of all valuations. The actual values of the variables, as given in the above information base, are shown by the black nodes: For example, the fact that Alice is 30 years old and does not have any children is provided by the lowest leftmost node, which is the valuation $\langle \text{Alice}, 30, 0 \rangle$. The properties that a candidate is married or not (= single) and that she or he is already employee of the company or not, are expressed by relations, i. e. sets of valuations. The relation R_1 designates the person's marital status, whereas relation R_2 describes whether the applicant is internal or external. In both cases, we are in a yes-no-situation which is captured by the fact that a node may or may not be in such a set, i. e. does or does not fulfill the relation. In Figure 2.2, R_1 is given by the set marked with a dotted line, R_2 by the one delimited by a dashed line.

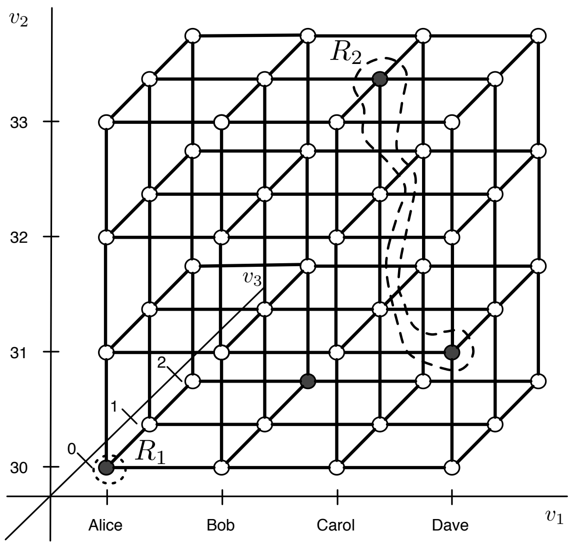


Figure 2.2: Visualization of domain-free semantic information

The other approach which was presented above is the labeled way of considering semantic information, using x-tuples. The advantage of x-tuples is that not all the information of the present system has to be taken into account, but only the details relating to x we are interested in.[2] Suppose that the human resources manager cares with regard to the job interview particularly about the applicants' names and their age. From a semantic point of view, the information she is interested in is given by $\{v_1, v_2\}$-tuples, and the information of the data base from Figure 2.1 corresponds to the following set of tuples: $\{\langle \text{Alice}, 30\rangle, \langle \text{Bob}, 30\rangle, \langle \text{Carol}, 33\rangle, \langle \text{Dave}, 31\rangle\}$. The relations R_1 and R_2 which were introduced above are translated to sets of tuples: $R_1 = \{\langle \text{Alice}, 30\rangle\}$ and $R_2 = \{\langle \text{Carol}, 33\rangle, \langle \text{Dave}, 31\rangle\}$. $\odot$

Summing up, we have seen that semantic pieces of information are provided by sets of values of variables, either by sets of tuples (labeled approach) or by sets of valuations (domain-free point of view), which may or may not be member of a set expressing a certain property.

2.2 Questions

Everything that should be said about the semantic nature of questions has already been said by the Hamblin quotation preceding this thesis. A question allows us to choose between different propositions that we call pieces of information. Each of these pieces of information is an answer to the question. So questions are represented by their possible answers. A straightforward realization of this idea is given in Section 2.2.1, where the labeled approach is chosen for understanding the semantics of questions. The domain-free way of capturing the semantics of questions is more sophisticated, as depicted in Section 2.2.2. Note that this is just an informal introduction to the wide field of questions, which will be investigated in depth in Chapter 3.

2.2.1 Questions in the Labeled Approach

The best way of understanding the semantics of questions in the labeled case is by means of an example.

Example 2.2.1 (Question, labeled) Continuing the example of the foregoing section, one may ask the question *"Who will get the job?"*. So we are interested in the candidates' names, especially in the person's name who will get the job, and we have the choice between 4 persons. In other words, the actual value of the variable v_1 is still unknown, but the 4 possible values are already given. The set of possible answers to our question is $\{\langle \text{Alice}\rangle, \langle \text{Bob}\rangle, \langle \text{Carol}\rangle, \langle \text{Dave}\rangle\}$. Each answer is a $\{v_1\}$-tuple and the set of possible answers is the set of all $\{v_1\}$-tuples provided by the

[2] As the above system is very small, it is actually no problem to look at all the given information, but in huge systems composed of hundreds of variables, it is definitely useful to utilize x-tuples.

system. Semantically, the question *"Who will get the job?"* is represented by the set of possible answers it allows. As this corresponds to the set of all $\{v_1\}$-tuples, we say that the question is given by the variable v_1.

Another question might be *"How old is the person who will get the job and how many children does he/she have?"*. This is a combined question, asking for the value of the variables v_2 and v_3. The set of possible answers to this question is given by the Cartesian product $\mathfrak{D}_{v_2} \times \mathfrak{D}_{v_3} = \{\langle 30, 0\rangle, \langle 30, 2\rangle, \langle 33, 1\rangle, \langle 31, 0\rangle\}$. Each possible answer is a $\{v_2, v_3\}$-tuple, and the question is said to be given by the set $\{v_2, v_3\}$ of variables. $\odot$

Questions are represented by their possible answers. An answer is a piece of information, which is semantically captured by a set of tuples, possibly a singleton, as in the above example. The common feature of all possible answers to a question is that they are related to the same set x of variables, so they are all x-tuples. This is why in general, a question is given by a set of variables x, which can be seen as a placeholder for all x-tuples, describing the question semantically.

2.2.2 Questions in the Domain-Free Approach

In the domain-free approach, a question is also given by a set x of variables. The possible answers are determined by the frames $\mathfrak{D}_{v_j}$ for all the variables $v_j \in x$, but they do not correspond to the Cartesian product $\bigtimes_{v_j \in x} \mathfrak{D}_{v_j}$, as in the labeled case.

Example 2.2.2 (Question, domain-free) In the above example on labeled questions, we have seen that the question *"Who will get the job?"* means asking for the value of the variable v_1. The same holds for the domain-free case. It is however important to take care that only the information about the name of the applicant is provided, and not more than this. So the answer that Alice is the one who gets the job is not just the valuation $\langle \text{Alice}, 30, 0\rangle$, since, in that case, we would also ascribe values to v_2 and v_3. This can only be avoided by listing all possible values for v_2 and v_3, expressing that their values are still unknown, but the one of v_1 is fixed to Alice. So one possible answer to the question v_1 is consequently the set of valuations

$$\begin{array}{llll} \{\langle \text{Alice}, 30, 0\rangle, & \langle \text{Alice}, 31, 0\rangle, & \langle \text{Alice}, 32, 0\rangle, & \langle \text{Alice}, 33, 0\rangle, \\ \langle \text{Alice}, 30, 1\rangle, & \langle \text{Alice}, 31, 1\rangle, & \langle \text{Alice}, 32, 1\rangle, & \langle \text{Alice}, 33, 1\rangle, \\ \langle \text{Alice}, 30, 2\rangle, & \langle \text{Alice}, 31, 2\rangle, & \langle \text{Alice}, 32, 2\rangle, & \langle \text{Alice}, 33, 2\rangle\}. \end{array}$$

Similar sets have to be constructed for Bob, Carol and Dave. Those 4 sets then make up the set of possible answers to the question v_1 (*"Who will get the job?"*), which is a set of sets of valuations. $\odot$

As already mentioned, a question is given by a set x of variables, and the semantic nature of a question is captured by the set of its possible answers, as in the labeled approach. The example shows that such a possible answer consists of a set of valuations which agree in the values of the variables asked for.

This leads to the following consideration: A question x expresses an equivalence relation π_x over $\mathfrak{D}_{Vbl}$. Recall that $\mathfrak{D}_{Vbl}$ is the Cartesian product of all frames of all variables and thus contains all the possible sequences of values which may be attributed to the variables. What the equivalence relation π_x does is to group the valuations together by the values they provide relative to the question x. Two valuations $\omega, \theta \in \mathfrak{D}_{Vbl}$ are equivalent relative to a question x (denoted by $\omega \equiv_x \theta$) if the values of the variables in x are the same, i. e. if $\omega_j = \theta_j$ for all $v_j \in x$. $\mathfrak{D}_{Vbl}$ is therefore partitioned into sets of equivalent valuations $[\omega]_{\pi_x} = \{\theta \in \mathfrak{D}_{Vbl} : \theta \equiv_x \omega\}$. A set of equivalent valuations will be called block in the following. The block or equivalence class $[\omega]_{\pi_x}$ contains all those valuations which have the same values for the variables in x. The set of all such blocks is denoted by $\mathfrak{D}_{Vbl}/\pi_x$. Each $[\omega]_{\pi_x} \in \mathfrak{D}_{Vbl}/\pi_x$ is a possible answer to the question x. Such an answer is exhaustive and assigns a unique value to each variable in question. Thus, a question induces a partition of $\mathfrak{D}_{Vbl}$ and may be seen as the set of possible exhaustive answers it allows.

Example 2.2.3 (Partition) In the our example there are three variables: v_1, which may take four possible values, v_2, which may also take four possible values, and v_3, which may take three possible values. As $\mathfrak{D}_{Vbl}$ consists of all possible valuations, there are $4 \cdot 4 \cdot 3 = 48$ valuations. They correspond to the nodes in the coordinate system and will be grouped together according to a given question. Figure 2.3 shows two possible partitions of $\mathfrak{D}_{Vbl}$, but there are lots of other ways of partitioning.

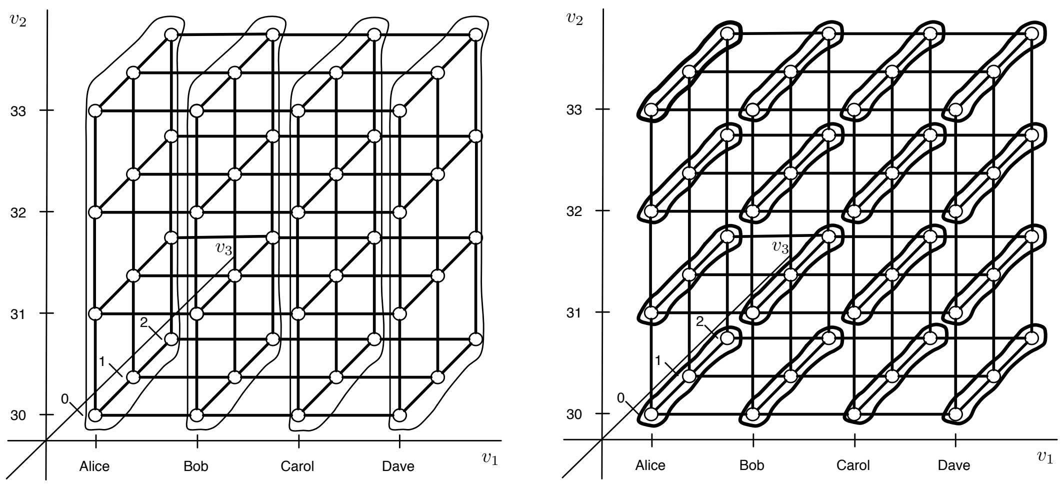


Figure 2.3: Partitions of $\mathfrak{D}_{Vbl}$ induced by the questions $\{v_1\}$ and $\{v_1, v_2\}$

When partitioning $\mathfrak{D}_{Vbl}$ relative to the variable v_1, those valuations which have the same v_1-value form a block. For example, the valuations $\langle \text{Alice}, 30, 0 \rangle$ and $\langle \text{Alice}, 31, 2 \rangle$ are equivalent relative to v_1. But there are still a lot of other equivalent valuations, namely all those which fit into the scheme $\langle \text{Alice}, *_2, *_3 \rangle$, where $*_2$ and $*_3$ may be any values of $\mathfrak{D}_{v_2}$ or $\mathfrak{D}_{v_3}$, respectively. The set of all valuations which are equivalent to $\langle \text{Alice}, 30, 0 \rangle$ is denoted by $[\langle \text{Alice}, 30, 0 \rangle]_{\pi_{\{v_1\}}} \in \mathfrak{D}/\pi_{\{v_1\}}$. The situation is depicted on the left hand side of Figure 2.3: The question $\{v_1\}$ "*Who will get* *the job?*" may be exhaustively answered either by *Alice, Bob, Carol* or *Dave*. Since $|\mathfrak{D}_{v_1}| = 4$, there are four equivalence classes containing $|\mathfrak{D}_{v_2}| \cdot |\mathfrak{D}_{v_3}| = 4 \cdot 3 = 12$ valuations each. $\mathfrak{D}_{Vbl}$ is divided into four blocks of the same size, and equivalent valuations are regrouped together.

One may also ask a question about more than one variable, e. g. about $\{v_1, v_2\}$ "*Who is getting the job and how old is this person?*" In this case, $\mathfrak{D}_{Vbl}$ is partitioned relative to v_1 and v_2 which results in more, but smaller equivalence classes. $\langle \text{Alice}, 30, 0\rangle$ has only two equivalent valuations, namely $\langle \text{Alice}, 30, 1\rangle$ and $\langle \text{Alice}, 30, 2\rangle$. The scheme is obvious: The values of v_1 and v_2 are fixed, the value of v_3 varies. The question may be exhaustively answered by $|\mathfrak{D}_{v_1}| \cdot |\mathfrak{D}_{v_2}| = 4 \cdot 4 = 16$ different pieces of information. Each answer or equivalence class contains $|\mathfrak{D}_{v_3}| = 3$ valuations and is delimited on the right hand side of Figure 2.3 by a bold line. $\odot$

Summing up, we have seen that different questions may induce different partitions of the set of all valuations $\mathfrak{D}_{Vbl}$. Questions are represented by their possible answers and are given by sets of variables.

2.3 Operations

Until now, a semantic description of information has been given. Furthermore, it was shown how to capture the semantic nature of questions. So the first two sections of this chapter dealt with the semantics in our "algebraic theory of semantic information". But the algebraic part is still missing. It will be examined in this section. An algebra is a set together with a collection of operations on this set. In the foregoing sections, we already spoke about pieces of information. This is an indicator for the fact that usually information comes piecewise. The set of the algebra which will now come up is constituted of pieces of information. We will denote them by Greek lower case letters like ϕ, ψ, ζ. Associated with the set of pieces of information is a set D of questions. As seen before, a question is a set of variables. Questions will be denoted in the following by lower case latin letters like x, y, z. So we are actually dealing with a two-sorted algebra, called information algebra. In this section we informally introduce the operations involving pieces of information and questions. In Section 2.4, they are given in a more formal way. Depending on whether the labeled or the domain-free approach is considered (see Section 2.3.1 and 2.3.2, respectively), there are slightly different operations.

2.3.1 Operations in a Labeled Information Algebra

As seen above, the algebra we are dealing with is constituted of a set of pieces of information. It is therefore termed information algebra. When the pieces of information are given in a tuple fashion (see Section 2.1), the labeled way of representing information is chosen. In order to point out this property, we speak of a labeled information algebra. The set of pieces of information of a labeled information algebra will be denoted by Φ in this thesis. But there is not only the set Φ which constitutes a labeled information algebra. Each piece of information is also connected to a question. Questions have been introduced to be sets of variables. Thus the set D of questions is a set of sets of variables. Note that questions will even turn out to form a lattice, see Chapter 3, where also the restriction to sets of variables will be dropped. So a labeled information algebra is denoted by (Φ, D) and the following three operations are defined on it: labeling, combination and marginalization.

Labeling

Labeling is an operation specifying the question a piece of information $\phi \in \Phi$ pertains to. The operator is d and stands for a mapping from the set of all pieces of information Φ to the lattice of questions D, formally

$$d : \Phi \to D \quad \text{and} \quad \phi \mapsto d(\phi).$$

The element $d(\phi) \in D$ is called the domain of ϕ. For that reason, D is also referred to as the lattice of domains. Obviously, if ϕ is given by a set of x-tuples, where $x \in D$ is a set of variables, then $d(\phi) = x$.

Example 2.3.1 (Labeling) Let us continue our well-known example from the previous sections. The piece of information $\phi = \{\langle \text{Alice, 30} \rangle, \langle \text{Carol, 33} \rangle\}$ provides information relative to the names (described by variable v_1) and the age (captured by variable v_2) of some candidates. So $d(\phi) = \{v_1, v_2\}$, as ϕ is constituted of $\{v_1, v_2\}$-tuples. ⊙

Marginalization

When one is not interested in all the information one disposes of, but only in a given field of interest, marginalization comes into play. This operation is heavily used in our everyday life, even if we are not aware of it. The most intuitive example of marginalization might be an internet search request, using a search engine. In the beginning, one disposes of the whole information the internet provides, but by typing a key word, e. g. "*information algebra*", only the information relating to the key word will be made available. Clearly, the key word plays the role of the domain of the information we are interested in. So marginalization involves a piece of information $\phi \in \Phi$ and a domain $x \in D$ which contains less variables than the domain of ϕ, i. e. $x \subseteq d(\phi)$. We want to extract information from ϕ about the field of interest x. The result is the reduced information $\phi^{\downarrow x}$ which is again an element of Φ. The binary operation of marginalization is captured by

$$\downarrow : \Phi \times D \to \Phi \quad \text{and} \quad (\phi, x) \mapsto \phi^{\downarrow x}, \text{ for } x \subseteq d(\phi).$$

When marginalizing a piece of information $\phi \in \Phi$ to $x \in D$, the tuples making up ϕ are restricted to x. Consequently, $\phi^{\downarrow x}$ is composed of those x-tuples which have the same values for the variables in x as the tuples in ϕ. It is apparent that marginalization corresponds to projection in relational algebra.

Example 2.3.2 (Marginalization) Consider a piece of information ϕ which gives a detailed description of who is invited to the job interview. $\phi = \{\langle\text{Alice, 30, 0}\rangle, \langle\text{Bob, 30, 2}\rangle, \langle\text{Carol, 33, 1}\rangle, \langle\text{Dave, 31, 0}\rangle\}$ is obviously bearing on the domain $d(\phi) = \{v_1, v_2, v_3\}$. The human resources manager is only interested in the candidates' names. Her field of interest is described by the variable v_1. As $\{v_1\} \subseteq \{v_1, v_2, v_3\}$, marginalization can be performed on ϕ. The result $\phi^{\downarrow\{v_1\}}$ is obtained by dropping the values in the tuples relating to the other variables. Thus $\phi^{\downarrow\{v_1\}} = \{\langle\text{Alice}\rangle, \langle\text{Bob}\rangle, \langle\text{Carol}\rangle, \langle\text{Dave}\rangle\}$. $\odot$

There is a very natural relation between marginalization and questions: Each piece of information $\phi \in \Phi$ is related to some question $d(\phi)$ which means that it tells a fact about a certain domain or field of interest. If one is only interested in a subdomain of $d(\phi)$, then marginalization is used to obtain the information by extraction. Thus marginalization allows to relate ϕ to another question x, as long as $x \subseteq d(\phi)$.

Combination

Combination is a binary operation. It is used for information aggregation and maps to any two pieces of information $\phi, \psi \in \Phi$ the combined information $\phi \otimes \psi$ which is again in Φ. This is expressed by

$$\otimes : \Phi \times \Phi \rightarrow \Phi \quad \text{and} \quad (\phi, \psi) \mapsto \phi \otimes \psi.$$

The domain of the resulting piece of information is the union of the domains of the factors: $d(\phi \otimes \psi) = d(\phi) \cup d(\psi)$. Furthermore, if $d(\phi) \cup d(\psi) = x$, the combined piece of information $\phi \otimes \psi$ consists of x-tuples. Such a tuple is an element of $\phi \otimes \psi$ if it has the following two properties: marginalized to $d(\phi)$, it belongs to ϕ and marginalized to $d(\psi)$, it is contained in ψ. Here, it is already quite evident that the combination operation corresponds to the join in relational algebra.

In a sequence of pieces of information to be combined it does not matter in which order the combinations are performed, as long as the sequence is not changed. In other words, combination is associative:

$$(\phi \otimes \psi) \otimes \zeta \quad = \quad \phi \otimes (\psi \otimes \zeta).$$

Associativity allows to drop parentheses when several pieces of information are combined, so we can simply write $\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_{n-1} \otimes \phi_n$. Furthermore, the result of combining the first piece of information with the second is the same as doing it the other way round. This property is called commutativity:

$$\phi \otimes \psi \quad = \quad \psi \otimes \phi.$$

The fact that combination is associative and commutative is a very natural requirement when the aggregation of information is modeled. When adding several pieces of information, one expects to obtain always the same result, no matter in which order the information is combined.

Example 2.3.3 (Combination, labeled) Reconsider the job interview example. Suppose that someone got to know that two of the candidates are invited for a second interview, whereas the two others did not attract interest. This insider disposes of the information $\phi = \{\langle \text{Alice, 30}\rangle,\ \langle \text{Carol, 33}\rangle\}$, concerning their names and their ages ($d(\phi) = \{v_1, v_2\}$). Now somebody tells this person how the two candidates are called and how many children they have; this is the piece of information $\psi = \{\langle \text{Alice, 0}\rangle, \langle \text{Carol, 1}\rangle\}$ with $d(\psi) = \{v_1, v_3\}$. Combining both pieces of information results in $\phi \otimes \psi = \{\langle \text{Alice, 30, 0}\rangle,\ \langle \text{Carol, 33, 1}\rangle\}$ bearing on the domain $d(\phi) = \{v_1, v_2, v_3\}$. ⊙

Transport

There is actually a fourth operation defined on a labeled information algebra (Φ, D), called transport. As it can be derived from combination and marginalization, we stated above that there are only 3 operations necessary to set up a labeled information algebra. However, the transport operation is very meaningful, since it allows to relate a piece of information $\phi \in \Phi$ to any domain $x \in D$, and not only to those domains x which satisfy $x \subseteq d(\phi)$, as in the case of marginalization. Why is the transport operation so interesting? It allows us to consider the same piece of information under different questions, even under all possible questions that may be asked. Furthermore, the operation shows that the same piece of information may be represented equivalently relative to different domains. This leads to domain-free information algebras, which will be treated afterwards. The transport operation is formally described by

$$\rightarrow: \Phi \times D \rightarrow \Phi \quad \text{and} \quad (\phi, x) \mapsto \phi^{\rightarrow x}.$$

In order to simplify the definition of the transport of a piece of information from one domain to another, we will introduce a further operation, called vacuous extension:

$$\uparrow: \Phi \times D \rightarrow \Phi \quad \text{and} \quad (\phi, x) \mapsto \phi^{\uparrow x}, \text{ for } x \supseteq d(\phi).$$

The goal of this operation is to extend a piece of information $\phi \in \Phi$ to some domain x containing more variables than $d(\phi)$ without changing the meaning of ϕ. This is done by assuming all possible values for the variables which are in x, but not in $d(\phi)$. Vacuous extension is defined for any $x \supseteq d(\phi)$ by means of combination:

$$\phi^{\uparrow x} \quad := \quad \phi \otimes \mathfrak{D}_x.$$

Example 2.3.4 (Vacuous Extension) In the job interview example we could imagine a situation where the information $\phi = \{\langle \text{Alice}\rangle, \langle \text{Bob}\rangle\}$ is available. This is information relative to the candidates' name, so $d(\phi) = \{v_1\}$. If we want to extend ϕ vacuously to the domain $\{v_1, v_2\}$, this is done by combining it with the set of all $\{v_1, v_2\}$-tuples which results in $\phi^{\uparrow \{v_1, v_2\}} = \{\langle \text{Alice, 30}\rangle, \langle \text{Alice, 31}\rangle, \langle \text{Alice, 32}\rangle, \langle \text{Alice, 33}\rangle, \langle \text{Bob, 30}\rangle, \langle \text{Bob, 31}\rangle, \langle \text{Bob, 32}\rangle, \langle \text{Bob, 33}\rangle\}$. No information is added to ϕ, as all possible values of $\mathfrak{D}_{v_2}$ are assumed to hold for the age of the two candidates which are described by ϕ. ⊙

Vacuous extension, which does not change the information it is applied to, is now used in the definition of the transport operation. The marginalization operation is also involved in the definition, since transporting $\phi \in \Phi$ from $d(\phi)$ to some other domain x consists of extending ϕ vacuously to some domain which covers both $d(\phi)$ and x. Thereafter, this piece of information which still provides the same information as ϕ, but relative to more variables, is marginalized to the field of interest x:

$$\phi^{\rightarrow x} \quad := \quad (\phi^{\uparrow d(\phi)\cup x})^{\downarrow x}.$$

Note that the smallest domain which covers both $d(\phi)$ and x is their union. This is why in a first step, ϕ is vacuously extended to $d(\phi) \cup x$.

Example 2.3.5 (Transport) Now, the piece of information $\phi = \{\langle 30, 0\rangle, \langle 30, 2\rangle\}$ is given, which provides only information about the age of the candidates and their number of children, so $d(\phi) = \{v_2, v_3\}$. If we are now interested in what ϕ tells us about the names and the age of these candidates, we have to transport it to the domain $\{v_1, v_2\}$. The smallest common domain is $\{v_1, v_2, v_3\}$. Extending ϕ vacuously to the domain $\{v_1, v_2, v_3\}$ results in $\phi^{\uparrow\{v_1,v_2,v_3\}} = \{\langle \text{Alice, 30, 0}\rangle, \langle \text{Bob, 30, 0}\rangle, \langle \text{Alice, 30, 2}\rangle, \langle \text{Bob, 30, 2}\rangle\}$. This intermediate result can now be marginalized to the domain of interest: $(\phi^{\uparrow\{v_1,v_2,v_3\}})^{\downarrow\{v_1,v_2\}} = \phi^{\rightarrow\{v_1,v_2\}} = \{\langle \text{Alice, 30}\rangle, \langle \text{Bob, 30}\rangle\}$. When some piece of information provides only the age of some of the candidates and their number of children, but we are interested in their names and their age, the transport operation allows to bring this information into focus. $\odot$

Note that marginalization and vacuous extension can be seen as special cases of the transport operation: $\phi^{\rightarrow x} = \phi^{\downarrow x}$, if $x \subseteq d(\phi)$ and $\phi^{\rightarrow x} = \phi^{\uparrow x}$, if $x \supseteq d(\phi)$. But the latter operation furthermore allows to transport a piece of information ϕ to any domain.

2.3.2 Operations in a Domain-Free Information Algebra

From the previous section it is known that an information algebra consists of a set of pieces of information and a set of questions D and that operations are defined on them. In a domain-free information algebra, the set of pieces of information consists of sets of valuations, as introduced in the beginning of Section 2.1. The domain-free approach of representing information is based on the idea that a piece of information tells something about the values of every variable of the system. This is why a valuation provides a value for every variable considered. Therefore, there is no labeling operation in domain-free information algebras. It is nevertheless possible to relate a piece of information to a question $x \in D$. For that purpose, the focusing operation will be introduced. When aggregating information, the combination operation is used, producing again a domain-free piece of information. There will be only slight changes regarding the combination operation of Section 2.3.1.

The set of pieces of information of a domain-free information algebra will be denoted by Ψ in this thesis. The set of questions or domains is the same as before, so D

is still used to name it. As before, the elements of D are sets of variables. In the sequel, (Ψ, D) will always denote a domain-free information algebra, whereas (Φ, D) points out that we are dealing with a labeled information algebra. We now look in more detail at the operations defined on the two-sorted algebra (Ψ, D).

Focusing

In a domain-free information algebra, information extraction is done by focusing. Focusing involves a domain-free piece of information $\psi \in \Psi$ and a domain $x \in D$. There is not any more a restriction on x as in the case of marginalization, since ψ is domain-free and carries information about any domain in D. Focusing is similar to the transport operation in the labeled approach, this is why it is also denoted by a right arrow, but this time by a double right arrow $\Rightarrow$. It allows to extract information from ψ relative to any field of interest. Furthermore, by means of focusing, a piece of information may be related to every question that may be asked. The operation of focusing is given by

$$\Rightarrow: \Psi \times D \to \Psi \quad \text{and} \quad (\psi, x) \mapsto \psi^{\Rightarrow x}.$$

Focusing a piece of information $\psi \in \Psi$ on some domain $x \in D$ gives rise to a piece of information $\psi^{\Rightarrow x}$. A valuation of $\mathfrak{D}_{Vbl}$ is an element of $\psi^{\Rightarrow x}$ if it provides the same values for the variables of x as at least one of the valuations in ψ.

Example 2.3.6 (Focusing) In the current job interview example, consider the piece of information $\psi = \{\langle \text{Dave}, 31, 0\rangle, \langle \text{Dave}, 31, 1\rangle, \langle \text{Dave}, 31, 2\rangle\}$ which states that Dave is 31 years old. No information about his number of children is given (all possible values are assumed). ψ is depicted in Figure 2.4 by a solid line. Clearly, it refers to the question $\{v_1, v_2\}$.

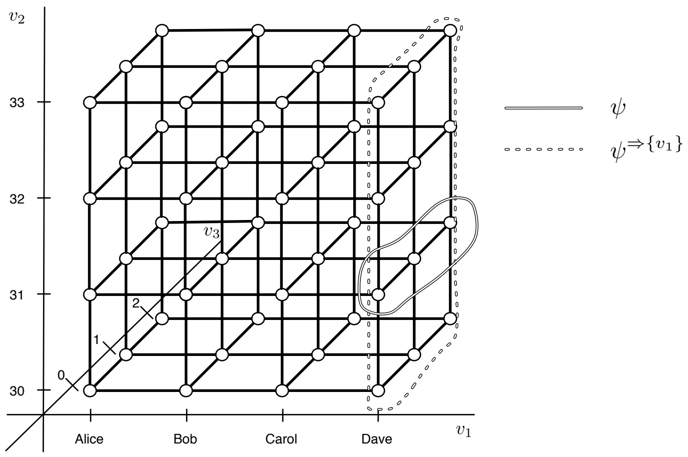


Figure 2.4: Focusing of information

In case we are not interested in the whole information provided, but only in a part of it, ψ can be focused on the field of interest. If one is only interested in the candidate's name and not in the age, ψ has to be focused on $\{v_1\}$. This means extending the set of valuations: v_1 = "Dave" is fixed by ψ; for the other variables, however, every possible value is assumed. $\psi^{\Rightarrow\{v_1\}}$ is depicted by a dotted line in the Figure 2.4. $\odot$

Combination

The combination of two domain-free pieces of information has the same motivation as in the labeled approach. It is also denoted by $\otimes$, but it is applied to domain-free pieces of information $\phi, \psi \in \Psi$:

$$\otimes : \Psi \times \Psi \to \Psi \quad \text{and} \quad (\phi, \psi) \mapsto \phi \otimes \psi.$$

Performing the combination of $\phi, \psi \in \Psi$ is very easy, as domain-free pieces of information are sets of valuations. The combined piece of information $\phi \otimes \psi$ is obtained by intersecting ϕ and ψ. The combination operation is again associative and commutative.

Example 2.3.7 (Combination, domain-free) Consider that there are two pieces of information given: ψ saying that "The person who is getting the job is 31 years old and has no children" and ϕ telling that "Dave is getting the job". In Figure 2.5, they are given by the solid and the dotted line, respectively.

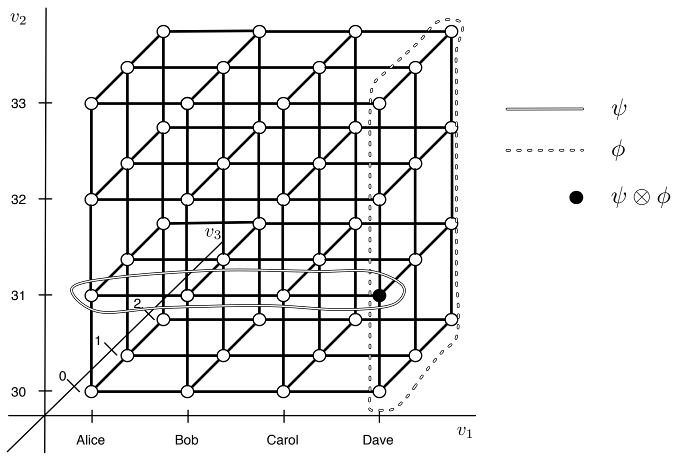


Figure 2.5: Combination of information

When combining both pieces of information, $\phi \otimes \psi$, we get an overall picture of the situation. ϕ tells that Dave is the selected person, but says nothing about his age and the number of his children. On the other hand, ψ states that a 31-year-old person with no children will get the job, where this very person may either be Alice, Bob, Carol or Dave. So when ψ is known, the values of v_2 and v_3 are unequivocally given. By combination with ϕ, the complete information about v_1 is added. The intersection of both pieces of information is taken, which results in the valuation $\langle \text{Dave}, 31, 0\rangle$, given by the black node in the above figure. Thus $\phi \otimes \psi = \{\langle \text{Dave}, 31, 0\rangle\}$. $\odot$

2.4 Information Algebras of Tuples and Valuations

In this chapter, an algebraic view on semantic information has been introduced. Labeled pieces of information have been represented by sets of tuples, domain-free pieces of information by sets of valuations. As these formalisms will be used again and again to capture semantic information, they are given here in compact form, together with the corresponding information algebra operations. See (Kohlas, 2003, Chapter 6.3) and (Kohlas & Schneuwly, 2009) for more details.

The information algebra framework, which is used for a general, abstract description of the nature of semantic information, its properties, its associated operations and its measure, consists of a set of pieces of information and a lattice D of questions or domains. A frame $\mathfrak{D}_x$ is related to each question $x \in D$. $\mathfrak{D}_x$ provides the possible answers to the question x. In the labeled case, the set of pieces of information is denoted by Φ, in the domain-free case by Ψ. As pieces of information are different in the two versions of information algebras, so are the frames, which are made up of the possible answers, which in turn are pieces of information. The same lattice D is considered in both cases, but the frames $\mathfrak{D}_x$ for $x \in D$ are not the same. There is a heavily used special case where the elements of D are finite subsets of a (possibly countable) set of variables Vbl. We will consider this case in this résumé, as it will be important for the comparison in the second and the third part of this thesis.

We will now look at semantic pieces of information in both information algebra versions. First, in Section 2.4.1, we examine the labeled case where semantic pieces of information are sets of tuples. Then, in Section 2.4.2, domain-free semantic pieces of information being sets of valuations are considered.

2.4.1 Sets of Tuples

Consider a set of variables $x = \{v_1, \ldots, v_{|x|}\} \subseteq Vbl$. Then, the corresponding frame $\mathfrak{D}_x$ may be seen as the Cartesian product of the variables in x. Assuming for technical reasons an ascending order regarding the indices in Vbl, we can write without loss of generality $\mathfrak{D}_x = \mathfrak{D}_{v_1} \times \mathfrak{D}_{v_2} \times \ldots \times \mathfrak{D}_{v_{|x|}}$.

A labeled piece of information $\phi \in \Phi$ is a set of tuples. An x-tuple is a function

$$f : x \to \mathfrak{D}_x. \tag{2.1}$$

Obviously, the frame $\mathfrak{D}_x$ is the set of all x-tuples. A labeled piece of information $\phi \in \Phi$ regarding x is a set of x-tuples $\phi \subseteq \mathfrak{D}_x$. The set of all x-tuples is

$$\Phi_x = \mathfrak{P}(\mathfrak{D}_x). \tag{2.2}$$

Then, the set of all pieces of information is

$$\Phi \subseteq \bigcup_{x \in D} \Phi_x. \tag{2.3}$$

For every domain $x \in D$, there is a neutral element $e_x \in \Phi$. It expresses ignorance with respect to $x \in D$. The neutral piece of information e_x is the whole frame $\mathfrak{D}_x$. Furthermore, there is a null element z_x for every domain $x \in D$. It is incompatible with any other piece of information of Φ and upon combination (see below), leads to a contradiction. For every frame $\mathfrak{D}_x$, the incompatible piece of information is the empty subset of $\mathfrak{D}_x$. One assumes by convention that every frame $\mathfrak{D}_x$ has its own empty subset $\emptyset_x$.

The labeling operation d is used to determine the question or domain a piece of information $\phi \in \Phi$ pertains to:

$$d(\phi) = x, \text{ if } \phi \subseteq \mathfrak{D}_x. \tag{2.4}$$

Restricting an x-tuple f to a coarser question $y \leq x$ involves a y-tuple g:

$$f[y] = g \text{ such that } g(v) = f(v) \text{ for all } v \in y. \tag{2.5}$$

This leads to the marginalization operation $\downarrow$ which serves for extracting information from $\phi \in \Phi$, regarding a coarser domain $y \leq d(\phi)$. It corresponds to the projection π of relational algebra.

$$\phi^{\downarrow y} \quad = \quad \{f[y] : f \in \phi\}. \tag{2.6}$$

The combination operation $\otimes$ is used for information aggregation and applied to two pieces of information ϕ and ψ with $d(\phi) = x$ and $d(\psi) = y$. It corresponds to the natural join $\bowtie$ of relational algebra.

$$\phi \otimes \psi \quad = \quad \{f \in \mathfrak{D}_{x \cup y} : f[x] \in \phi, f[y] \in \psi\}. \tag{2.7}$$

This defines the labeled information algebra (Φ, D) of semantic pieces of information as sets of tuples. The associated operations are labeling, marginalization and combination.

2.4.2 Sets of Valuations

In the domain-free approach, only the whole set Vbl of variables is considered. The corresponding frame is denoted by $\mathfrak{D}_{Vbl}$. We assume again an arbitrary order in Vbl and can therefore write without loss of generality $\mathfrak{D}_{Vbl} = \mathfrak{D}_{v_1} \times \mathfrak{D}_{v_2} \times \ldots$.

A domain-free piece of information $\psi \in \Psi$ is a set of valuations. A valuation is a function

$$f : Vbl \rightarrow \mathfrak{D}_{Vbl}, \tag{2.8}$$

where $\mathfrak{D}_{Vbl}$ is the universe or the set of all valuations. Domain-free pieces of information are sets of valuations and thereby subsets of $\mathfrak{D}_{Vbl}$. In particular, if the set Vbl of variables is finite, the set of all pieces of information is

$$\Psi \quad = \quad \mathfrak{P}(\mathfrak{D}_{Vbl}), \tag{2.9}$$

Otherwise, Ψ is only constituted of those subsets of $\mathfrak{D}_{Vbl}$ which refer to a finite subset of Vbl.[3] The neutral element $e \in \Psi$ is $\mathfrak{D}_{Vbl}$ itself. The null element of Ψ is $z = \emptyset$.

By means of the focusing operation $\Rightarrow$, a piece of information $\psi \in \Psi$ is considered relative to some question $x \in D$. This involves x-equivalent valuations. Two valuations f and g are said to be x-equivalent, denoted by $f \equiv_x g$, if they provide the same values for all variables $v_i \in x$. The set of all valuations which are x-equivalent to f is

$$f^{\Rightarrow x} = \{g \in \mathfrak{D}_{Vbl} : g \equiv_x f\}. \tag{2.10}$$

This leads to the focusing operation, which extracts the information of ϕ which is relative to $x \in D$:

$$\phi^{\Rightarrow x} = \bigcup_{f \in \phi} f^{\Rightarrow x}. \tag{2.11}$$

Two pieces of information ϕ and ψ of Ψ may be joined using the combination operation $\otimes$:

$$\phi \otimes \psi = \phi \cap \psi. \tag{2.12}$$

This defines the domain-free information algebra (Ψ, D) of semantic pieces of information as sets of valuations. Its associated operations are focusing and combination.

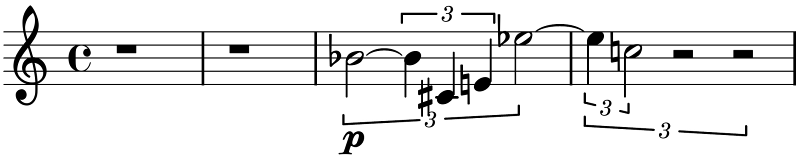
2.5 Conclusion

An information algebra is a two-sorted algebraic structure, involving a set of pieces of information and a set D of questions which are also called domains. The elements of D are finite subsets of the set Vbl of variables, for the time being. (Later in this thesis, this idea will be generalized.) The first component of information algebras, the set of pieces of information, has two possible forms of representation: labeled pieces of information and domain-free pieces of information. The set of labeled pieces of information is denoted by Φ, the set of domain-free pieces of information by Ψ. We look at pieces of information from a semantic point of view. In the labeled approach, this leads to a piece of information being a set of x-tuples, where x is a finite subset of Vbl, that is, a domain in D. The domain-free way of representing a piece of information is by a set of valuations. A valuation carries a value for every variable considered in a system, whereas an x-tuple only provides values for the variables in x. As every piece of information is related to a question, each question can be represented by its possible answers. This is best seen in the domain-free approach, where a question induces a partition of the set of all valuations. Different operations come along with an information algebra. In the labeled case, three operations are

[3] Later on in this thesis, pieces of information which have this property will be said to have a finite support.

defined on (Φ, D): labeling d, combination $\otimes$ and marginalization $\downarrow$. Labeling is used for determining the question a piece of information relates to, information aggregation is done by combination, and information can be extracted by means of marginalization. As in the domain-free approach, information is considered not related to a specific domain, but globally, involving the variables of the whole system, there are only two operations defined on (Ψ, D): combination $\otimes$ (for information aggregation) and focusing $\Rightarrow$ (for information extraction). The operations defined on an information algebra are very basic and are heavily used in our everyday life. Thus information algebras allow to give a formal description of the nature of information and its processing. In Chapters 4 and 5, sets of axioms are imposed on (Φ, D) and (Ψ, D), respectively, leading to a mathematic framework for describing the nature of information.

3 Modeling Questions



Charles Edward Ives (1874 – 1954)
The Unanswered Question, Trumpet Solo

A basic element of the algebraic theory of semantic information introduced in this first part of the thesis are questions. This chapters provides a mathematical structure that models questions. Section 3.1 presents some fundamental requirements we impose on questions leading to a lattice of questions. However, the properties presented are quite abstract. In the second part of this chapter, things get more concrete and refinements between questions are introduced which correspond to partitions (see Section 3.2.1). An important special case are partitions induced by sets of variables (see Section 3.2.2).

Let us first look at a single question. The best characterization of a question and its semantic impact is probably given in (Hamblin, 1973). As it is so convincing, we will cite again the quotation preceding this thesis:

> *[...] a question sets up a choice-situation between a set of propositions, namely, those propositions which count as answers to it.*

According to the above quotation, a question gives rise to several pieces of information. They can be seen as statements having in common that they may all be given as an answer to the same question. A question is thus characterized by the possible answers it allows. The set of possible answers will also be referred to as frame and will be denoted by Latin lower case letters like x, y, z in Section 3.1 and by Greek upper case letters like Ω, Λ, Γ in Section 3.2.[1]

[1] The reason for this notational difference is that in Section 3.1, as in most of this thesis, questions will only be looked at from the exterior, as entites having certain properties in relation with other

3.1 The Lattice of Questions

Usually we do not ask only one question, but often several of them. So it is very natural to consider a set containing all the questions one could imagine to ask in a given situation. This set is hereinafter known as D, its elements, the questions, will be denoted by x, y, z, as already seen in Chapter 2. The questions in D are somehow related. Clearly, there are questions in D about completely different subjects. But there are also questions which, regarding one subject, ask for more or less detailed answers. So there exists a certain granularity of questions. A question may be finer or coarser than another, depending on the level of detail of the possible answers one is aiming at. A finer question gives rise to finer answers which differentiate between more details than those answering a coarser question. Therefore, the frame of a finer question provides more possible answers than the frame of a coarser question. For two questions $x, y \in D$, we denote the fact that x is a coarser question than y (or, equivalently, that y is a finer question than x) by $x \leq y$. In that case y is asking for more detailed answers than x. So an order relation $\leq$ is established in D. The order theoretic properties of the set D are captured by a partial order:

Definition 3.1 (Partial Order)

Let D be a nonempty set. A partial order *on D is a binary relation $\leq$ on D such that, for all $x, y, z \in D$,*

1. $x \leq x$,
2. $x \leq y$ *and* $y \leq x$ *imply* $x = y$,
3. $x \leq y$ *and* $y \leq z$ *imply* $x \leq z$.

These three conditions are referred to as reflexivity, antisymmetry *and* transitivity, *respectively.*

Our set D of questions equipped with the binary order relation $\leq$ ("coarser question than") is said to be a partially ordered set, written $\langle D; \leq \rangle$. But there are further properties.

As already pointed out above, we may also want to ask two questions $x, y \in D$ at the same time. Asking two arbitrary questions $x, y \in D$ simultaneously means asking the question $z \in D$ which may be described in terms of the order relation $\leq$ as follows: Both $x \leq z$ and $y \leq z$ hold. Such a question z is finer than both x and y. It is called an upper bound of the set $\{x, y\} \subseteq D$. However, there are lots of questions in D which are upper bounds of the set $\{x, y\}$. Together they constitute the set $Z^{xy} = \{z \in D : x \leq z, y \leq z\}$. But one does not want to ask for more details than really required as this would lead to superfluous information. So we are actually interested in a very specific question z, namely the coarsest one of the

questions. However, in Section 3.2, we will examine the interior of questions and look at how a question is constituted, always in comparison with other questions.

whole set Z^{xy}. The question $z \in D$ we are looking for is consequently the $z \in Z^{xy}$ such that for all $z' \in Z^{xy}$ we have $z \leq z'$. In other words, it is the coarsest question which is finer than both x and y and is often denoted by $\sup\{x, y\}$. This specific question $z \in D$ is called least upper bound of the set of questions $\{x, y\}$. We prefer to use a neater notation also found in the literature and write $x \vee y$ (read as "x join y") in place of $\sup\{x, y\}$.

As there is an upper bound, one could also look for a lower bound. The lower bound of two arbitrary questions $x, y \in D$ expresses the common part of them. Dually, there are possibly several lower bounds of $\{x, y\}$, constituting the set $Z_{xy} = \{z \in D : z \leq x, z \leq y\}$ which contains all questions which are coarser than x and y at a time. Again, there is a specific question $z \in Z_{xy}$, which is the closest to x and y, so it is the finest question in Z_{xy}. This question z is referred to as greatest lower bound and has the property $z' \leq z$ for all $z' \in Z_{xy}$. Thus, it is the finest question which is coarser than both x and y, denoted by $\inf\{x, y\}$. As before, we favor another notation, namely $x \wedge y$ (read as "x meet y").

We now suppose that for two arbitrary questions $x, y \in D$, a coarsest question which is finer than both always exists, i. e. $x \vee y \in D$. Similarly, there is always a finest question which is coarser than any two question $x, y \in D$, so $x \wedge y \in D$. In other words, this amounts to say that D is a lattice.

Definition 3.2 (Lattice)

A partially ordered set $\langle D; \leq \rangle$ is a lattice *if $x \vee y$ and $x \wedge y$ exist for all $x, y \in D$.*

Obviously, the notions of (least) upper bound and (greatest) lower bound can be generalized to arbitrary subsets of D. So for a subset $T \subseteq D$, $\bigvee T$ is the least upper bound of the set T, the "join of T". Dually, the greatest lower bound of some set $T \subseteq D$ is denoted by $\bigwedge T$ and referred to as the "meet of T". Note that $\bigvee T$ and $\bigwedge T$ do not always exist for all $T \subseteq D$. But if they do, the lattice is even complete:

Definition 3.3 (Complete Lattice)

A partially ordered set $\langle D; \leq \rangle$ is a complete lattice *if $\bigvee T$ and $\bigwedge T$ exist for all $T \subseteq D$.*

A complete lattice can be obtained by supposing that our lattice $\langle D; \leq \rangle$ of questions has always a bottom element $\bot$, which has the property $\bot \leq x$, for all $x \in D$. It is thus the least element of the lattice which is coarser than every other question. We can furthermore assume the existence of a top element $\top$, i. e. $\top \geq x$, for all $x \in D$, which is the greatest element of the lattice and embodies the question which is finer than every other question. So we claim the existence of a finest question $\top$ and a coarsest question $\bot$, having the following properties for all $x \in D$: $\top \vee x = \top$, $\top \wedge x = x$, $\bot \vee x = x$, $\bot \wedge x = \bot$. In order to avoid such strong assumptions, we can consider another approach allowing to obtain a complete lattice. It is applied when the questions of D are finite subsets of a countable set (of variables, as in Section

3.2.2). Then the existence of $\perp$ and $\top$ is not required. But we can nevertheless find for every set T of questions, where T is a set of finite sets, a greatest lower and a least upper bound, which is again a finite set and thus in D.

Sometimes, one does not want to look at the whole set of all possible questions D, but only at a smaller set of questions $T \subseteq D$. So a sublattice $\langle T; \leq \rangle$ can be derived from $\langle D; \leq \rangle$ in the following way:

Definition 3.4 (Sublattice)

Let $\langle D; \leq \rangle$ be a lattice and $\emptyset \neq T \subseteq D$. Then $\langle T; \leq \rangle$ is a sublattice *of $\langle D; \leq \rangle$ if $x, y \in T$ implies $x \vee y \in T$ and $x \wedge y \in T$.*

The above definition says that a non-empty subset T of D which is closed under finite applications of $\vee$ and $\wedge$ is itself a lattice. The notion of complete sublattice can be defined analogous to Definition 3.3. See also (Davey & Priestley, 2002) and (Grätzer, 2003) for a more detailed presentation of ordered sets and lattices. Note furthermore that we do not require for the time being that the lattice of questions is distributive.[2]

Until now, we have seen that a single question is modeled by its possible answers. The set of possible answers a question allows is called a frame. When more than one question is looked at, the questions form a partially ordered set D which is considered together with the binary order relation $\leq$ expressing that a question is coarser than another, if it provides less detailed answers. We require from now on that the set D of questions forms a lattice, i. e. for any two questions in D, there is always a coarsest question which is finer than both and a finest question which is coarser than both in D. So this section established some formal requirements concerning the structure inside the set of all questions. In the next section we will look at how these rather abstract ideas are translated in concrete question structures.

3.2 Concrete Question Structures

The previous section has shown that there are coarser and finer questions. This is why we will now focus on coarsenings and refinements of questions. At the beginning of this chapter, a question was introduced by setting up a choice-situation between its possible answers. This is why a question will now be represented by the set of possible answers it allows. According to (Shafer, 1976), such a set of possible answers is called a frame of discernment, or simply frame, and will be denoted by a Greek upper case letter, like Ω, Λ, Γ. When we want to draw the attention to the possible answers of a question, frames Ω, Λ, Γ are used, in contrast to the preceding section where questions (denoted by x, y, z) as such are looked at. The content of this section is a short résumé of (Shafer, 1976, Chapter 6), or at least of those parts which are important for our semantic theory of algebraic information (of which questions are a basic element). Furthermore, it is based on (Kohlas & Monney, 1995, Chapter 7), which itself goes back to Shafer.

[2] The properties of a distributive lattice are given in Equations 3.8 and 3.9.

3.2.1 Compatible Frames of Discernment

Let us start with an introductory example taken from (Kohlas & Monney, 1995). It is an adaptation of the "car that will not start" example described by Shafer.

Example 3.2.1 Suppose your car does not start any more. You want to determine the cause of its failure, i. e. get an answer to the question "Why does my car not start?". This processes is depicted by a so-called tree of diagnoses, see Figure 3.1, copied from (Kohlas & Monney, 1995). The tree structure already shows that the diagnoses-nodes are mutually exclusive and collectively exhaustive, i. e. exactly one is correct.

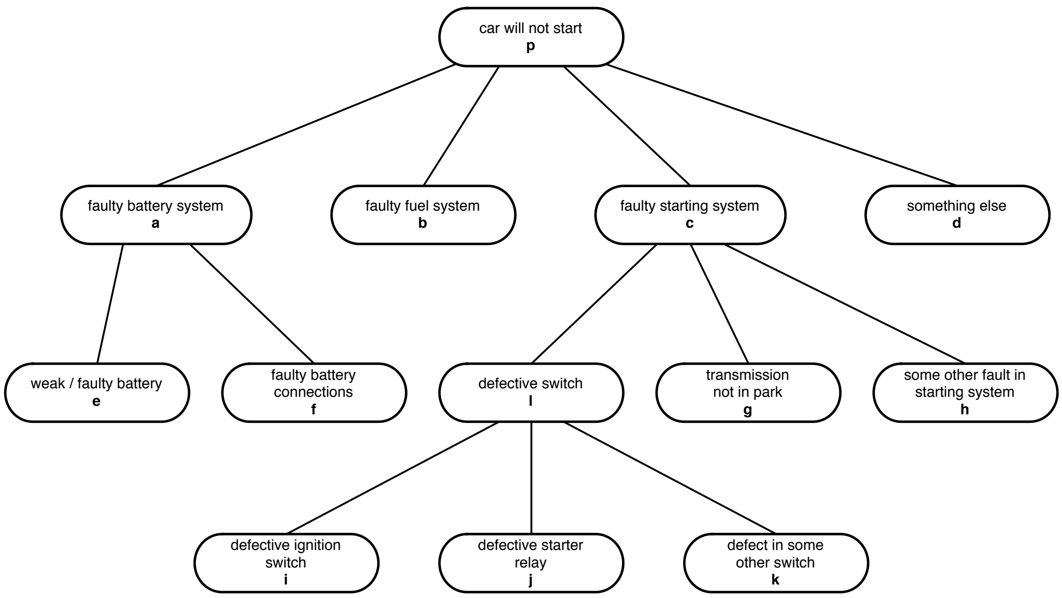


Figure 3.1: A tree of diagnoses for the "car will not start" problem

At the beginning, the list of diagnoses will by fairly crude. The different failure possibilities, i. e. the possible answers to the question, are given in Figure 3.1 by the nodes a, b, c, d. But it might be the case that this state of diagnoses is not precise enough and (some of) the diagnoses have to be decomposed. Such a decomposition is necessary in order to get finer answers regarding one diagnosis. So in the original list of diagnoses, the diagnosis a will be replaced by the list of diagnoses e, f and the diagnosis c will be replaced by the list of diagnoses l, g, h. This process can be repeated again and again, until the desired stage of precision is reached. In the above tree, the diagnosis l is decomposed into the list of diagnoses i, j, k, which gives rise to a third list of possible diagnoses. On the top of the tree, the degenerate diagnosis p, telling that the car will not start, is added.

There are different lists of diagnoses, they are all frames of discernment for the three increasingly finer defined questions: "Which element of the list is the cause of the problem?". Explicitly, the frames are

$$\begin{aligned}
\Omega_1 &= \{\emptyset\}, \\
\Omega_2 &= \{a, b, c, d\}, \\
\Omega_3 &= \{e, f, b, l, g, h, d\}, \\
\Omega_4 &= \{e, f, b, i, j, k, g, h, d\},
\end{aligned}$$

where Ω_1 is the artificial frame corresponding to the question "What is the problem with the car?". Passing from a frame Ω_i to a finer frame Ω_{i+1} is done in this example by substituting a diagnosis by a set of more precise diagnoses. This is described by a mapping

$$\tau_i : \Omega_i \to \mathfrak{P}(\Omega_{i+1}).$$

For example, Ω_3 is obtained from Ω_2 by application of $\tau_2 : \Omega_2 \to \mathfrak{P}(\Omega_3)$, given by

$$\tau_2(x) = \begin{cases} \{x\} & \text{if } x \in \{b, d\} \\ \{e, f\} & \text{if } x = a \\ \{l, g, h\} & \text{if } x = c. \end{cases}$$

This shows that the passage from one frame to a finer frame can be expressed by a mapping τ. Hereinafter, such a mapping will be called refining.

Obviously, the frames Ω_1 to Ω_4 given above are related. Instead of considering a mapping between them, we can see them as a collection of partitions[3] of the frame Ω_4. In this alternative but equivalent point of view the blocks of a partition are sets of elements of Ω_4. The blocks of the partition corresponding to $\Omega_2 = \{a, b, c, d\}$ are $\{e, f\}$, $\{b\}$, $\{i, j, k, g, h, \}$, $\{d\}$ and the blocks of the partition corresponding to $\Omega_3 = \{e, f, b, l, g, h, d\}$ are $\{e\}$, $\{f\}$, $\{b\}$, $\{i, j, k\}$, $\{g\}$, $\{h\}$, $\{d\}$. The partitions corresponding to the two remaining frames are the two extreme cases. There is only one block in the partition corresponding to Ω_1, namely $\{e, f, b, i, j, k, g, h, d\}$, i. e. Ω_4 itself. The partition corresponding to Ω_4, however, has as many blocks as there are elements in Ω_4, each being a singleton: $\{e\}$, $\{f\}$, $\{b\}$, $\{i\}$, $\{j\}$, $\{k\}$, $\{g\}$, $\{h\}$, $\{d\}$. We have seen that another approach for expressing the refinement from one frame to another is looking at the partitions which are related to them. To each frame Ω_i corresponds a partition $\mathcal{P}_i$ and the above mapping τ can now be seen as a mapping from a partition to the powerset of the partition of the next finer partition. Formally, $\tau_i : \mathcal{P}_i \to \mathfrak{P}(\mathcal{P}_{i+1})$. ⊙

Refinements and Coarsenings

The above example shows that there are different frames, providing more or less precise statements. Each frame is a set of possible answers. By decomposing the elements (answers) of a frame into more precise statements, a finer frame is obtained.

[3] See Definition 3.7 below for a formal introduction of partitions.

This must be done when the possible answers to the question considered so far prove to be too coarse to express the required level of detail of the desired information. In order to formalize this idea, let Ω, Λ be two frames, where Λ represents the new set of finer possible answers. Λ is constituted of the elements obtained by splitting up (some of) the elements of Ω. The mapping $\tau : \Omega \rightarrow \mathfrak{P}(\Lambda)$ represents this subdivision of the elements of Ω. It assigns a subset $\tau(\omega) \subseteq \Lambda$ to each element $\omega \in \Omega$. Some $\omega \in \Omega$ might not be decomposed. Such an ω is simply placed in Λ and the mapping is defined to be $\tau(\omega) = \{\omega\}$. Following (Shafer, 1976), the mapping τ is called a refining, if it satisfies the following requirements in order to represent indeed a decomposition of the elements of Ω:

Definition 3.5 (Refining)

Let Ω, Λ be two frames. The mapping

$$\tau : \Omega \rightarrow \mathfrak{P}(\Lambda)$$

is called a refining*, if*

1. $\tau(\omega) \neq \emptyset$ *for all* $\omega \in \Omega$.
2. *For any* $\omega, \omega' \in \Omega$, $\tau(\omega) \cap \tau(\omega') = \emptyset$ *whenever* $\omega \neq \omega'$.
3. $\bigcup_{\omega \in \Omega} \tau(\omega) = \Lambda$.

An element in Ω can never be decomposed to the empty set. The decomposition of two different elements of Ω will also result in two disjoint sets in Λ. Finally, the union of the decomposition of all elements of Ω equals Λ, the new set of possible answers. A refining τ can be extended from Ω to $\mathfrak{P}(\Omega)$ by defining

$$\tau(A) = \bigcup_{\omega \in A} \tau(\omega) \tag{3.1}$$

for all non-empty subsets of $A \subseteq \Omega$ and by letting $\tau(\emptyset) = \emptyset$ by convention. See (Shafer, 1976) for more properties of refinings.

We have seen that a refining creates a relation between two frames. In order to express how one frame is related to another, the following notions are introduced. They will make things easier when talking about more than one frame.

Definition 3.6 (Refinement, Coarsening)

Let the mapping $\tau : \Omega \rightarrow \mathfrak{P}(\Lambda)$ be a refining. Then, the frame Λ is called a refinement *of the frame Ω and Ω is called a* coarsening *of Λ. In other words, Λ is a* finer *frame than Ω and Ω is a* coarser *frame than Λ.*

Partition

The refining introduced in Definition 3.5 represents a decomposition of the elements of Ω. The family of sets $\tau(\omega)$, $\omega \in \Omega$ is a partition of the frame Λ. It is worth having a closer look at partitions, as this concept will be important below.

Consider a set U of elements. U is also called universe. A partition $\mathcal{P}$ divides U into different parts by regrouping its elements.

Definition 3.7 (Partition)

A partition of U is a set $\mathcal{P}$ of pairwise disjoint non-empty subsets of U whose union is U. The elements of a partition $\mathcal{P}$ are called the blocks *of $\mathcal{P}$.*

Clearly, the refining mapping of Definition 3.5 above induces a partition where the set U which is partitioned corresponds to the frame Λ and $\mathcal{P}$, the set of pairwise disjoint subsets of Λ, is $\{\tau(\omega) : \omega \in \Omega\}$. Some important characteristics of partitions are listed below. We refer to (Grätzer, 2003) for more details.

- A singleton as a block of a partition $\mathcal{P}$ is called trivial.
- If $a, b \in U$ belong to the same block of partition $\mathcal{P}$, we write $a \equiv b \pmod{\mathcal{P}}$, which is an equivalence relation on U. There is a one-to-one correspondence between the partitions of U and the equivalence relations on U.
- The set of all partitions of U is denoted by $Part(U)$.

This set $Part(U)$ of all partitions will now be the center of interest. U may be partitioned in many different ways. In particular, the partitions in $Part(U)$ are partially ordered, see (Grätzer, 2003). It is important to note that the partial order introduced in the following is the reverse of that which is usually considered with regard to partitions. However, the one given below is more convenient for our purposes, as we will see in a moment. The partial order between two partitions $\mathcal{P}_j, \mathcal{P}_k \in Part(U)$ is established by the relation $\leq$:

$$\begin{aligned}\mathcal{P}_j \leq \mathcal{P}_k \quad \text{iff} \quad & \forall a, b \in U, \\ & a \equiv b \pmod{\mathcal{P}_k} \text{ implies that } a \equiv b \pmod{\mathcal{P}_j}.\end{aligned} \tag{3.2}$$

$\mathcal{P}_k$ is then called a subpartition of $\mathcal{P}_j$. An alternative, but equivalent formulation is

$$\begin{aligned}\mathcal{P}_j \leq \mathcal{P}_k \quad \text{iff} \quad & \text{for every block } B \in \mathcal{P}_k \text{ there exists a block } B' \in \mathcal{P}_j \\ & \text{such that } B \subseteq B'.\end{aligned} \tag{3.3}$$

Figure 3.2 shows two partitions $\mathcal{P}_0$ and $\mathcal{P}_1$, where $\mathcal{P}_1$ is a subpartition of $\mathcal{P}_0$.

$Part(U)$, the set of all partitions, together with the partial ordering $\leq$ forms a complete lattice, called the partition lattice, see (Grätzer, 2003). The blocks of the

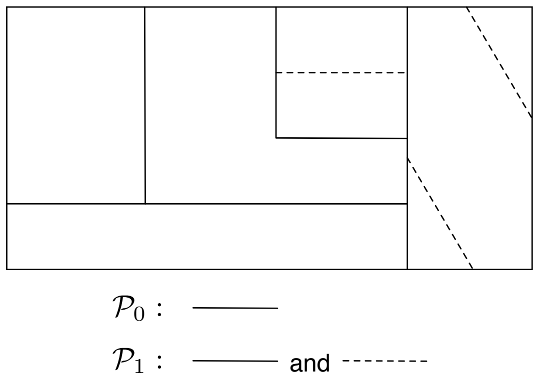


Figure 3.2: $\mathcal{P}_0 \leq \mathcal{P}_1$

join of an arbitrary collection of partitions $\{\mathcal{P}_j : \mathcal{P}_j \in Part(U), j \in J\}$ are precisely the non-empty intersections of the blocks in the partitions $\mathcal{P}_j$:

$$\bigvee_{j \in J} \mathcal{P}_j \quad = \quad \Big\{ \bigcap_{j \in J} B_j : B_j \in \mathcal{P}_j, \bigcap_{j \in J} B_j \neq \emptyset \Big\}. \tag{3.4}$$

The meet of an arbitrary collection of partitions $\{\mathcal{P}_j : \mathcal{P}_j \in Part(U), j \in J\}$ is defined by means of its join:

$$\bigwedge_{j \in J} \mathcal{P}_j \quad = \quad \bigvee \Big\{ \mathcal{P} : \mathcal{P} \in Part(U), \mathcal{P} \leq P_j \quad \forall j \in J \Big\}. \tag{3.5}$$

As $(Part(U), \leq)$ forms a complete lattice, the join and the meet of any arbitrary collection of partitions always exists. The bottom element $\perp$ of $Part(U)$ has only one block, namely U itself. The top element $\top$ has only trivial blocks. See (Grätzer, 2003) for the proofs.

Finally, if $\mathcal{P}_j \leq \mathcal{P}_k$, let

$$d : \mathcal{P}_j \to \mathfrak{P}(\mathcal{P}_k) \tag{3.6}$$

denote the mapping which assigns to each block $B \in \mathcal{P}_j$ the set of blocks in $\mathcal{P}_k$ whose union is B. It is called decomposition mapping. This definition of d can be extended to sets of blocks $X \subseteq \mathcal{P}_j$ if we define $d(X) = \bigcup \Big\{ d(B) : B \in X \Big\}$. This is a refining in the sense of Definition 3.5.

In order to relate partitions to questions and frames, let us sum up what we have seen until now: A question is given by the set of its possible answers. This set is called frame. Considering more than one frame gives rise to the notion of coarser and finer frames. A frame is coarser than another (which is then finer) if it provides less detailed answers. In particular, a coarser frame is a partition of a finer frame. Frames can thus be seen as partitions, which are known to form a lattice. We will now see under which conditions a set of frames is a lattice.

Family of Compatible Frames

The lattice structure introduced in Section 3.1 will now be related with the finer and coarser frames seen at the beginning of this section. This will be done by means of partitions, introduced just before. A frame represents a question and contains the possible answers to this question. The passage from one frame to a finer frame (providing more detailed answers) is done by a refining τ, see Definition 3.5. So the lattice of questions will be given by a family of frames and refinings. If the frames and the refinings linking the frames fit in the lattice structure presented above, they are called compatible:

Definition 3.8 (Family of Compatible Frames)

A collection of frames $\mathcal{F}$ toghether with a collection $\mathcal{R}$ of refinings between them forms a family of compatible frames*, abbreviated by f.c.f., denoted by $(\mathcal{F}, \mathcal{R})$, if the following four conditions hold:*

1. *For each pair Ω, Λ of frames in $\mathcal{F}$, there is at most one refining $\tau : \Omega \rightarrow \mathfrak{P}(\Lambda)$ in $\mathcal{R}$.*
2. *There exists a set U and a mapping $\sigma : \mathcal{F} \rightarrow Part(U)$ such that $\sigma(\Omega_1) \neq \sigma(\Omega_2)$ whenever $\Omega_1 \neq \Omega_2$ (i. e. σ is injective) and such that $\sigma(\mathcal{F}) = \{\sigma(\Omega) : \Omega \in \mathcal{F}\}$ is a sublattice of $Part(U)$ containing the least partition $\perp$ and the greatest partition $\top$.*
3. *For each frame $\Omega \in \mathcal{F}$, there is a bijective mapping $b : \Omega \rightarrow \sigma(\Omega)$. By this mapping b, exactly one block $b(\omega) \in \sigma(\Omega)$ corresponds to each possible answer $\omega \in \Omega$.*
4. *There is a refining $\tau : \Omega \rightarrow \mathfrak{P}(\Lambda)$ in $\mathcal{R}$ iff $\sigma(\Omega) \geq \sigma(\Lambda)$. In this case, $d(b(F)) = b(\tau(F))$ for all $F \subseteq \Omega$, i. e. the diagram depicted in Figure 3.3 is commutative.*

$$\begin{tikzcd}
\Omega \arrow[r, "b"'] \arrow[d, "\tau"] & \sigma(\Omega) \arrow[d, "d"'] \\
\Lambda \arrow[r, "b"] & \sigma(\Lambda)
\end{tikzcd}$$

Figure 3.3: Correspondence between refining and partitions

The first condition of the above definition is rather intuitive. Condition 2 tells that the frames in $\mathcal{F}$ correspond to partitions of a set U. The mapping σ is actually an embedding of $\mathcal{F}$ into $Part(U)$. It is σ which is used to make $\mathcal{F}$ a lattice. $\sigma(\mathcal{F})$ is a sublattice of $Part(U)$. Not only the least partition $\perp$ of $Part(U)$, having only one block which is U itself, is in $\sigma(\mathcal{F})$, but also the greatest partition $\top$, having only trivial blocks, the singletons of U. Therefore, there exists exactly one frame Ω such that $\sigma(\Omega)$ is this greatest partition. By the mapping b of condition 3 we know that every block ω in this very frame Ω is in bijective correspondence with the blocks of the greatest partition of U. Finally, according to condition 4, to each refining τ between the frames Ω and Λ, there is a corresponding decomposition mapping d (see Equation 3.6) between the partitions $\sigma(\Omega)$ and $\sigma(\Lambda)$.

There are two important theorems for a family of compatible frames, taken from (Kohlas & Monney, 1995, Chapter 7):

Theorem 3.9 *If $\tau : \Omega \to \Lambda$ and $\mu : \Lambda \to \Gamma$ are two refinings in a family of compatible frames $(\mathcal{F}, \mathcal{R})$, then, there is exactly one refining $\rho : \Omega \to \Gamma$ in $\mathcal{R}$ and it satisfies $\rho = \mu \circ \tau$.*

The above theorem states that the refinings relating the frames in a f.c.f. are indeed defined in a compatible way. Furthermore, by considering the corresponding partitions, the frames of a f.c.f. can be partially ordered by

$$\Omega \leq \Lambda \quad \text{iff} \quad \sigma(\Omega) \leq \sigma(\Lambda).$$

This means that $\Omega \leq \Lambda$ iff Λ is finer than Ω (see Definition 3.6). Then, the following theorem asserts that a set $\mathcal{F}$ of frames or questions, together with the ordering relation $\leq$, actually forms a lattice, as it is a sublattice of $Part(U)$:

Theorem 3.10 *In a family of compatible frames $(\mathcal{F}, \mathcal{R})$, the partially ordered set $(\mathcal{F}, \leq)$ is a lattice. If Ω, Λ are two frames in $\mathcal{F}$, then*

$$\begin{aligned} \Omega \vee \Lambda &= \sigma^{-1}(\sigma(\Omega) \vee \sigma(\Lambda)), \\ \Omega \wedge \Lambda &= \sigma^{-1}(\sigma(\Omega) \wedge \sigma(\Lambda)). \end{aligned}$$

From a semantic point of view, questions are represented by their possible answers. The set of possible answers is called frame. So a question is given by a frame. To each frame (and thus to each question), there is a corresponding partition. The blocks of the partition are the possible answers to the question. As questions may be finer or coarser, the answers to these questions have different levels of detail. The granularity of questions gives rise to different partitions which form a lattice. This lattice, which has already been introduced and motivated in Section 3.1, is now concretized. Thereby, meaning is given to the purely syntactic structure of the previous section.

3.2.2 Questions Induced by Sets of Variables

One basic component of our algebraic theory of information is the lattice of questions, given by a family of compatible frames. A f.c.f. is a very general concept capturing the nature of questions. In particular, it models questions which are given by sets of variables, as already seen in Section 2.2. Questions as sets of variables are obviously a special case of what has presented so far in this chapter. But as this special case takes place very often, we will have a closer look at it.

Lattice of Subsets

Let Vbl denote the countable set of all variables considered. In order to name finite subsets of Vbl, lower-case letters x, y, z are used. Variables will be designated by the lower-case letter v_j with indices in $j \in \mathbb{N}$. We are interested in sets of variables, as they represent questions. This is why we are considering the set of all finite subsets of Vbl and call it D. Since every $x \in D$ is a set of variables representing a question, D can be seen as the set of all possible questions. An order $\leq$ can be defined on D by the set inclusion relation $\subseteq$. $x \leq y$ means that x is a coarser question than y and that the latter is finer than the former. x being a coarser question than y may be paraphrased by x asking for the values of less variables than y. The order on D is

$$x \leq y \quad \text{iff} \quad x \subseteq y, \tag{3.7}$$

which is actually a partial order: The family of subsets of a set ordered by set inclusion is the most important example of a partial order, see (Davey & Priestley, 2002). The three conditions of Definition 3.1 are obviously satisfied by the properties of set inclusion:

1. *Reflexivity:* $x \subseteq x$ for all $x \subseteq Vbl$.
2. *Antisymmetry:* $x \subseteq y$ and $y \subseteq x$ imply $x = y$.
3. *Transitivity:* $x \subseteq y$ and $y \subseteq z$ imply $x \subseteq z$.

Every pair of elements $x, y \in D$ has a supremum in this partial order, which is the union of the two sets of variables and means that we are asking two questions simultaneously and therefore combine both questions. This supremum $x \cup y$ represents the coarsest question finer than both x and y. Furthermore, for every two questions $x, y \in D$, there exists an infimum in this partial order, given by the intersection of the two corresponding sets of variables. $x \cap y$ expresses the common part of both questions and is therefore the unique finest question, which is both coarser than x and y. Sets of variables form a lattice with $\cup$ as join and $\cap$ as meet. This lattice has as bottom element $\emptyset$ and, if Vbl is finite, Vbl is the top element. Otherwise, there is no top element, but it is not necessary, since we can consider the least upper bound of a set of finite subsets of Vbl. As mentioned before, this results again in a finite subset of Vbl. Then, D is of course a complete lattice. Furthermore, it is distributive, as the following laws are satisfied:

$$x \cup (y \cap z) = (x \cup y) \cap (x \cup z), \tag{3.8}$$
$$x \cap (y \cup z) = (x \cap y) \cup (x \cap z). \tag{3.9}$$

D is therefore also a modular lattice, as

$$\forall x, y, z \in D : y \supseteq z \text{ implies } x \cap (y \cup z) = (x \cap y) \cup z. \tag{3.10}$$

Example 3.2.2 (Lattice of Questions) Consider a situation described by three variables v_1, v_2 and v_3. The lattice of questions may be depicted as given in Figure 3.4. $\{v_1, v_2, v_3\}$, the set of all variables considered, is the finest possible question. The empty question $x = \emptyset$ is the coarsest one.

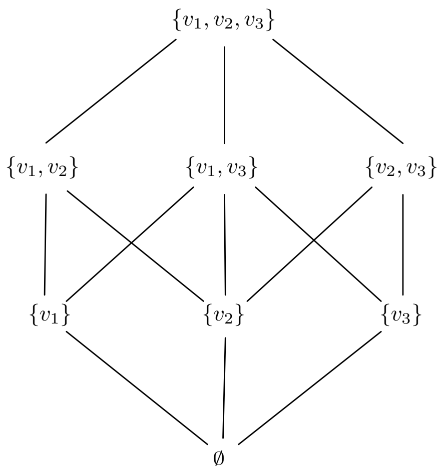


Figure 3.4: Lattice of questions for the variables v_1, v_2, v_3

Every set of questions has a supremum (join) in this partial order, which is the union of the sets of variables. The join $\{v_1\} \vee \{v_3\}$ is $\{v_1\} \cup \{v_3\} = \{v_1, v_3\}$. It represents the crudest question finer than both $\{v_1\}$ and $\{v_3\}$. Furthermore, for every set of questions, there exists an infimum (meet) in this partial order, given by the intersection of the corresponding sets of variables. The meet $\{v_1, v_3\} \wedge \{v_1, v_2\}$ is $\{v_1, v_3\} \cap \{v_1, v_2\} = \{v_1\}$ and expresses the common part of both questions $\{v_1, v_3\}$ and $\{v_1, v_2\}$. $\{v_1\}$ is therefore the unique finest question, which is coarser than both $\{v_1, v_3\}$ and $\{v_1, v_2\}$. $\odot$

Frames

A variable v_j can take different values, maybe other ones than another variable v_i, $i, j \in \mathbb{N}$. Therefore, $\mathfrak{D}_{v_j}$ designates the set of possible values of a variable v_j. If we are asking which value the variable v_j actually takes, the question is represented by the set $\{v_j\}$ and the set of possible answers to this question is $\mathfrak{D}_{v_j}$. This is why $\mathfrak{D}_{v_j}$ is called the frame of v_j. When asking for the value of more than one variable, the Cartesian product of their frames has to be considered. This finally leads to the Cartesian product of all possible frames $\mathfrak{D}_{v_j}$, $v_j \in Vbl$, which is the frame of the set Vbl and denoted by

$$\mathfrak{D}_{Vbl} = \bigtimes_{v_j \in Vbl} \mathfrak{D}_{v_j}. \tag{3.11}$$

The elements of $\mathfrak{D}_{Vbl}$ are called valuations. They are sequences $\langle \omega_1 \omega_2 \ldots \rangle$, where $\omega_1 \in \mathfrak{D}_{v_1}$, $\omega_2 \in \mathfrak{D}_{v_2}$, etc.

Definition 3.11 (Valuation)

A sequence $\omega \in \mathfrak{D}_{Vbl}$, providing for each $v_j \in Vbl$ one value, is called a valuation*.*

We will now look at partitions of $\mathfrak{D}_{Vbl}$, where the blocks are sets of valuations.

Partitions Induced by Sets of Variables

Following Definition 3.7, for every $x \in D$ we are interested in a set $\mathcal{P}_x$ of pairwise disjoint subsets of $\mathfrak{D}_{Vbl}$ whose union is $\mathfrak{D}_{Vbl}$. Such a set $\mathcal{P}_x$ is termed partition of $\mathfrak{D}_{Vbl}$, induced by the set of variables x. For any $x \in D$ we can now define the partition $\mathcal{P}_x$ by describing its blocks, the pairwise disjoint subsets of $\mathfrak{D}_{Vbl}$. Two valuations $\omega, \theta \in \mathfrak{D}_{Vbl}$ are in the same block of the partition $\mathcal{P}_x$ if their values of the variables in x are the same. Formally, an equivalence relation π_x is required. Any two valuations $\omega = \langle \omega_1\, \omega_2\, \ldots \rangle$ and $\theta = \langle \theta_1\, \theta_2\, \ldots \rangle$ are equivalent relative to π_x iff $\omega_j = \theta_j$ for all $v_j \in x$. All elements θ which are equivalent to ω make up the equivalence class $[\omega]_{\pi_x} = \{\theta \in \mathfrak{D}_{Vbl} : (\theta, \omega) \in \pi_x\}$. The partition $\mathcal{P}_x$ is therefore constituted of all the equivalence classes which are determined by π_x: $\mathcal{P}_x = \{[\omega]_{\pi_x} : \omega \in \mathfrak{D}_{Vbl}\}$.

3.3 Conclusion

Questions are represented by the possible answers they allow. The set of possible answers to a question is called frame. A refinement mapping between two frames expresses the granularity of their possible answers. The elements of a finer frame are more detailed answers than those of a coarser frame and are obtained by the decomposition of the answers of the coarser frame. In other words, the coarser frame corresponds to a partition of the finer frame, each answer (block) of the coarser frame has been decomposed into the blocks making up the finer frame. So a partition is related to every frame or question and vice versa. The set $Part(U)$ of all partitions forms a lattice. The frames we are considering, in particular the frames and partitions induced by finite subsets of some set of variables, are a sublattice of $Part(U)$.

4
Labeled Information Algebras

Information is a central concept of science, especially of computer science. [...] The algebra allows for a generic study of the structure of information.

Jürg Kohlas
Information Algebras: Generic Structures for Inference

We now come to the core of the algebraic theory of semantic information. A mathematical characterization of information is given in this chapter, applying the labeled point of view which has already been introduced in Section 2.1. (Chapter 5 will provide the domain-free point of view.) We are going to formalize the ideas which were intuitively introduced and exemplified in Chapter 2.

The two main features of information induce an algebraic structure behind it:

1. The fact that information is always considered relative to a specific question leads to the necessity to focus the available information on the field of interest and to extract the information relevant to the actual question.

2. The fact that information comes in pieces, maybe from different sources, shows that there is need for aggregation of information, in order to get an overall view of the combined information.

It is therefore natural to consider a two-sorted algebra of a set of pieces of information and a set of questions. The basic operations of this algebra are combination of information and focusing of information. The first two sections of this chapter are devoted to an introduction to this algebraic structure; Sections 4.3 and 4.4 show variants for information focusing. Section 4.5 is a digression from the main, rather theoretical, subject of this thesis, giving a very short insight into the inference mechanism which is associated with information algebras. Finally, information algebras with additional properties are presented in Sections 4.6 and 4.7.

This chapter is based on an abstract, axiomatic system, which was first introduced by Shenoy and Shafer (Shenoy, 1989; Shenoy & Shafer, 1990a). A changed version of this axiomatic formulation is used, namely the idempotency property (see Section 4.2) has been added; the mathematical structure defined by these axioms is called information algebra. Its concepts and ideas are taken from (Kohlas, 2003). All omitted proofs can be found there.

4.1 The Framework

Information algebras are made up of a lattice D of questions and a set Φ of pieces of information. The lattice of questions has already been discussed in detail in Chapter 3. As every piece of information $\phi \in \Phi$ relates to a question $x \in D$, this question x will be called the domain of ϕ. Given a domain $x \in D$, there is a set Φ_x regrouping all pieces of information with domain x:

$$\Phi_x \quad = \quad \{\phi \in \Phi : d(\phi) = x\}. \tag{4.1}$$

D is also referred to as the lattice of domains. Contrary to the intuitive introduction given in Section 2.3.1 and the examples given in Chapter 2, domains are not restricted to subsets of variables, but a useful generalization is assumed, as given in Chapter 3: Domains of pieces of information are elements of a lattice D provided with a partial order $\leq$ and the operations of join ($\vee$) and meet ($\wedge$). In summary, it can be stated that we dispose of

- a lattice D of domains and
- a set $\Phi = \bigcup_{x \in D} \Phi_x$ of pieces of information.

For Φ and D, three basic operations are defined. See Section 2.3.1 for a detailed explanation of their meaning.

Definition 4.1 (Labeling)

The labeling *operation*

$$d : \Phi \rightarrow D$$

tells to which domain or question $x \in D$ a piece of information $\phi \in \Phi$ refers to:

$$\phi \mapsto d(\phi).$$

Definition 4.2 (Combination)

The combination *operation*

$$\otimes : \Phi \times \Phi \rightarrow \Phi$$

is used to add two pieces of information $\phi, \psi \in \Phi$ together in order to get the aggregated piece of information:

$$(\phi, \psi) \mapsto \phi \otimes \psi.$$

Definition 4.3 (Marginalization)

The marginalization *operation*

$$\downarrow : \Phi \times D \to \Phi$$

can be understood as focusing a piece of information $\phi \in \Phi$, which pertains to the domain $d(\phi)$, on some less precise (coarser) domain of interest $x \in D$:

$$(\phi, x) \mapsto \phi^{\downarrow x}, \text{ for } x \leq d(\phi).$$

4.2 Axiomatics

The following set of axioms is now imposed on (Φ, D) with the three operations of Definitions 4.1, 4.2 and 4.3:

1. *Semigroup*:
   The operation of combination is commutative and associative, i. e. for any $\phi, \psi \in \Phi$
   $$\begin{aligned} \phi \otimes \psi &= \psi \otimes \phi \\ (\phi \otimes \psi) \otimes \zeta &= \phi \otimes (\psi \otimes \zeta) \end{aligned}$$
   For all $x \in D$, there is an element $e_x \in \Phi$ with $d(e_x) = x$ such that for all $\phi \in \Phi$ with $d(\phi) = x$,
   $$e_x \otimes \phi = \phi \otimes e_x = \phi. \tag{4.2}$$
   e_x is called the neutral element of domain x.
2. *Labeling*:
   For any $\phi, \psi \in \Phi$,
   $$d(\phi \otimes \psi) = d(\phi) \vee d(\psi).$$
3. *Marginalization*:
   For $\phi \in \Phi$, $x \in D$ such that $x \leq d(\phi)$,
   $$d(\phi^{\downarrow x}) = x.$$
4. *Transitivity*:
   For any $\phi \in \Phi$ and $x \leq y \leq d(\phi)$,
   $$(\phi^{\downarrow y})^{\downarrow x} = \phi^{\downarrow x}.$$
5. *Combination*:
   For $\phi, \psi \in \Phi$ with $d(\phi) = x$, $d(\psi) = y$,
   $$(\phi \otimes \psi)^{\downarrow x} = \phi \otimes \psi^{\downarrow x \wedge y}.$$

6. *Stability*:
   For $x, y \in D$, $x \leq y$,

$$e_y^{\downarrow x} \quad = \quad e_x.$$

7. *Idempotency*:
   For any $\phi \in \Phi$ and $x \in D$, $x \leq d(\phi)$,

$$\phi \otimes \phi^{\downarrow x} \quad = \quad \phi.$$

The semigroup axiom tells that pieces of information can be combined in any order without changing the result; the neutral element does not contain any information relative to domain x, it is the empty information on that domain. There is a unique neutral element e_x for each sub-semigroup Φ_x. The labeling axiom says that the combination of two pieces of information leads to a piece of information whose domain is the join of the two original domains. Marginalization of a piece of information to a domain x returns a piece of information regarding that very domain x, as stated by the third axiom. The fourth axiom is about the transitivity of marginalization; this operation can be done stepwise, by passing through intermediate domains. Furthermore, marginalization is distributive over combination. This is postulated by the fifth axiom. Instead of combining two pieces of information and marginalizing afterwards to one of their domains, one can marginalize in a first step the second piece of information to the meet of both domains and carry out the combination in a second step. The sixth axiom, the stability axiom, expresses the fact that the marginalization of an empty piece of information remains empty. Finally, the idempotency axiom ensures that the combination of a piece of information with itself (or a marginalized version of itself) gives nothing new. Idempotency is one of the most basic and intuitive properties of information, since by repeating a piece of information over and over again, nothing new will appear. The consequences of this idempotency property will be discussed in detail in Section 6.2. For the time being, we just point out that idempotency introduces a partial order of being less informative in Φ. For any two pieces of information $\phi, \psi \in \Phi$ it is expressed in the following way:

$$\phi \leq \psi \quad \text{iff} \quad \phi \otimes \psi = \psi. \tag{4.3}$$

Definition 4.4 (Labeled Information Algebra)

A system (Φ, D) where Φ is a set of pieces of information and D a lattice of questions or domains, together with the three operations of labeling, combination and marginalitziation (Definitions 4.1 to 4.3), satisfying the preceding seven axioms, is called a (labeled) information algebra.

Some elementary consequences of the above set of axioms are listed in the following lemma, taken from (Kohlas, 2003, Chapter 2.2).

Lemma 4.5 *Let (Φ, D) be a labeled information algebra. Then, combination and marginalization behave in the following way:*

1. *If $\phi, \psi \in \Phi$ such that $d(\phi) = x \in D$, $d(\psi) = y \in D$, then*
$$(\phi \otimes \psi)^{\downarrow x \wedge y} = \phi^{\downarrow x \wedge y} \otimes \psi^{\downarrow x \wedge y}.$$

2. *If $\phi, \psi \in \Phi$ such that $d(\phi) = x \in D$, $d(\psi) = y \in D$ and $z \in D$ with $z \leq y$, then*
$$(\phi \otimes \psi)^{\downarrow z} = (\phi \otimes \psi^{\downarrow x \wedge y})^{\downarrow z}.$$

3. *If $\phi, \psi \in \Phi$ such that $d(\phi) = x \in D$, $d(\psi) = y \in D$ and $z \in D$ with $x \leq z \leq x \vee y$, where D is a modular lattice (see Equation 3.10), then*
$$(\phi \otimes \psi)^{\downarrow z} = \phi \otimes \psi^{\downarrow y \wedge z}.$$

The importance of the above statements lies in their computational implications, which are essential for the inference mechanism of Section 4.5, where also more details about the required modularity can be found. The lemma, together with the combination axiom, allow to restrict computations to smaller domains by using the right hand side of the identities.

4.2.1 Labeled Information Algebras with Null Elements

A labeled information algebra (Φ, D) may have pieces of information that are incompatible with others. When combining them, they create something contradictory. Such a contradiction is expressed by the null element of combination. The null element is "absorbing", as every further combination with it remains contradictory. A piece of information $z_x \in \Phi$ is called the null element for x if for all $\phi \in \Phi$ with $d(\phi) = x$ it holds that

$$\phi \otimes z_x = z_x. \tag{4.4}$$

Note that such a null element is always idempotent. If a labeled information algebra has null elements, it has a unique null element z_x for each domain $x \in D$. Then, we impose an eighth axiom, namely the nullity axiom:

8. *Nullity*:
For all $x \in D$ there is an element $z_x \in \Phi$ such that $z_x \otimes \phi = z_x$ for all $\phi \in \Phi$ with $d(\phi) = x$. If $x, y \in D$ and $x \leq y$, then
$$z_x \otimes e_y = z_y.$$

The following Lemma on null elements is (Kohlas, 2003, Lemma 6.1) and can be proven by means of the stability and the idempotency axiom.

Lemma 4.6 *Let (Φ, D) be an information algebra with null elements.*

1. *For $x, y \in D$ and $x \leq y$ it holds that $z_y^{\downarrow x} = z_x$.*
2. *If $\phi \in \Phi$ such that $d(\phi) = y \geq x \in D$ and $\phi^{\downarrow x} = z_x$, then $\phi = z_y$.*
3. *For $x, y \in D$ it holds that $z_x \otimes z_y = z_{x \vee y}$.*

Two pieces of information $\phi, \psi \in \Phi$ with $d(\phi) = x \in D$, $d(\psi) = y \in D$ are called contradictory if

$$\phi \otimes \psi \quad = \quad z_{x \vee y}. \tag{4.5}$$

4.3 Variable Elimination

Consider a labeled information algebra (Φ, D). When D is a lattice of finite subsets of a set of variables Vbl (and it is important to emphasize this!), as in Section 3.2.2 and in the examples shown in Chapter 2, then the operation of marginalization (Definition 4.3) can be substituted by another primitive operation, namely variable elimination, which also servers for information extraction.[1]

Definition 4.7 (Variable Elimination)

The variable elimination *operation*

$$- : \Phi \times Vbl \rightarrow \Phi$$

is also used for focusing a piece of information $\phi \in \Phi$, which pertains to the domain $d(\phi)$, on some less precise (coarser) domain of interest $d(\phi) \setminus \{X\}$, where $X \in d(\phi)$:

$$(\phi, X) \mapsto \phi^{-X}, \text{ for } X \in d(\phi).$$

It is defined by means of marginalization:

$$\phi^{-X} \quad = \quad \phi^{\downarrow d(\phi) \setminus \{X\}}.$$

Replacing marginalization by variable elimination results in a system (Φ, D) with the following operations:

[1] As already indicated by its name, this operation eliminates variables from the domain of the piece of information it is applied to. This results in a piece of information bearing on less variables. An analogy to logics can be drawn, since the existential quantifier $\exists$ has the same function as the operation of variable elimination. When the quantifier $\exists$ precedes a formula, the quantified variables are already eliminated, they do not influence any more the meaning of the formula. See also Chapter 9 for more details on the relation between variable elimination and existential quantification.

1. *Labeling:* $\Phi \to D$; $\phi \mapsto d(\phi)$,
2. *Combination:* $\Phi \times \Phi \to \Phi$; $(\phi, \psi) \mapsto \phi \otimes \psi$,
3. *Variable Elimination:* $\Phi \times Vbl \to \Phi$; $(\phi, X) \mapsto \phi^{-X}$, for $X \in d(\phi)$.

When considering the operation of variable elimination instead of marginalization, the set of axioms for an information algebra has to be rewritten:

1. *Semigroup*:
   The operation of combination is commutative and associative, i. e. for any $\phi, \psi \in \Phi$

   $$\begin{aligned} \phi \otimes \psi &= \psi \otimes \phi \\ (\phi \otimes \psi) \otimes \zeta &= \phi \otimes (\psi \otimes \zeta) \end{aligned}$$

   For all $x \in D$, there is an element $e_x \in \Phi$ with $d(e_x) = x$ such that for all $\phi \in \Phi$ with $d(\phi) = x$,

   $$e_x \otimes \phi = \phi \otimes e_x = \phi.$$

   e_x is called the neutral element of domain x.

2. *Labeling*:
   For any $\phi, \psi \in \Phi$,

   $$d(\phi \otimes \psi) = d(\phi) \cup d(\psi).$$

3. *Variable Elimination*:
   For $\phi \in \Phi$ and $X \in d(\phi)$,

   $$d(\phi^{-X}) = d(\phi) \setminus \{X\}.$$

4. *Commutativity*:
   For $\phi \in \Phi$ and $X, Y \in d(\phi)$,

   $$(\phi^{-X})^{-Y} = (\phi^{-Y})^{-X}. \tag{4.6}$$

5. *Combination*:
   For $\phi, \psi \in \Phi$ with $d(\phi) = x$, $d(\psi) = y$ and $Y \notin x, Y \in y$,

   $$(\phi \otimes \psi)^{-Y} = \phi \otimes \psi^{-Y}.$$

6. *Stability*:
   For $X \in x \in D$,

   $$e_x^{-X} = e_{x \setminus \{X\}}. \tag{4.7}$$

7. *Idempotency*:
For any $\phi \in \Phi$ and $X \in d(\phi)$,

$$\phi \otimes \phi^{-X} = \phi.$$

Definition 4.8 (Φ, Vbl)

A system of pieces of information Φ over finite subsets of variables Vbl, together with the operations of labeling (Definition 4.1), combination (Definition 4.2) and variable elimination (Definition 4.7), which satisfies the above seven axioms is denoted by (Φ, Vbl).

Until now, it was only shown how to eliminate a single variable $X \in Vbl$. One could be interested in eliminating a set of variables $x \subseteq Vbl$. However, we do not want to be concerned with the elimination order, but the result should always be the same, independent of the the actual elimination sequence. This is made possible by the commutativity of elimination. So we do not have to worry about the actual elimination sequence and can define the elimination of several variables $X_1, X_2, \ldots, X_n \in d(\phi)$ as

$$\phi^{-\{X_1,X_2,\ldots,X_n\}} = (\cdots((\phi^{-X_1})^{-X_2})\cdots)^{-X_n} \tag{4.8}$$

The following lemma sums up a few elementary results on variable elimination. It is (Kohlas, 2003, Lemma 2.2), where the proofs can be found:

Lemma 4.9 *Consider a system (Φ, Vbl) of Definition 4.8 and variable set elimination according to Equation 4.8.*

1. *If $x \subseteq d(\phi)$ for some $\phi \in \Phi$, then*

$$d(\phi^{-x}) = d(\phi) \setminus x.$$

2. *If x and y are two disjoint subsets of $d(\phi)$ for some $\phi \in \Phi$, then*

$$(\phi^{-x})^{-y} = \phi^{-(x \cup y)}.$$

3. *If $\phi, \psi \in \Phi$ and $y \subseteq d(\psi)$, disjoint to $d(\phi)$, then*

$$(\phi \otimes \psi)^{-y} = \phi \otimes \psi^{-y}.$$

In Definition 4.7, the operation of variable elimination was expressed in terms of marginalization. Conversely, marginalization can also be expressed in terms of variable elimination: For $x \subseteq d(\phi)$, we have

$$\phi^{\downarrow x} = \phi^{-(d(\phi) \setminus x)}. \tag{4.9}$$

Until now, we have seen two ways of defining a labeled information algebra. The first approach is involving marginalization, the second one is considering variable elimination instead. This is however only possible if the lattice D is constituted of finite subsets of the set of variables Vbl. If this is the case, the two ways of defining algebras (Φ, D) and (Φ, Vbl) are in fact equivalent, as stated by the following theorem. The associated proof is that of (Kohlas, 2003, Theorem 2.3).

Theorem 4.10 *If (Φ, D) is an information algebra and variable elimination is defined as in Definition 4.7, then the system of axioms, given at the beginning of this section, is satisfied.*
If (Φ, Vbl) satisfies these axioms and marginalization is defined by Equation 4.9, then (Φ, D) is an information algebra.

4.4 Transport of Information

The operations of marginalization and variable elimination are used for projecting information to a subdomain of its domain. We now introduce a further operation which serves to transport information from its domain to an arbitrary other domain. A piece of information ϕ can be brought to a domain $x \in D$ using marginalization or variable elimination, but it is required that $x \leq d(\phi)$. The transport operation, however, will allow to bring ϕ to every $x \in D$. In order to describe the transport of information, we define in a first step a further operation, called vacuous extension.

Definition 4.11 (Vacuous Extension)

In a labeled information algebra (Φ, D), *the* vacuous extension *operation*

$$\uparrow : \Phi \times D \rightarrow \Phi$$

is used for expressing a piece of information $\phi \in \Phi$, which pertains to the domain $d(\phi)$, by means of some more precise (finer) domain of interest $d(\phi) \leq x \in D$:

$$(\phi, x) \mapsto \phi^{\uparrow x}, \text{ for } x \geq d(\phi).$$

The vacuous extension of ϕ is performed without adding further information:

$$\phi^{\uparrow x} \quad = \quad \phi \otimes e_x.$$

The name vacuous extension is justified, since for $d(\phi) = x \leq y \in D$ it holds by the combination and stability axioms that

$$(\phi^{\uparrow y})^{\downarrow x} = (\phi \otimes e_y)^{\downarrow x} = \phi \otimes e_y^{\downarrow x} = \phi \otimes e_x = \phi.$$

This now allows to define the transport operation.

Definition 4.12 (Transport)

In a labeled information algebra (Φ, D)*, the* transport *operation*

$$\rightarrow : \Phi \times D \rightarrow \Phi$$

permits to transport a piece of information $\phi \in \Phi$ *from its original domain* $d(\phi)$ *to any other domain* $x \in D$*:*

$$(\phi, x) \mapsto \phi^{\rightarrow x}.$$

It is defined by means of vacuous extension and marginalization. ϕ *is not extended to unnecessarily fine domains, the coarsest domain which is finer than both the original and the destination domain is taken:*

$$\phi^{\rightarrow x} \quad = \quad (\phi^{\uparrow x \vee d(\phi)})^{\downarrow x}. \tag{4.10}$$

Obviously, marginalization and vacuous extension can be seen as special cases of this transport operation, namely for $x \leq d(\phi)$ or $x \geq d(\phi)$, respectively.

Note further that for $d(\phi) = x$ and $y \in D$, we have $\phi^{\uparrow x \vee y} = \phi \otimes e_{x \vee y} = \phi \otimes e_y \otimes e_{x \vee y} = \phi \otimes e_y$, hence

$$\phi^{\rightarrow y} = (\phi \otimes e_y)^{\downarrow y} = \phi^{\downarrow x \wedge y} \otimes e_y = (\phi^{\downarrow x \wedge y})^{\uparrow y}. \tag{4.11}$$

Equations 4.10 and 4.11 give rise to an interesting situation in the lattice of domains, depicted in Figure 4.1.

$$\begin{tikzcd}
& x \vee y \arrow[dr] & \\
x \arrow[ur] \arrow[dr] & & y \\
& x \wedge y \arrow[ur] &
\end{tikzcd}$$

Figure 4.1: Two ways of transporting a piece of information

When a piece of information ϕ with domain $d(\phi) = x \in D$ is transported to another domain $y \in D$, there are two possible ways: One is to "go up" in the lattice D and to vacuously extend ϕ to the join of the original domain x and the destination domain y, before it is marginalized to y. This first approach is provided by Equation 4.10 in Definition 4.12 and corresponds to the upper part of Figure 4.1. The other possibility is to "go down" in the lattice D of domains. Then, ϕ is first projected to the meet of the original domain x and the destination domain y and it is vacuously extended to y afterwards. This second approach is described by Equation 4.11 and depicted in the lower part of Figure 4.1. Both approaches lead to the same result, which is especially important for computational purposes (see Section 4.5), where it is always advantageous to pass through smaller domains, using the second approach.

Some of the elementary results about the transport operation collected in (Kohlas, 2003, Lemma 3.2) are listed in the following lemma.

Lemma 4.13 *Consider a labeled information algebra* (Φ, D)*.*

1. *For each* $\phi \in \Phi$ *and all* $x, y \in D$*, it holds that* $(\phi^{\to x})^{\to y} = (\phi^{\to x \wedge y})^{\to y}$*.*
2. *If* $d(\phi) = x$*, then* $(\phi \otimes \psi)^{\to x} = \phi \otimes \psi^{\to x}$*.*
3. *If* $d(\phi) = x$*, then* $\phi^{\to x} = \phi$*.*
4. *If* $d(\phi) = x$ *and* $y \in D$*, then* $\phi \otimes \phi^{\to y} = \phi^{\to x \vee y}$*.*

These properties of transport are similar to the transitivity, combination, marginalization and idempotency axioms of the information algebra. In fact, they could replace them (Kohlas, 2003).

4.5 Local Computation

A knowledge base is a set of pieces of information $\{\phi_1, \ldots, \phi_n\}$. Each ϕ_i is an element of the set of all pieces of information Φ of an information algebra (Φ, D) and its domain will be denoted by $x_i = d(\phi_i)$.

Query processing is an important task on knowledge bases, done by aggregation of information and focusing on a given field of interest $y \in D$. Answering a query means telling what is known in the knowledge base on a domain y. It is quite straightforward to compute a global aggregation of the knowledge for that purpose, i. e.

$$\phi \quad = \quad \phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_n,$$

with $d(\phi) = x_1 \vee x_2 \vee \ldots \vee x_n$. Thereafter, the result ϕ is focused on the domain $y \leq d(\phi)$ of the query, i. e. the field of interest:

$$\phi^{\downarrow y} \quad = \quad (\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_n)^{\downarrow y}. \tag{4.12}$$

So we are dealing with a projection problem, sometimes also referred to as inference problem, see (Haenni, 2004). The projection problem with respect to the empty set is often of particular interest. It has even a proper name.

Example 4.5.1 (Satisfiability Problem)
The knowledge base ϕ is given by a set of propositional formulae, where every formula describes a piece of information ϕ_i. In order to call the knowledge base satisfiable, every element ϕ_i has to be satisfiable in the sense of Section 8.2. By marginalizing the knowledge base ϕ to the empty domain,

$$\phi^{\downarrow\emptyset} = (\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_n)^{\downarrow\emptyset},$$

it can be decided whether ϕ is satisfiable or not. $\phi^{\downarrow\emptyset} = \emptyset$ stands for a satisfiable knowledge base. If the marginalization results in $\{\perp\}$, the knowledge expressed by ϕ is contradictory (and thus not satisfiable). Many important problems in propositional logic are reduced to such a satisfiability problem. See (Langel, 2004) for further details. $\odot$

Theoretically, the very simple approach outlined in Equation 4.12 above leads to the answer of the query, but from a computational point of view, it is not very advisable. The complexity of combination and projection can grow exponentially with the size of the domains. Therefore, a projection performed on the maximum domain $d(\phi) = x_1 \vee x_2 \vee \ldots \vee x_n$ is not feasible for most applications.

As an alternative, a so-called local computation algorithm has originally been proposed for the case of Bayesian networks by (Lauritzen & Spiegelhalter, 1988). The algorithm is called local since it allows to solve the projection problem without considering domains any larger than the domains appearing in the ϕ_i. (Shenoy & Shafer, 1990b) showed how to apply this algorithmic method to more general cases, namely for underlying structures such as information algebras. That is where the computational importance of information algebras lies.

Local computation thus limits the operations for solving a projection problem to smaller domains. The local computation algorithm exploits the elementary consequences of the semigroup, the transitivity and the combination axiom of a labeled information algebra listed in Lemma 4.5.[2]

Let (Φ, D) be an information algebra. As already stated, the knowledge $\phi \in \Phi$ consists of the aggregation of certain pieces of information

$$\phi = \phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_n,$$

where every ϕ_i belongs to Φ. We say that ϕ factorizes into the ϕ_i or, in other words, that ϕ is a factorization or combination of pieces of information.

Lemma 4.5 is the crucial point which renders local computation possible since a factorization of the information ϕ is considered. The factorization of ϕ is related to a graph structure, called a join tree. Each node of the join tree carries one or more pieces of information (factors ϕ_i) and a nesting of combination and projection is performed on the nodes, on domains smaller than or equal to the nodes' domain. Every node communicates its result to its neighbors by a message passing algorithm. At the end of the algorithm, the node whose domain is the domain of the query contains the answer of the query. See (Shenoy & Shafer, 1990b), (Lauritzen & Spiegelhalter, 1988) and (Jensen *et al.*, 1990) for details on the principal local compuation mechanisms based on join trees. Due to their idempotency property, information algebras allow, from a computational point of view, a clear simplification of these architectures for the solution of the projection problem (Kohlas & Wilson, 2008; Kohlas, 2003)[3]. See also (Pouly, 2008) for a generic architecture for local computation.

[2]Information has often a modular structure, which allows a decomposition of the knowledge base. The third point of Lemma 4.5 requires a modular lattice of domains, but it is also possible to perform local computation with a general partition lattice, see (Mellouli, 1988).

[3]In (Kohlas & Wilson, 2008), information algebras are called idempotent valuation algebras.

4.6 Boolean Information Algebras

The information algebra framework presented in the foregoing sections provides a generic toolbox for modeling information. However, there are many formalisms which have more characteristics than the ones that are formalized by an information algebra. These characteristics may be expressed by further operations or properties. In this section we consider the case where a set of pieces of information bearing on a domain x forms a Boolean lattice. This will finally lead to a Boolean information algebra. The information algebras associated with propositional logic (treated in Chapter 8) and predicate logic (presented in Chapter 9) are Boolean. Based on (Davey & Priestley, 2002, Chapter 4), we will first specify what a Boolean lattice is (Section 4.6.1). Then, in Section 4.6.2, we will look at labeled Boolean information algebras in the way proposed in (Schneuwly, 2007, Chapter 7.1).[4]

4.6.1 Boolean Lattice

A Boolean lattice has, compared to a general lattice, additional structure. Its elements each have a unique complement, which is also an element of the Boolean lattice.

Definition 4.14 (Complement)

Let L be a lattice with bottom element $\perp$ and top element $\top$. The complement operator

$$^c : L \to L$$

attributes to each element $a \in L$ a unique element $a^c \in L$:

$$a \mapsto a^c,$$

such that

$$\begin{aligned} a \wedge a^c &= \perp \text{ and} \\ a \vee a^c &= \top. \end{aligned}$$

Such an element a^c is called the complement *of a.*

It is known that in a distributive lattice (satisfying Equations 3.8 and 3.9) an element can have at most one complement. This property is very desirable for information modeling, and leads us straight to the definition of Boolean lattices.

[4] In Chapter 5 information algebras will be looked at from a domain-free point of view. A Boolean version also exists. However, we restrict ourselves to the presentation of labeled Boolean information algebras as they are important for the measure of information in Chapter 7. For a description of domain-free Boolean information algebras, refer to (Kohlas & Schneuwly, 2009) or to (Kohlas, 2003, Chapter 6.5.2).

Definition 4.15 (Boolean Lattice)

A lattice L is called a Boolean lattice *if*

1. *L is distributive,*
2. *L has $\perp$ and $\top$,*
3. *each $a \in L$ has a (necessarily unique) complement $a^c \in L$.*

4.6.2 Labeled Boolean Information Algebra

As mentioned above, a Boolean information algebra (Φ, D) involves Boolean lattices. Actually, we consider a family of Boolean lattices $\langle \Phi_x, \leq_x \rangle$ with indices x out of the lattice of domains D. In this family there is for every $x \in D$ the set Φ_x whose elements bear on the domain x: $\Phi_x = \{\phi \in \Phi : d(\phi) = x\}$. The partial order $\leq_x$ is defined by Equation 4.3. This is looked at in more detail in Section 6.2. $\langle \Phi_x, \leq_x \rangle$ is a Boolean lattice having the following properties:

- The meet $\wedge_x$ and the join $\vee_x$ operators are mutually distributive.
- Φ_x has a least informative piece of information $\perp_x$ and a most informative piece of information $\top_x$, i. e. for all $\phi \in \Phi_x$ it holds that $\perp_x \leq_x \phi \leq_x \top_x$.
- Each piece of information $\phi \in \Phi_x$ has its complement $\phi^{c_x} \in \Phi_x$:

$$\begin{aligned} \phi \wedge \phi^c &= \perp, \\ \phi \vee \phi^c &= \top. \end{aligned}$$

  Since the complement operator is unary, we write without loss of generality ^{c} instead of ^{c_x}.

The set of all pieces of information Φ of a labeled Boolean information algebra is then the union of this family of Boolean lattices:

$$\Phi = \bigcup_{x \in D} \Phi_x. \tag{4.13}$$

The operations of labeling, combination and marginalization are defined as before in Section 4.1.

Further Operations and Further Axioms

A labeled Boolean information algebra (Φ, D) is a labeled algebra with additional properties, operations and axioms. We consider a lattice of domains D and a family of Boolean lattices Φ_x with indices x out of the lattice D. Φ is then given by Equation 4.13. Not only the usual operations of labeling ($d : \Phi \to D$, see Definition 4.1), combination ($\otimes : \Phi \times \Phi \to \Phi$, see Definition 4.2) and marginalization ($\downarrow$: $\Phi_x \times D \to \Phi_y$, for $y \leq x$, following Definition 4.3) are given, but further operations exist in labeled Boolean information algebras:

- for each $x \in D$ a complement operation ^{c} inside Φ_x (see Definition 4.14),
- for each $x \in D$ a meet operation $\wedge_x : \Phi_x \times \Phi_x \to \Phi_x$.

In addition to the eight axioms listed in Section 4.2, labeled Boolean information algebras are required to fulfill three further axioms. Note that the nullity axiom of Section 4.2 is always satisfied, as each Boolean lattice Φ_x has a top element which is the null element, so $z_x = \top_x$. Moreover, we identify the neutral element e_x with the bottom element $\perp_x$.

9. *Implied Distributive Lattices*:
   For all $x \in D$, the set
   $$\Phi_x \quad = \quad \{\phi \in \Phi : d(\phi) = x\}$$
   is a distributive lattice with meet $\wedge_x$ and join $\otimes$.
10. *Boolean*:
    For all $\phi \in \Phi_x$, the complement $\phi^c \in \Phi_x$ has the properties
    $$\begin{aligned} \phi \wedge_x \phi^c &= e_x, \\ \phi \otimes \phi^c &= z_x. \end{aligned}$$
11. *Weak Extended Distributivity*:
    For $x, y \in D$, where $x \leq y$ and $\phi, \psi \in \Phi_x$
    $$(\phi \wedge_x \psi) \otimes e_y \quad = \quad (\phi \otimes e_y) \wedge_y (\psi \otimes e_y).$$

So labeled Boolean information algebras satisfy additional axioms, expressing further properties about the two extra operations on each domain $x \in D$, complement and meet. It is important to stress that their use is restricted on the respective domain.[5]

4.7 Atomic Information Algebras

In this section, we look at a further special case of a labeled information algebras (Φ, D) which is essentially met by information algebra instances which can be expressed as tuple systems, see (Kohlas, 2003, Chapter 6.3) for a detailed discussion. Again, the two examples of propositional and predicate logic in Chapters 8 and 9 fulfill the additional properties listed below. An information algebra which is atomic allows interesting measures of information, as explained in Chapter 7. When modeling questions (Chapter 3), it turned out that a domain $x \in D$ can be seen as a

[5]Being interested in the meet of two pieces of information ϕ, ψ, bearing on two different domains, $d(\phi) = x, d(\psi) = y$, leads to the application of the meet-operation of the domain $x \vee y$, as both pieces of information have to be vacuously extended: $\phi \wedge' \psi := \phi^{\uparrow x \vee y} \wedge_{x \vee y} \psi^{\uparrow x \vee y}$. Here, $\wedge'$ is a kind of pseudo-meet, since it is applied to two pieces of information which are not elements of the same Boolean lattice. This is why ϕ and ψ have to be vacuously extended at first.

question. It may be the case that there exist some finest, non-contradictory possible answers to this question. A piece of information which is a most informative answer, relative to the domain or question $x \in D$, is called an atomic piece of information or simply an atom. An atomic information algebra is an information algebra with atomic pieces of information.

4.7.1 Atoms

An atom tells the most that can be said, besides contradiction. This means that an atom in a domain x is a most informative piece of information which is different from the null element z_x, among the pieces of information bearing on domain x. In the order expressed by Equation 4.3, it is a maximal element of $\Phi_x = \{\phi \in \Phi : d(\phi) = x \in D\}$:

Definition 4.16 (Atom)

Consider a labeled information algebra (Φ, D) with null elements. An element $\alpha \in \Phi_x$ is called an atom on *x, $x \in D$, if*

1. *$\alpha \neq z_x$ and*
2. *for all $\phi \in \Phi_x$, $\alpha \leq \phi$ implies either $\alpha = \phi$ or $\alpha = z_x$.*

Here are a few elementary properties of atoms. The corresponding proof is that of (Kohlas, 2003, Lemma 6.14):

Lemma 4.17 *Consider a labeled information algebra (Φ, D) with null elements.*

1. *If α is an atom on $x \in D$, then for every $y \in D$ with $y \leq x$, $a^{\downarrow y}$ is an atom on y.*
2. *If α is an atom on $x \in D$ and $\phi \in \Phi_x$, then either $\phi \leq \alpha$ or $\alpha \otimes \phi = z_x$.*
3. *If α, β are atoms on $x \in D$, then either $\alpha = \beta$ or $\alpha \otimes \beta = z_x$.*

So there might be more than one atom on a domain x. The following notation will hereinafter be used. $At_x(\Phi)$ denotes the set of all atoms on the domain x in the information algebra (Φ, D):

$$At_x(\Phi) \quad = \quad \{\alpha \in \Phi_x : \alpha \text{ atom on } x\}. \tag{4.14}$$

The set of all atoms of the information algebra (Φ, D) is given by

$$At(\Phi) \quad = \quad \bigcup_{x \in D} At_x(\Phi). \tag{4.15}$$

Furthermore, one may define the set

$$At(\phi) \quad = \quad \{\alpha \in At(\Phi) : d(\alpha) = d(\phi), \phi \leq \alpha\} \tag{4.16}$$

which contains all the atoms relative to a piece of information ϕ. If $\alpha \in At(\phi)$ we often say that α is an atom in ϕ or contained in ϕ. Now we have all the necessary tools for defining labeled atomic information algebras and its variants.

4.7.2 Information Algebras with Atoms

There are three different types of labeled information algebras with atoms:

Definition 4.18 (Atomic Information Algebra)

A labeled information algebra (Φ, D) with null elements is called

1. atomic *if for all $\phi \in \Phi$, if ϕ is not a null element, the set of atoms $At(\phi)$ is not empty.*
2. atomic composed *if it is atomic and if for all $\phi \in \Phi$*

$$\phi \quad = \quad \bigwedge At(\phi).$$

3. atomic closed *if it is atomic composed and if for every subset $A \subseteq At_x(\Phi)$ the meet (infimum) exists and belongs to Φ.*

This concludes our short excursion to atomic information algebras. An information-theoretical interpretation of what we have seen so far can be found in Section 6.4.

4.8 Conclusion

Information usually comes piecewise. There are two basic operations which are performed on pieces of information: They are aggregated in order to get an overall picture, and they may be focused on a given field of interest. In particular, each piece of information has a kind of label, called the domain of that piece of information or the question it refers to. Questions or domains form a lattice D. Together with a set of pieces of information Φ, to which the operations of combination, marginalization and labeling are applied, this leads to a two-sorted algebraic structure, called labeled information algebra, denoted by (Φ, D). The information algebra framework is given by a set of axioms and presents a generic way of modeling information. If the domains are sets of variables, an alternative to the marginalization operation is variable elimination. It is heavily used in the inference mechanism customized for information algebras, the so-called local computation algorithms. The information algebra axioms are fulfilled by many formalisms; some of them have even further common properties. Boolean information algebras and atomic information algebras are both special cases of the information algebra framework, capturing even more, but different, characteristics about the information representing formalism.

5 Domain-Free Information Algebras

The elements of this algebra group all valuations together which represent the same knowledge or information, even on different domains.

Jürg Kohlas
Information Algebras: Generic Structures for Inference

There is another version of information algebras, namely the so-called domain-free version, taking up the ideas of the transport operation (see Definition 4.12). As already seen in the introductory Section 2.1, domain-free pieces of information are not labeled, i. e. they do not belong to a certain domain, but one postulates that every piece of information tells something about every domain. This “something” might also be the neutral information, when no precise statement can be made about any domain. Domain-free does not mean that there is no lattice of domains D needed for the definition, it just tells that we do not dispose of a labeling operation, but look at each piece of information globally, on all domains. Labeled information algebras seen in the previous chapter are generally taken for practical, computational purposes, whereas domain-free information algebras are best used for understanding theoretical concerns. They are completely inappropriate for applications exploiting the local computation algorithms of Section 4.5. Note that the domain-free and the labeled information algebra versions are equivalent. We will now explain how the pieces of information of a labeled information algebra can be transformed into domain-free ones (Section 5.1), leading to a domain-free information algebra (Ψ, D), whose set of axioms is given (Section 5.2). The way back is sketched in Section 5.3.

5.1 From the Labeled to the Domain-Free Information Algebra

In order to convert a labeled information algebra into a domain-free one, a well-known technique in mathematics is used, the quotient construction associated with an equivalence relation on a set. In a labeled information algebra (Φ, D), different pieces of information sometimes tell the same fact, but it is not immediately seen that they do so. Such pieces of information are then called equivalent, expressed by the following equivalence relation σ in Φ, which holds for $\phi, \psi \in \Phi$ with $d(\phi) = x$ and $d(\psi) = y$:

$$\phi \equiv \psi \pmod{\sigma} \quad \text{iff} \quad \phi^{\uparrow x \vee y} = \psi^{\uparrow x \vee y}. \tag{5.1}$$

By the stability axiom and the transport operation (Definition 4.12), it follows from the above Equation 5.1 that

$$\phi \equiv \psi \pmod{\sigma} \quad \text{implies} \quad \psi^{\to x} = \phi \text{ and } \phi^{\to y} = \psi, \tag{5.2}$$

for $\phi, \psi \in \Phi$ with $d(\phi) = x$ and $d(\psi) = y$. If two pieces of information which are related to two different domains are equivalent then the first information transported to the domain of the second one yields in the second piece of information. Moreover, the second piece transported to the domain of the first one yields in the first piece of information.

Elements of Φ which are equivalent regarding σ can be grouped together and form an equivalence class. An equivalence class is constituted by a piece of information $\phi \in \Phi$ and contains all pieces $\psi \in \Phi$ which are equivalent to the former: $[\phi]_\sigma = \{\psi \in \Phi : \phi \equiv \psi \pmod{\sigma}\}$. The family of equivalence classes is called the quotient set of Φ and is written Φ/σ.

In (Kohlas, 2003, Theorem 3.4), the equivalence relation σ is proven to be compatible with the operations of combination and transport, so it is a congruence relative to those two operations. Consider a labeled information algebra (Φ, D). Then, for all $\phi, \psi \in \Phi$, the following two statements hold:

1. If $\phi_1 \equiv \psi_1 \pmod{\sigma}$ and $\phi_2 \equiv \psi_2 \pmod{\sigma}$, then
$$\phi_1 \otimes \phi_2 \quad \equiv \quad \psi_1 \otimes \psi_2 \pmod{\sigma}.$$
2. If $\phi \equiv \psi \pmod{\sigma}$, then for all $x \in D$,
$$\phi^{\to x} \quad \equiv \quad \psi^{\to x} \pmod{\sigma}.$$

So far, a congruence σ in an information algebra (Φ, D) leads to the quotient algebra $(\Phi/\sigma, D)$. The elements of this algebra, the equivalence classes $[\phi]_\sigma$, group all pieces of information together which state the same facts, even on different domains. Therefore, we call $[\phi]_\sigma$ domain-free pieces of information. Obviously, the quotient algebra $(\Phi/\sigma, D)$ does not dispose any more of a labeling operation, as its elements are domain-free. But it has the operation of combination, denoted as before by the symbol $\otimes$, and a second operation, which is called focusing, denoted by $\Rightarrow$. The definitions of these two operations are as follows:

$$[\phi]_\sigma \otimes [\psi]_\sigma \quad := \quad [\phi \otimes \psi]_\sigma, \tag{5.3}$$
$$[\phi]_\sigma^{\Rightarrow x} \quad := \quad [\phi^{\to x}]_\sigma \tag{5.4}$$

Due to the fact that σ is a congruence relative to the operations of combination and transport in the labeled information algebra (Φ, D), the operations given in Equations 5.3 and 5.4 are unambiguously defined. The properties of the two operation are resumed in (Kohlas, 2003, Theorem 3.5). The set of axioms, imposed on a domain-free information algebra in the next section, is an abstraction of the basic properties of these two operations.

So by regrouping pieces of information of a labeled information algebra Φ in equivalence classes, which leads to a quotient algebra Φ/σ, domain-free pieces of information are obtained. As the equivalence relation σ is compatible with the combination and transport operations of Φ, two operations between equivalence classes are defined. These new operations are called combination and focusing and have some interesting and basic properties, which give rise to the domain-free information algebra axioms of the next section.

5.2 Axiomatics of Domain-Free Information Algebras

From Section 4.1, it is known that an information algebra is made up of a lattice of questions or domains D and a set of pieces of information. In the case of domain-free information algebras, this set is denoted by Ψ, the pieces of information contained in Ψ are again designated by Greek lower-case letters such as ϕ, ψ, ζ, but this time they are domain-free. For

- Ψ, a set of domain-free pieces of information and
- D, a lattice of domains,

two basic operations are defined. See Section 2.3.2 for a detailed explanation of their meaning.

Definition 5.1 (Combination)

The combination *operation*

$$\otimes : \Psi \times \Psi \to \Psi, \quad (\phi, \psi) \mapsto \phi \otimes \psi,$$

is used to add two pieces of information $\phi, \psi \in \Psi$ together in order to get the aggregated piece of information.

Definition 5.2 (Focusing)

The focusing *operation*

$$\Rightarrow : \Psi \times D \to \Psi, \quad (\psi, x) \mapsto \psi^{\Rightarrow x},$$

is used to extract from a piece of information ψ the part relevant to an arbitrary question $x \in D$. ψ is focused on the domain of interest x and thereby, information is extracted.

The following set of five axioms is now imposed on (Ψ, D), with the two operations of Definitions 5.1 and 5.2:

1. *Semigroup*:
   The operation of combination is commutative and associative, i. e. for any $\phi, \psi, \zeta \in \Psi$
   $$\begin{aligned} \phi \otimes \psi &= \psi \otimes \phi, \\ (\phi \otimes \psi) \otimes \zeta &= \phi \otimes (\psi \otimes \zeta). \end{aligned}$$
   There is a neutral element $e \in \Psi$ such that $e \otimes \psi = \psi$ for all $\psi \in \Psi$.
2. *Transitivity*:
   For any $\psi \in \Psi$ and $x, y \in D$, $(\psi^{\Rightarrow x})^{\Rightarrow y} = \psi^{\Rightarrow x \wedge y}$.
3. *Combination*:
   For $\psi, \phi \in \Psi$ and $x \in D$, $(\psi^{\Rightarrow x} \otimes \phi)^{\Rightarrow x} = \psi^{\Rightarrow x} \otimes \phi^{\Rightarrow x}$.
4. *Idempotency*:
   For any $\psi \in \Psi$ and $x \in D$, $\psi \otimes \psi^{\Rightarrow x} = \psi$.
5. *Support*:
   For $\psi \in \Psi$, there is an $x \in D$ such that $\psi^{\Rightarrow x} = \psi$.

The first axiom tells that pieces of information can be combined in any order without changing the result; the unique neutral element e does not contain any information, it is the empty or vacuous information. Focusing is transitive as it can be done stepwise. Furthermore, focusing is some sort of distributive over combination, i. e. combining a focused piece of information with any other one and then focusing the result may also be achieved by focusing both pieces of information separately and combining them afterwards. The idempotency axiom ensures that the combination of a piece of information with itself (or a focused version of itself) gives nothing new. Similar to the labeled case, idempotency gives rise to a partial oder in Ψ, which is examined in detail in Section 6.2. Just note that for any two pieces of information $\phi, \psi \in \Psi$ it is expressed in the following way:

$$\phi \leq \psi \quad \text{iff} \quad \phi \otimes \psi = \psi. \tag{5.5}$$

The last axiom states that any piece of information $\psi \in \Psi$ has a support; this is a domain $x \in D$ such that focusing on this domain results in ψ itself. If some piece of information ψ is supported by the domain x, one does not loose information when focusing on x.

Definition 5.3 (Support)

A domain $x \in D$ is called a support *of $\psi \in \Psi$, if*

$$\psi^{\Rightarrow x} = \psi.$$

If ψ has two supports x and x', it has also the support $x \wedge x'$, if this join exists. If D is a complete lattice, ψ has a unique least support y, such that, for any other support $x \in D$ with $x \neq y$, we have $y \leq x$. The neutral element e, for example, has least support $\perp$. We denote the least support of ψ by $\Delta(\psi)$. It is also called dimension set of ψ. The cardinality $|\Delta(\psi)|$ of the least support is called the dimension of ψ. The support axiom has two purposes, a technical one and a semantic one. It is needed for technical reasons, namely for passing from the domain-free version of information algebras to the labeled one. However, the semantic sense of the support axiom is more important, as it states that any piece of information refers to at least one domain or question. See (Kohlas, 2003, Section 3.2) for proofs and more details about (least) supports.

Definition 5.4 (Domain-Free Information Algebra)
A system (Ψ, D) where Ψ is a set of domain-free pieces of information and D a lattice of questions or domains, together with the two operations of combination and focusing (Definitions 5.1 and 5.2), satisfying the preceding five axioms, is called a domain-free information algebra.

5.2.1 Domain-free Information Algebras with Null Element

The contradictory or impossible information may be included in Ψ. It is the null element of combination z. Combining it with any $\psi \in \Psi$ results in z, i. e. contradiction cannot be removed. If the domain-free information algebra (Ψ, D) possesses a null element, a sixth axiom is added, namely the nullity axiom.

6. *Nullity*:
   There is one element $z \in \Psi$ such that for all $\psi \in \Psi$
   $$z \otimes \psi \quad = \quad z$$
   holds. Moreover, for all $x \in D$, $z^{\Rightarrow x} = z$.

5.2.2 Variable Elimination

Just as in the case of labeled information algebras, focusing may be replaced by variable elimination. It is again only possible to introduce the operation of variable elimination in a domain-free information algebra (Ψ, D), if D is a lattice of finite subsets of a set Vbl of variables, as in Section 3.2.2 and in the examples shown in Chapter 2. Taking up Definition 4.7, we express variable elimination in the domain-free case by means of focusing. For $\psi \in \Psi$ and $X \in Vbl$ define

$$\psi^{-X} \quad = \quad \psi^{\Rightarrow Vbl \setminus \{X\}}. \tag{5.6}$$

Three important properties of variable elimination are proven in (Kohlas, 2003, Lemma 3.8):

Lemma 5.5 *Consider a domain-free information algebra (Ψ, D) with D the lattice of all finite subsets of the set of variables Vbl and the variable elimination operation according to Equation 5.6.*

1. Commutativity of Variable Elimination*: For $\psi \in \Psi$ and $X, Y \in Vbl$,*

$$(\psi^{-X})^{-Y} \quad = \quad (\psi^{-Y})^{-X}.$$

2. Combination*: For $\phi, \psi \in \Psi$ and $X \in Vbl$,*

$$(\psi^{-X} \otimes \phi)^{-X} \quad = \quad \psi^{-X} \otimes \phi^{-X}.$$

3. Neutrality*: For $X \in Vbl$,*

$$e^{-X} \quad = \quad e.$$

This allows us to extend the operation of variable elimination to sets of variables $\{X_1, X_2, \ldots, X_n\} \subseteq Vbl$, so that we do not have to be concerned with the actual elimination sequence:

$$\psi^{-\{X_1, X_2, \ldots, X_n\}} \quad = \quad (\cdots((\psi^{-X_1})^{-X_2})\cdots)^{-X_n}. \tag{5.7}$$

Let us conclude with a word on variable elimination and focusing. Variable elimination is only defined for finite sets of variables, whereas it is theoretically possible to focus a piece of information on some countable set of variables $x \in D$. Therefore, in general, variable elimination is less powerful than focusing. However, for information algebras where D is a lattice of finite subsets of the set of variables Vbl, both operations are interchangeable without loss of generality. In this case, the set of axioms formulated by means of variable elimination is equivalent to the one presented in the beginning of this section. It can be obtained in the same way as it was shown for labeled information algebras in Section 4.3.

5.3 Reconstruction of the Labeled Information Algebra

In the Section 5.1, we have explained how the pieces of information of a labeled information algebra can be transformed into domain-free ones. They are part of a quotient algebra whose operations correspond to the operations of a domain-free information algebra. So a domain-free information algebra (Ψ, D) could be constructed from a labeled information algebra (Φ, D). Based on (Kohlas, 2003, Chapters 3.2 & 3.3), we will sketch in this section a procedure which allows to reconstruct again a labeled information algebra (Φ^*, D) from a domain-free one.

Contrary to domain-free pieces of information, labeled pieces of information are related to a specific domain $x \in D$. Therefore, we will now "label" a domain-free piece of information $\psi \in \Psi$ by its support $x \in D$. This leads to a pair (ψ, x), where $\psi^{\Rightarrow x} = \psi$. These pairs are considered as the pieces of information of the reconstructed labeled information algebra.

$$\Phi^* \quad := \quad \{(\psi, x) : \psi \in \Psi, \psi^{\Rightarrow x} = \psi\}. \tag{5.8}$$

The following operations are defined on Φ^* and D, by means of the operations of combination (see Definition 5.1) and of focusing (see Definition 5.2) on domain-free information algebras, which are used on the right hand side:

1. *Labeling:* For $(\psi, x) \in \Phi^*$,
$$d((\psi, x)) \quad := \quad x.$$
2. *Combination:* For $(\phi, x), (\psi, x) \in \Phi^*$,
$$(\phi, x) \otimes (\psi, y) \quad := \quad (\phi \otimes \psi, x \vee y).$$
3. *Marginalization:* For $(\psi, x) \in \Phi^*$ and $y \leq x$,
$$(\psi, x)^{\downarrow y} \quad := \quad (\psi^{\Rightarrow y}, y).$$

Combination and marginalization give indeed results in Φ^* and it can be verified that they define, together with labeling, a labeled information algebra (Φ^*, D).

Furthermore, when we start with a labeled information (Φ, D), as introduced in Chapter 4, and construct the corresponding domain-free information algebra, as explained in Section 5.1, the above procedure allows us to obtain again a labeled information algebra (Φ^*, D), which is up to isomorphism the same as the original one, (Φ, D).

5.4 Conclusion

Domain-free pieces of information are not related directly to some domain, but they express information about the over-all situation. They can be focused on any possible question and can be combined. This leads to a two-sorted algebraic structure (Ψ, D), called domain-free information algebra. Ψ is a set of domain-free pieces of information and D is a lattice of domains or questions. The two basic operations which are used to manipulate elements of Ψ are combination and focusing. If the elements of D are sets of variables, an alternative to the focusing operation is variable elimination. Domain-free information algebras and labeled information algebras are two equivalent ways of modeling information and one can switch between them as one wishes. The domain-free point of view is more appropriate for theoretical considerations, whereas the labeled variant is more convenient for computational issues.

6
Natural Properties of Information

To live effectively is to live with adequate information.

Norbert Wiener (1894 –1964)
The Human Use of Human Beings

In the foregoing Chapters 4 and 5, the properties of information have been captured in a mathematical, algebraic way by presenting the labeled and the domain-free version of information algebras. Based on this description, we will now point out further properties which are very natural in the sense that everybody agrees with them, even people without a mathematical background. However, these properties are anchored in the information algebra framework, they are part of the algebraic theory of semantic information. Thus, we will deduce them from the framework introduced before. We start with the often neglected fact that information always relates to a question and maybe also prior information (Section 6.1). The information algebra framework, the combination operation in particular, allows to order information partially. This partial order of information is described in Section 6.2. Impossible information takes a prominent position in this partial order. This is why we will cast a glance at it in Section 6.3. At the end (Section 6.4), atomic information is examined, which relies on a specialized version of information algebras.

6.1 Relativity of Information

There are two basic principles of relativity of information mentioned in (Kohlas, 2002):

1. Information is always relative to a question.
2. The content of information is always relative to prior information.

The first principle is often forgotten. (Shannon, 1948) only implicitly says that information pertains to a question. In his theory, the question is fixed, one asks for the symbol sent over a communication channel. Our extension of this theory of information states explicitly that a piece of information always relates to a certain question. However, the question is not fixed. When doing information processing like query answering, where information is retrieved from an information base, whole systems of interrelated questions are considered. So, questions are a core element of the present theory. Section 6.1.1 is entirely dedicated to questions. The second principle points out that a piece of information is not to be considered separately, but always in relation to what has been known before, in particular when its content is to be measured. This will be looked at in more detail in Section 6.1.2. The content of the following subsections is based on Chapters 3 to 5. So the above two principles are fundamental properties of our algebraic theory of semantic information.

6.1.1 Relativity to a Question

Chapter 3 has already been dedicated to questions. It turned out that a question is represented by the set of its possible answers, called its frame. There is a refinement relation between questions; the elements of a finer frame provide more detailed answers than those of a coarser frame. This can be seen as a partition where each answer (block of the partition) of the coarser frame has been decomposed into some of the blocks making up the finer frame. Partitions are known to form a lattice, and this also holds for frames since the set of frames is embedded in the set of partitions, see Section 3.2.1. In the context of information algebras, this lattice of questions is called D. It will turn out that the first principle of relativity of information to a question is closely related to one of the operations of the information algebra framework. In the labeled case, it is the transport operation $\rightarrow$, in the domain-free case, the focusing operation $\Rightarrow$. We will now look at the labeled and domain-free pieces of information and their relation to questions separately.

Labeled Information

In the labeled version of the information algebra framework (see Chapter 4), it is very obvious that information always pertains to a question. Given a labeled information algebra (Φ, D), the label or domain $d(\phi)$ of a piece of information $\phi \in \Phi$ is the question this piece of information refers to. The fact that ϕ provides information concerning some precise question turns it into a (possibly partial) answer to this question.

Marginalizing a piece of information ϕ means relating it to a coarser question, which gives rise to less detailed answers. So information extraction is actually the consideration of another question $x \leq d(\phi)$. The transport operation allows to relate a piece of information ϕ to any possible question $x \in D$, not only to a coarser one. In the extreme case, when $d(\phi)$ and the chosen x have nothing in common, ϕ bears no information with respect to x, $\phi^{\rightarrow x} = e_x$, as seen in Example 2.3.5.

Domain-Free Information

Things get more involved in the domain-free case. As explained in Chapter 5, in a domain-free information algebra (Ψ, D), a piece of information $\psi \in \Psi$ does not dispose of an associated domain which could be retrieved by means of a labeling operation. There is no such labeling operation as the underlying idea is that each piece of information tells something about every domain. This might also be the neutral information if no precise statement besides "nothing is known about this domain" can be made. Nevertheless, the support axiom of domain-free information algebras allows to determine the questions the domain-free piece of information ψ is related to. Definition 5.3 states that a question $x \in D$ is a support of ψ if focusing ψ on this domain x does not change ψ. If D is a complete lattice, each $\psi \in \Psi$ has a least support $\Delta(\psi)$ which is coarser than every other support of ψ. ψ does not state anything which does not refer to $\Delta(\psi)$.

The focusing operation allows to relate a domain-free piece of information $\psi \in \Psi$ to an arbitrary question $x \in D$. It allows to extract information relative to any field of interest. If x is not a support of ψ, this results in a less informative piece of information $\psi^{\Rightarrow x}$, as we will see in Section 6.2.

6.1.2 Relativity to Prior Information

We have seen above that the first principle of relativity of information is linked to the focusing or the transport operation of information algebras. The second principle of relativity of information relates it to prior information. This principle is associated with the combination operation $\otimes$ of either version of information algebra. Thus we do not have to distinguish between labeled and domain-free pieces of information.

Information should not be taken as a fixed entity since it is likely to change over time. If the current knowledge is given by the piece of information ψ and a piece of information ϕ is learned, the current knowledge changes and is now described by $\psi' = \psi \otimes \phi$. The fact that information usually comes in a sequence of pieces of information gives rise to the concept of prior information, which is of particular importance when information is measured. In Section 7.4 we will see that information content is measured by reduction of uncertainty. Depending on which information ψ is already available, the arrival of another piece of information ϕ will more or less reduce the uncertainty of the specific situation. Obviously, if $\phi = \psi$, the uncertainty is not reduced at all. This is why the measure of information content is always relative to prior information.

6.2 Partial Order of Information

A fundamental characteristics of the combination of information is that it is idempotent. Knowing some piece of information and learning the same piece of information again, or only a part of it, does not change what is known. The idempotency axiom of information algebras permits to introduce a partial order in the set of pieces of information. In this section, we will only take the domain-free information algebra into account. However, at the end, references to the partial order in labeled information algebras are given.

Let (Ψ, D) be a domain-free information algebra. A piece of information $\phi \in \Psi$ is less informative than another piece of information ψ from Ψ, written $\phi \leq \psi$, if their combination yields ψ. Formally, for $\phi, \psi \in \Psi$,

$$\phi \leq \psi \quad \text{iff} \quad \phi \otimes \psi = \psi. \tag{6.1}$$

The idea is that ϕ does not add anything new to ψ, since it is already contained in ψ. The $\leq$ relation of being less informative defines a partial order in Ψ:

1. *Reflexivity:* $\phi \leq \phi$.
2. *Antisymmetry:* $\phi \leq \psi$ and $\psi \leq \phi$ imply $\phi = \psi$.
3. *Transitivity:* $\phi \leq \psi$ and $\psi \leq \zeta$ imply $\phi \leq \zeta$.

Reflexivity holds by the idempotency axiom: $\phi \otimes \phi = \phi$. Antisymmetry is given by definition: $\psi = \phi \otimes \psi = \phi$. Transitivity follows from $\zeta = \psi \otimes \zeta = ((\phi \otimes \psi) \otimes \zeta) = (\phi \otimes (\psi \otimes \zeta)) = \phi \otimes \zeta$. So, not only are questions partially ordered as seen in Chapter 3, but pieces of information, too. Some important results concerning this ordering are listed in the following lemma. The corresponding proofs are those of (Kohlas, 2003, Lemma 6.2):

Lemma 6.1 *For a domain-free information algebra (Ψ, D), any $\phi, \psi \in \Psi$, $x, y \in D$ and the partial order $\leq$ defined by Equation 6.1, the following properties hold:*

1. $\phi, \psi \leq \phi \otimes \psi = \sup\{\phi, \psi\}$.
2. $\psi^{\Rightarrow x} \leq \psi$.
3. $x \leq y$ *implies* $\psi^{\Rightarrow x} \leq \psi^{\Rightarrow y}$.
4. $\phi \leq \psi$ *implies* $\phi^{\Rightarrow x} \leq \psi^{\Rightarrow x}$.
5. $e \leq \psi \leq z$ *for all* $\psi \in \Psi$.

The first property states that any finite set of elements from Ψ has a least upper bound in Ψ. We can derive that Ψ is a semilattice. The combination of two pieces of information is their join. Another quite natural attribute is expressed by property 2 of this lemma, namely that a part of a piece of information is less informative than the whole of it. Similarly, the third property tells that a piece of information relative to a coarser question is less informative than the same piece of information pertaining to a finer question. Moreover, the fourth property states that focusing does not affect an earlier established order between pieces of information, the order is maintained. Finally, the fifth and last property introduces the bottom element of this partial order, the neutral element e. If (Ψ, D) is an algebra satisfying the nullity axiom, there is a null element $z \in \Psi$, which is then the top element.

Note that a partial order may be introduced in a similar way into a labeled information algebra (Φ, D). In Section 4.2, the definition of the partial order $\leq$ was already given (Equation 4.3), which is exactly the same as that of Equation 6.1. However, in the labeled case, the partial order makes not so much sense in the whole set Φ of all pieces of information, but rather in the set Φ_x of all pieces of information bearing on a specific domain $x \in D$. The partial order in Φ would allow statements like $\phi \leq \phi^{\uparrow x}$, for $x > d(\phi)$, which is questionable, since the vacuous extension does not add any information. Some properties of this partial order can be found in Section 7.2, but a more comprehensive listing is available in (Kohlas, 2002, Lemma 13) and (Kohlas, 2003, Lemma 6.3), where also the corresponding proofs can be found.

6.3 Contradictory Information

In Section 5.2.1, we have seen the nullity axiom of domain-free information algebras, which is only satisfied by some information algebra instances, as the ones introduced in Chapters 8 and 9. If this sixth axiom is satisfied, the null element z is included in the set of all pieces of information Ψ. However, this is not really a piece of information in the sense of having a certain information content. It is somehow no information and is merely included in Ψ for technical reasons and computational purposes. z is often called the contradictory information and it was stated that $\psi \otimes z = z$ for all $\psi \in \Psi$. By the definition of $\leq$, this also means that every piece of information is less "informative" than the contradictory information z: $\psi \leq z$ for all $\psi \in \Psi$. The contradictory information z is the most "informative" piece of information of the partial order.[1] Everything that is known is contained in z (idempotency axiom). The same holds for the null element z_x of every domain $x \in D$ of a labeled information algebra (Φ, D).

The idempotency property of information explains the absorbing nature of the contradictory information. The contradictory piece of information z may result from the combination of two pieces of information (see Equation 4.5). By $z \otimes \psi = z$ we know that if information becomes contradictory, it remains so and cannot be changed by further combinations. Contradiction is a state one can never get out.

6.4 Atomic Information

In Section 4.7.1, atomic pieces of information (or atoms for short) have been introduced. The concept of atom makes only sense in a labeled information algebra (Φ, D), as we are always considering an atom on some domain $x \in D$. By definition,

[1] This situation is very well reflected in logics, where everything can be deduced from the contradiction. The contradiction is the most informative statement, from which any other, and therefore less informative statement can be derived.

an atom α on a domain x cannot be the null element z_x of this domain. No piece of information of this domain can be more informative than the atom, except the null information. Semantically, an atom is the most informative piece of information on this domain which is not yet contradictory. It is not unique, there might be several atoms on a domain $x \in D$; together, they form the set $At_x(\Phi)$. Furthermore, for any piece of information $\phi \in \Phi$, the set which contains all the atoms relative to ϕ is denoted by $At(\phi)$, see Equation 4.16. In addition to the properties of Lemma 4.17, there are some results on atoms which will play an important role later on, when information is measured Chapter 7. The properties listed in the following lemma are proven in (Schneuwly, 2007, Lemma 8.10):

Lemma 6.2 *Let (Φ, D) be an atomic composed information algebra. Then, the following properties hold for any $x \in D$:*

1. $At(e_x) = At_x(\Phi)$.
2. $\bigwedge At_x(\Phi) = e_x$.
3. $At(z_x) = \emptyset$.

The set of atoms relative to the neutral information e_x is the set of all atoms on x. The infimum of this set exists and it is actually equal to e_x. Clearly, there is no atom relative to the contradictory information z_x.

6.5 Conclusion

The first principle of relativity of information tells that a piece of information is always relative to a question. In the case of labeled information algebras, this question is determined by the labeling operation. A domain-free piece of information refers to the question indicated by any of its supports. In order to change the question, transport and focusing are used. The second principle of relativity of information tells that the measure of information content is always relative to prior information. This principle is related to the operation of combination. In an information algebra, the pieces of information can be partially ordered. This partial order $\leq$ is made possible by the idempotency axiom of information algebras. Idempotency is perhaps the most basic property of information. The contradictory information z is not a piece of information in the proper sense as it does not provide any information content. Its raison d'être is computational and technical. It is e. g. used for the definition of atoms which are the most informative pieces of information apart from contradiction in the partial order $\leq$.

7 Measure of Information

Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

Claude Elwood Shannon (1916 – 2001)
A Mathematical Theory of Communication

There are two different viewpoints or interpretations of pieces of information. They have been first pointed out in (Schneuwly, 2007). We call them disjunctive and conjunctive point of view. These terms tell how the set of tuples or valuations making up a semantic piece of information have to be interpreted. A disjunctive interpretation of a piece of information can be viewed as a scheme of choice. One of its elements, which is however unknown, is the right one, which describes the real world. The other possibility of interpretation, the conjunctive view, is an enumeration of everything that is possible in the situation described by a piece of information. These two points of view are examined and exemplified in Section 7.1.

In both ways of interpretation, information content can either be measured qualitatively or quantitatively. In the first, qualitative, case (Sections 7.2 and 7.3), one often speaks rather of an order of information than of a measure. The order expresses that some piece of information is more or less informative than another one. The second case, the quantitative measure, provides a numeric value for the information content of a piece of information (Sections 7.4 and 7.5).

The measures of information proposed in this chapter are all based on the idea of relativity of information. In the previous chapter, natural properties of information have been pointed out, which have also to be considered when information is measured. Information was shown to be relative to a question in Section 6.1, so the qualitative and the quantitative measures will be formulated regarding a question. Furthermore, when information is measured quantitatively, it will be considered relative to prior information. The quantitative measure can be traced back to the *change of uncertainty* in a choice situation, involving entropy. This measure is first established without attributing any probabilities. In Section 7.6 it is extended by also taking probabilities into account. It is important to point out that not only pieces of information carry information content, but also the probability distribution over the elements which a semantic piece of information is composed of. Probability distributions also represent information.

Unless otherwise stated, we will assume in this chapter pieces of information ϕ, ψ in a labeled information algebra (Φ, D). Sometimes, further properties are added, such as the information algebra being atomic and / or Boolean. See Chapter 4 for more details. The lattice D will always dispose of a top domain $\top$, and for all $x \in D$ the frames $\mathfrak{D}_x$ are supposed to be finite. The qualitative measures of Sections 7.2 and 7.3 could also be seen from a domain-free viewpoint, which will be sketched in footnotes. However, for establishing a quantitative measure, the information algebra has to be atomic, a concept which is only meaningful in labeled information algebras.

The closed world assumption (see below) is preconditioned in this chapter. A justification of this assumption is that the information representing formalisms we are interested in, especially the examples in Chapters 8 and 9, fulfill this condition and so we can restrict our measure to such cases. Furthermore, efficient computing can only be done when the stipulation is met.

7.1 Disjunctive and Conjunctive View

In Section 2.4, semantic pieces of information were given by sets of tuples, leading to an information algebra (Φ, D). It depends on the situation how this set of tuples is interpreted – disjunctively or conjunctively.[1] We will now give further insight into both viewpoints and end with an example. In order to be more explicit and with regard to the second and the third part of this thesis, we will assume in this section D to be a lattice of subsets of a set Vbl of variables. Recall that, for every set of variables $x \in D$, there is a corresponding frame $\mathfrak{D}_x$. The elements of this frame are called tuples. Such a tuple of $\mathfrak{D}_x$ provides a value for every variables in x. The semantic pieces of information considered in this section are sets of tuples.

As a tuple ascribes a unique value to every variable under consideration, it offers a detailed description of how the world (i. e. the situation modeled) is like. Often we simply say that the world is specified by a tuple. So a world, as we use the term, tells the actual values of the variables involved. Clearly, there are as many possible worlds as there are different settings of variables. In a given problem or question one may assume that there is some correct, but unknown description of the real world, provided by one tuple. When some semantic piece of information $\phi \in \Phi$ is given, its tuples are called the possible worlds. They are considered as the information relative to the unknown values of the variables under consideration, postulating that the real world must belong to the set of possible worlds. The tuples which are not in ϕ are excluded as correct descriptions of the real world. This is the so-called closed world

[1] The same holds for domain-free semantic pieces of information which are given by sets of valuations, also known from Section 2.4.

assumption. Therefore, a semantic piece of information $\phi \in \Phi$ induces a bipartition in $\mathfrak{D}_{d(\phi)}$, by selecting some tuples to be the possible worlds.[2]

7.1.1 Disjunctive View

A concrete problem consists in selecting one of the possible worlds of a semantic piece of information $\phi \in \Phi$. We are thus dealing with a situation of uncertainty and the piece of information constitutes a scheme of choice over the set ϕ of possible worlds.

Definition 7.1 (Scheme of Choice / Choice System)

In a case where n different possibilities exist, denoted by $e_1, e_2, \ldots, e_n$, $n \in \mathbb{N}$, the set of possibilities $S = \{e_1, e_2, \ldots, e_n\}$ is called a (finite) scheme of choice *or a* (finite) choice system*.*

The idea is that somebody or something (like a process or a mechanism) selects one of these possibilities. A choice system describes a situation of uncertainty, as we do not know which one of the n possibilities will be selected. The disjunctive view of a semantic piece of information ϕ means considering it as a scheme of choice. The possible worlds are the tuples of ϕ. One of the tuples $f_i \in \phi$, $1 \leq i \leq n \in \mathbb{N}$ will be selected to describe the real world: "either f_1, or f_2, or ... or f_n". By choosing exactly one tuple, the world is uniquely described.[3]

7.1.2 Conjunctive View

Sometimes, one is not so much interested in selecting one of the possibilities, but one wants to know all the possibilities provided. In this case, the piece of information ϕ is not a scheme of choice, but the tuples constituting ϕ are an enumeration of everything that is possible in this very situation. This is the conjunctive view of the possible worlds and $\phi = \{f_1, \ldots, f_n\}$ tells that "f_1 and f_2 and ... and f_n" are all given at the same time and describe only together what is known about the real world.[4]

7.1.3 Example

A semantic piece of information is given by a set of tuples. In the case of the disjunctive view of information, the tuples set up a scheme of choice. In the conjunctive

[2] A domain-free semantic piece of information $\psi \in \Psi$ is constituted of valuations. A valuation provides a value for every variable in Vbl and acts as a possible world. ψ induces a bipartition in $\mathfrak{D}_{Vbl}$.

[3] In a domain-free setting, a valuation of a semantic piece of information $\psi \in \Psi$ is chosen and then determines the real world.

[4] In the domain-free case, an enumeration of valuations constituting a piece of information ψ is considered.

view of information, they are an enumeration of possibilities. The following example takes up the job interview scenario which guided us through Chapter 2.

Example 7.1.1 (Disjunctive / Conjunctive View) In Figure 2.2, the situation of four candidates invited to the interview is illustrated by the four black nodes, each of them representing one tuple. These four tuples are the ones which describe how the (real) world will possibly be. All the other descriptions of the world given by the remaining tuples (empty nodes) are excluded, according to the closed world assumption. The black nodes are the possible worlds and make up a bipartition of Φ.

Consider the concrete problem "Who will get the job?". Only one applicant will be employed. So we are dealing with the disjunctive view of the possible worlds and the set of four tuples constitutes a scheme of choice. The tuple describing the candidate that will be chosen is going to be the real world.

The conjunctive view of the possible worlds arises when the problem consists of knowing which candidates are interesting, among those people who sent their letter of application. The problem consists in asking "Who is invited to the job interview?". In this case, the set of possible worlds is an enumeration of the four possible tuples, each of them describing one candidate having been invited. The real world is the set of all four tuples. $\odot$

7.2 Qualitative Measure: Disjunctive View

If two pieces of information are considered, their information content may be compared regarding their informativeness. Based on (Kohlas, 2002) and (Kohlas, 2003), where omitted proofs can be found, we first present an absolute qualitative measure of information content, followed by a qualitative measure which compares the information content of two pieces of information relative to a fixed question.

7.2.1 Partial Order - an Absolute Measure

The partial order introduced in Section 6.2 defines a qualitative measure of information content.

Definition 7.2 (Partial Order: Qualitative Measure)

Let (Φ, D) be a labeled information algebra and $\phi, \psi \in \Phi$. ϕ is less informative *than ψ (denoted by $\phi \leq \psi$) if their combination yields ψ. Formally,*

$$\phi \leq \psi \quad \textit{iff} \quad \phi \otimes \psi = \psi.$$

The idea is that the first piece of information (the less informative one) does not add any information to the second one.[5] Some elementary results of this partial order

[5] As already pointed out in Section 6.2, the partial order between labeled pieces of information makes more sense when it is only considered between pieces of information bearing on the same domain.

are given in the following lemma. The proofs, as well as further properties, can be found in (Kohlas, 2002, Lemma 13) and (Kohlas, 2003, Lemma 6.3).

Lemma 7.3 *For two pieces of information ϕ, ψ in a labeled information algebra (Φ, D) and an arbitrary domain $x \in D$, the properties listed below hold:*

1. $\phi, \psi \leq \phi \otimes \psi$.
2. $\phi \otimes \psi = \sup\{\phi, \psi\}$.
3. $\phi^{\to x} \otimes \psi^{\to x} \leq (\phi \otimes \psi)^{\to x}$.
4. $\phi \leq \psi$ *implies* $\phi^{\to x} \leq \psi^{\to x}$.

Lemma 7.3 states that a single piece of information is always less informative than or as informative as its combination with another piece of information. It is pointed out that a finite set of elements from Φ has a least upper bound in Φ, i. e. Φ is a semilattice where the supremum of two elements is their combination. The combination of two pieces of information which have separately been transported to the same domain is less informative than the transport of their combination to this very domain. The transport operation is coherent with the proposed qualitative measure.[6]

By this partial order, the elements of an information algebra may be compared relative to their absolute information content. A qualitative measure of information content is established. It is absolute in the sense of being independent of any particular, fixed question.

7.2.2 Information Relative to a Question

According to the first principle of relativity of information (see Section 6.1), we may also want to compare the information content of the elements of an information algebra with respect to a determined question, i. e. a fixed domain $x \in D$. Therefore, the pieces of information have to be transported to x, according to Definition 4.12, and can then be compared using the partial order of Definition 7.2. This qualitative measure regarding a question x is expressed by $\leq_x$:

Definition 7.4 (Qualitative Measure Regarding a Question)

Let (Φ, D) be a labeled information algebra, $x \in D$ and $\phi, \psi \in \Phi$. ϕ is said to be less informative *than ψ* relative to question *x,*

$$\phi \leq_x \psi, \quad \textit{iff} \quad \phi^{\to x} \leq \psi^{\to x}.$$

[6]In a domain-free information algebra (Ψ, D), the partial order is given in exactly the same way. For $\phi, \psi \in \Psi$, we define $\phi \leq \psi$ iff $\phi \otimes \psi = \psi$. Similar properties as the ones given in Lemma 7.3 can be found for domain-free information algebras in Lemma 6.2 of (Kohlas, 2003). In most cases, the transport operation $\to$ of labeled information algebras is simply replaced by the focusing operation $\Rightarrow$ of domain-free information algebras.

Obviously, this definition may be rewritten by

$$\phi \leq_x \psi \quad \text{iff} \quad \phi^{\rightarrow x} \otimes \psi^{\rightarrow x} = \psi^{\rightarrow x}, \tag{7.1}$$

according to Lemma 7.3. Actually, $\leq_x$ is only a pre-order, and not a partial order, i. e. it has to be proven to be reflexive, transitive, but not antisymmetric.

Theorem 7.5 *$\leq_x$ is a pre-order in the labeled information algebra (Φ, D).*

Proof.

- Reflexivity: $\phi \leq_x \phi$.

  With Equation 7.1 and the idempotency axiom, it holds that

  $$\phi \leq_x \phi \Leftrightarrow \phi^{\rightarrow x} \otimes \phi^{\rightarrow x} = \phi^{\rightarrow x}.$$

- Transitivity: $\phi \leq_x \psi$ and $\psi \leq_x \zeta$ imply $\phi \leq_x \zeta$.

  By Equation 7.1 it holds that

  $$\begin{aligned} \phi \leq_x \psi \quad &\Leftrightarrow \quad \phi^{\rightarrow x} \otimes \psi^{\rightarrow x} = \psi^{\rightarrow x} \text{ and} \\ \psi \leq_x \zeta \quad &\Leftrightarrow \quad \psi^{\rightarrow x} \otimes \zeta^{\rightarrow x} = \zeta^{\rightarrow x}. \end{aligned}$$

  Making use of the associativity and commutativity properties of the combination operation and replacing $\psi^{\rightarrow x}$ and $\zeta^{\rightarrow x}$ by the above equalities, the following holds:

  $$\begin{aligned} \zeta^{\rightarrow x} &= \zeta^{\rightarrow x} \otimes \psi^{\rightarrow x} \\ &= \zeta^{\rightarrow x} \otimes (\phi^{\rightarrow x} \otimes \psi^{\rightarrow x}) \\ &= \zeta^{\rightarrow x} \otimes (\psi^{\rightarrow x} \otimes \phi^{\rightarrow x}) \\ &= (\zeta^{\rightarrow x} \otimes \psi^{\rightarrow x}) \otimes \phi^{\rightarrow x} \\ &= \zeta^{\rightarrow x} \otimes \phi^{\rightarrow x} \end{aligned}$$

  So we see that $\phi^{\rightarrow x} \otimes \zeta^{\rightarrow x} = \zeta^{\rightarrow x}$, which corresponds by Equation 7.1 to $\phi \leq_x \zeta$, q.e.d.

- Counter-Example for Antisymmetry: $\phi \leq_x \psi$ and $\psi \leq_x \phi$ do not imply $\phi = \psi$.

  By Equation 7.1 we know that

  $$\begin{aligned} \phi \leq_x \psi \quad &\Leftrightarrow \quad \phi^{\rightarrow x} \otimes \psi^{\rightarrow x} = \psi^{\rightarrow x} \\ \psi \leq_x \phi \quad &\Leftrightarrow \quad \psi^{\rightarrow x} \otimes \phi^{\rightarrow x} = \phi^{\rightarrow x} \end{aligned}$$

  However, this only implies that $\phi^{\rightarrow x} = \psi^{\rightarrow x}$, but nothing is said about the values for $Vbl \setminus x$. Consider the case where $\phi = \{(a_1, b_1, c_1), (a_1, b_1, c_2)\}$ and $\psi = \{(a_1, b_1, c_3), (a_1, b_1, c_4)\}$, with $d(\phi) = d(\psi) = \{v_1, v_2, v_3\}$. For $i = 1, \ldots, 4$, $a_i \in \mathfrak{D}_{v_1}$, $b_i \in \mathfrak{D}_{v_2}$, $c_i \in \mathfrak{D}_{v_3}$. If $x = \{v_1, v_2\}$, then $\phi^{\rightarrow x} = \psi^{\rightarrow x} = \{(a_1, b_1)\}$. But clearly, $\phi \neq \psi$. □

Some properties of $\leq_x$ are summarized in Lemma 7.6 below. See (Schneuwly, 2007, Lemmata 8.5 and 8.6) for its proof:

Lemma 7.6 *Consider a labeled information algebra (Φ, D), two pieces of information $\phi, \psi \in \Phi$ and three domains $x, y, z \in D$. Then, for an arbitrary, but fixed $x \in D$, the pre-order $\leq_x$ has the following properties:*

1. $e_y \leq_x \phi \leq_x z_y$, *if* $d(\phi) = y$.
2. $\phi, \psi \leq_x \phi \otimes \psi$
3. $\phi \leq \psi$ *implies* $\phi \leq_x \psi$
4. $\phi^{\rightarrow y} \otimes \psi^{\rightarrow y} \leq_x (\phi \otimes \psi)^{\rightarrow y}$
5. $y \leq z$ *implies* $\phi^{\rightarrow y} \leq \psi^{\rightarrow z}$ *and hence* $\phi^{\rightarrow y} \leq_x \psi^{\rightarrow z}$

The above lemma states several interesting and natural characteristics of information ordering relative to questions.: Relative to an arbitrary question, a piece of information is always more informative than the vacuous information and less informative than the contradictory information. As for the partial order, a single piece of information is always less informative than or as informative as its combination with another one. If two pieces of pieces of information obey the partial order, they also obey the pre-order. Another property which can also be found in Lemma 7.3 for the partial order is that combining after transporting to some question of interest is in general less informative than combining the information first and then transporting it to this very question, which is not necessarily the same question as the one which determines the information order. Finally, the order in the lattice D is respected not only by the partial order, but also by the pre-order.[7]

7.3 Qualitative Measure: Conjunctive View

Looked at from the conjunctive point of view, a piece of information is an enumeration of possibilities. One obviously feels to dispose of more information the longer the list of possibilities is. Here, we are in the inverse situation of the one described in Section 7.2. Information content seems to depend on how the pieces of information of an information algebra (Φ, D) are interpreted, in other words, whether a disjunctive answer (scheme of choice) or a conjunctive answer (enumeration) is expected. The second scenario only appears in cases where Φ provides more structure and for all $x \in D$, the set Φ_x of pieces of information bearing on x is a Boolean lattice.[8] Then, we are dealing with a Boolean information algebra (see Section 4.6). In the

[7] As for the partial order above, the pre-order in a domain-free information algebra (Ψ, D) is similar to the one proposed in the labeled case. It is defined for $\phi, \psi \in \Psi$ and $x \in D$ by $\phi \leq_x \psi$ iff $\phi^{\Rightarrow x} \leq \psi^{\Rightarrow x}$.

[8] In the domain-free case, the set Ψ of all pieces of information is a Boolean lattice.

following, we take advantage of the duality known and carried over from Boolean lattices. The information algebras associated with propositional and predicate logic (Chapters 8 and 9) are Boolean.

Based on Section 4.6 and (Schneuwly, 2007), where this conjunctive view has been pointed out for the first time, we start with the introduction of an absolute qualitative measure. It is $\leq^d$, the dual partial order, and accounts for the issue of comparing qualitatively pieces of information interpreted as enumerations. Such a piece of information ϕ is the more informative the less tuples are not included in ϕ. So $\leq^d$ involves the complement of a piece of information. In a second step, we relate the proposed measure to a question, which gives rise to the dual pre-order $\leq^d_x$, comparing two enumerations relative to a question x.

7.3.1 The Dual Partial Order

In the same manner as in Section 7.2, the dual partial order $\leq^d$ will be expressed by means of the combination of two pieces of information. However, $\leq^d$ is induced in the dual labeled Boolean information algebra, which is isomorphic to the original one by a mapping $\phi \mapsto \phi^c$, see (Schneuwly, 2007, Theorem 8.1). So the dual combination $\otimes^d$ has to be considered:

Definition 7.7 (Dual Combination)

If ϕ, ψ are two pieces of information in a labeled Boolean information algebra (Φ, D), then the dual combination *is defined by*

$$\phi \otimes^d \psi \quad = \quad (\phi^c \otimes \psi^c)^c.$$

Further dual operations are dual marginalization and dual meet, as given by the following two definitions:

Definition 7.8 (Dual Marginalization)

If ϕ is a piece of information in a labeled Boolean information algebra (Φ, D), $x \leq d(\phi)$, $x \in D$, then the dual marginalization *is defined by*

$$\phi^{\downarrow^d x} \quad = \quad ((\phi^c)^{\downarrow x})^c.$$

Definition 7.9 (Dual Meet)

If ϕ, ψ are two pieces of information in a labeled Boolean information algebra (Φ, D) and $d(\phi) = d(\psi) = x \in D$, then the dual meet *is defined by*

$$\phi \wedge^d_x \psi \quad = \quad (\phi^c \wedge_x \psi^c)^c \quad = \quad \phi \vee \psi.$$

Based on (Schneuwly, 2007, Theorem 8.1), where a mapping $h : \Phi \to \Phi$, $\phi \mapsto \phi^c$ is proven to be a Boolean isomorphism between a labeled Boolean information algebra (Φ, D) and its dual counterpart, we can formulate the following theorem:

Theorem 7.10 (Φ, D), *together with the labeling operator* d, *the dual combination operator* $\otimes^d$, *the dual marginalization operator* $\downarrow^d$, *the dual meet operators* $\wedge_x^d$ *(see the above definitions) and the complement operator* ^{c}, *as given in Definition 4.14, is a labeled Boolean information algebra. It is referred to as* dual labeled Boolean information algebra. *Its neutral element* $e_x^d = z_x$ *and its null element* $z_x^d = e_x$.

The partial order in this dual labeled Boolean information algebra is defined in the usual way, which is already known from Section 6.2. Obviously, the dual combination has to be used:

Definition 7.11 (Dual Partial Order)

Consider a labeled Boolean information algebra (Φ, D) *and its dual counterpart. Let* $\phi, \psi \in \Phi$. *The* dual partial order *is given by*

$$\phi \leq^d \psi \quad \textit{iff} \quad \phi \otimes^d \psi = \psi.$$

We know by definition that the relation $\leq^d$ of being less informative defines a partial order in the dual labeled Boolean information algebra. This allows to introduce a qualitative measure of conjunctively interpreted information among the elements of (Φ, D).

Definition 7.12 (Dual Qualitative Measure)

Consider a labeled Boolean information algebra (Φ, D) *and its dual counterpart. Let* $\phi, \psi \in \Phi$. ϕ *is said to be* less informative in the conjunctive interpretation *than* ψ *iff* $\phi \leq^d \psi$.

See (Schneuwly, 2007, Chapter 8.1.2) for a more detailed description of the properties of the dual partial order.[9]

7.3.2 Dual Order Relative to a Question

Similarly to Section 7.2.2, the dual qualitative measure relative to some question x is defined using the transport operation. This time, however, the dual transport operation is needed. It is defined in the same way as the foregoing dual operations:

Definition 7.13 (Dual Transport)

Consider a labeled Boolean information algebra (Φ, D) *and its dual counterpart. For* $\phi \in \Phi$ *and* $x \in D$, *the* dual transport $\rightarrow^d$ *is defined by*

$$\phi^{\rightarrow^d x} \quad = \quad ((\phi^c)^{\rightarrow x})^c.$$

[9] In domain-free Boolean information algebras, a duality theory also exists. From (Kohlas, 2003, Chapter 6.5.2) and (Schneuwly, 2007, Chapter 8.1.4) it is known that the partial order and its dual are opposite: $\phi \leq^d \psi$ iff $\phi \geq \psi$, as well as $\phi \leq_x^d \psi$ iff $\phi^{\Rightarrow^d x} \geq \psi^{\Rightarrow^d x}$, where $\phi^{\Rightarrow^d} = ((\phi^c)^{\Rightarrow x})^c$.

Applying Definition 7.4, we are now able to define a qualitative measure relative to an arbitrary question $x \in D$ in the dual algebra:

Definition 7.14 (Dual Qualitative Measure Regarding a Question)

Consider a labeled Boolean information algebra (Φ, D) *and its dual counterpart. Let* $\phi, \psi \in \Phi$ *and* $x \in D$. ϕ *is said to be* less informative *than* ψ in the conjunctive interpretation relative to the question x*,*

$$\phi \leq_x^d \psi, \quad \textit{iff} \quad \phi^{\rightarrow^d x} \leq^d \psi^{\rightarrow^d x}.$$

Based on Lemma 7.5 of the disjunctive case, the relation $\leq_x^d$ is by definition a pre-order in the dual labeled Boolean information algebra. It is called dual pre-order. Note also that the results of Lemma 7.6 can be brought to the dual algebra in a straightforward way.

As already stated before, the dual order compares pieces of information relative to what tuples they exclude. It is easily seen that the dual order relative to a question takes the complement of transported pieces of information ψ into account, as $\psi = (\phi^{\rightarrow x})^c = (\phi^c)^{\rightarrow^d x}$.

7.4 Quantitative Measure

In order to measure information not only qualitatively, but also quantitatively, the basic idea of Shannon's classical information theory (Shannon, 1948) is applied. However, this only works if the labeled information algebra (Φ, D) we are dealing with is atomic composed, see Section 4.7. Then, every piece of information $\phi \in \Phi$ is merely composed of atoms. Shannon established a theory of communication where symbols are transmitted over a channel. He measures the information carried by a transmitted symbol be the reduction of uncertainty stemming from the fact that this very symbol becomes known. From our point of view, Shannon considers a fixed question: "Which symbol out of an alphabet is selected for transmission?". Note that the alphabet is considered to be finite. Shannon proposes to measure the reduction of uncertainty by the entropy of the alphabet. Once the symbol transmitted becomes known, the uncertainty is reduced to zero. Entropy thus measures the information that is gained when the selected symbol becomes known.

Before going into details, some general remarks have to be made. As already pointed out above, we will always consider an atomic labeled information algebra (Φ, D). For the rest of this chapter, we furthermore assume that for all domains $x \in D$ the set of all atoms $At_x(\Phi)$ of the domain is finite.

The information content of a piece of information $\phi \in \Phi$ is first measured relative to its domain $d(\phi)$ in Section 7.4.1. Clearly, the domain is the question the piece of information refers to. As a piece of information ϕ contains possibly information about some other question $x \neq d(\phi)$, we will see in Section 7.4.2 how to measure ϕ's information content relative to a specified question. In order to cover not only the first principle of information, but also the second one, the information measure will also take prior information into account in Section 7.4.3. Even if this section is based on Shannon's ideas, we will not consider random experiments and avoid probabilistic considerations. The main ideas of a probability-based measure are subject of Chapter 7.6.

A detailed application of the theory proposed below can be found in (Langel & Kohlas, 2005) for propositional logic. The ideas in this section are motivated by (Hartley, 1928; Shannon, 1948; Kohlas, 2002; Schneuwly, 2007). Omitted proofs can be found in the last two references.

7.4.1 Hartley Measure and Reduction of Uncertainty

According to Sections 4.7 and 7.2, an atom of a domain $x \in D$ is the finest (maximal) information one may obtain about this domain. The coarsest (minimal) information about the domain x is set of all atoms $At(e_x) = At_x(\Phi)$, which is assumed to be finite. Thus, the total number of possible atoms is given by $|At(e_x)|$. This corresponds to what Hartley calls "the number of distinguishable sequences" and denotes by s^n in (Hartley, 1928):

> "Suppose that [...] one is provided [with a system] in which an arbitrary number s of different current values can be applied to the line and distinguished from each other at the receiving end. Then the number of symbols available at each selection is s and the number of distinguishable sequences is s^n."

The idea underlying the measure Hartley proposes is that the amount of information carried by an atom on x is determined by the number of yes-no-questions to be asked to find out an arbitrarily selected, but initially unknown element from $At(e_x)$. At least, $\lfloor \log |At(e_x)| \rfloor$ binary questions are needed, and at most $\lceil \log |At(e_x)| \rceil$. That is why $\log |At(e_x)|$ is a reasonable measure for the total uncertainty in a choice situation represented by the finite choice system $At(e_x)$. This measure is nowadays known as Hartley measure, as he was the first to take "as our practical measure of information the logarithm of the number of possible symbol sequences" (Hartley, 1928). As a consequence of these yes-no-questions, the logarithm is taken to base 2, but any other base could be used, too. This would only lead to a shift of scale in the measurement of uncertainty and information. Information is measured by a value out of $\mathbb{R}_0^+ \cup \{+\infty\}$. The unit of the measurement is the bit.

Now we will look at how this total uncertainty, given by the set of all possible atoms on x and measured by $\log |At(e_x)|$, can be changed. Consider a piece of information $\phi \in \Phi$, $d(\phi) = x$ and $At(\phi) \subseteq At(e_x)$. This piece of information is seen as a scheme of choice (without probabilities), and the question is which atom will be chosen. Clearly, it must be one of the atoms of $At(\phi)$. Knowing ϕ reduces the uncertainty to $\log |At(\phi)| \leq \log |At(e_x)|$.

The information content of ϕ is measured by $i(\phi)$. It is defined to be the reduction of uncertainty obtained by ϕ when initially, nothing is known. Implicitly, ϕ, with $d(\phi) = x$, is measured relative to the domain or question x, which is however not mentioned in the measure $i(\phi)$. Furthermore, when we say that initially, nothing is known, we mean that only the vacuous information e_x is given.

Definition 7.15 (Information Content Relative to Initial Ignorance) *Assuming initial ignorance, the content of a piece of information $\phi \in \Phi$ with $x = d(\phi)$ in an atomic composed information algebra (Φ, D) is*

$$i(\phi) \quad = \quad \log |At(e_x)| - \log |At(\phi)| \quad = \quad -\log \tfrac{|At(\phi)|}{|At(e_x)|}.$$

One may consider

$$p(\phi) \quad = \quad \frac{|At(\phi)|}{|At(e_x)|} \tag{7.2}$$

to be the probability that an atom in $At(\phi)$ is selected out of the atoms of $At(e_x)$, in case that all atoms have the same chance to be selected. We will also call it the probability of ϕ. Putting Definition 7.15 and Equation 7.2 together results in

$$i(\phi) \quad = \quad -\log p(\phi), \tag{7.3}$$

which corresponds to what is known to be the entropy of an event with uniform probability distribution observed in a random experiment. However, we do not want to focus on probabilistic considerations for the moment. Note the inverse relation of information to probability. The information content of a piece of information ϕ will be smaller the more probable it is.

There are special cases, such as measuring how much information content there is in the vacuous information:

$$\begin{aligned} i(e_x) &= -\log \tfrac{|At(e_x)|}{|At(e_x)|} \\ &= -\log 1 \\ &= 0 \end{aligned}$$

Expectedly, the vacuous information carries no information and has measure 0. A further special case is the other "extreme" piece of information, the null information z_x. The null information does not contain any atoms, so $At(z_x) = \emptyset$ and thus $|At(z_x)| = 0$. Measuring its information content results in

$$\begin{aligned} i(z_x) &= -\log \tfrac{|At(z_x)|}{|At(e_x)|} \\ &= -\log \tfrac{0}{|At(e_x)|} \\ &= -\log 0 \\ &= +\infty \end{aligned}$$

Here we use the convention that $\log 0 = -\infty$. In fact, z_x has been introduced before for technical reasons, but it is not really a piece of information about a possible atom (to be selected), since it contains no atom at all. Instead of $+\infty$, one often decides that $i(z_x)$ is not defined.

Note that the partial order $\leq$ of Section 7.2 is maintained. For $\phi, \psi \in \Phi$ with $d(\phi) = d(\psi)$,

$$\phi \leq \psi \quad \text{implies} \quad i(\phi) \leq i(\psi).$$

In particular, we can state that for a piece of information $\phi \in \Phi_x$ in an atomic composed information algebra we have

$$i(e_x) \quad \leq \quad i(\phi) \quad \leq \quad i(z_x),$$

which gives a lower bound for $i(\phi)$:

$$0 \quad \leq \quad i(\phi) \quad \leq \quad +\infty. \tag{7.4}$$

Let us now move on to a more general case, where the content of a piece of information is measured relative to an arbitrary, but fixed question.

7.4.2 Information Content Relative to a Question

A piece of information $\phi \in \Phi$ always contains information about a question $x \neq d(\phi)$, even if this is only the vacuous information e_x. Therefore, ϕ's information content will vary from one question to another. ϕ has to be transported to the question one is interested in before measuring its content relative to this question.

Definition 7.16 (Information Content Relative to a Specified Question)

Assuming initial ignorance, the information content relative to question $x \in D$ of a piece of information $\phi \in \Phi$ in an atomic composed information algebra (Φ, D) is

$$i(\phi\,;\,x) \quad = \quad i(\phi^{\rightarrow x}).$$

Reconsider the measure introduced in Definition 7.15. Clearly, the following equality holds:

$$i(\phi\,;\,d(\phi)) \quad = \quad i(\phi),$$

as $i(\phi)$ measures the information content of ϕ relative to the question it naturally refers to (its domain $d(\phi)$) and $\phi^{\rightarrow d(\phi)}$ is obviously equal to ϕ in an information algebra.

Let us look at further properties of $i(\phi\,;\,x)$. Consider $\phi \in \Phi$ and the just described case of $x = d(\phi)$. If, additionally, ϕ is the vacuous information e_x, then $i(e_x\,;\,x) = 0$. Since we have seen in Equation 7.4 that 0 is the lower bound of the measure proposed in Definition 7.16, we furthermore point out that $i(\phi\,;\,x) = -\log \frac{|At(\phi^{\rightarrow x})|}{|At(e_x)|} \geq 0$ for all $\phi \in \Phi$ and $x \in D$, as the negative logarithm of a number in the interval $]0;1[$ will always result in a positive value. Further properties are collected in the following lemma, see (Kohlas, 2002; Schneuwly, 2007; Kohlas & Schneuwly, 2009) for their respective proofs:

Lemma 7.17 *Let (Φ, D) be an atomic composed information algebra with finite sets of atoms $At_x(\Phi), \forall x \in D$. Then, for all $x, y, z \in D$ and $\phi, \psi \in \Phi$, it holds that:*

1. $\phi \leq \psi$ *implies* $\phi \leq_x \psi$ *implies* $i(\phi\,;\,x) \leq i(\psi\,;\,x)$.
2. $i(\phi^{\to y}\,;\,x) \leq i(\phi\,;\,x)$.
3. $i(\phi\,;\,x), i(\psi\,;\,x) \leq i(\phi \otimes \psi\,;\,x)$.
4. $\phi \leq \psi$ *implies* $i(\phi^{\to y}\,;\,x) \leq i(\psi^{\to y}\,;\,x)$.
5. $i(\phi^{\to y} \otimes \psi^{\to y}\,;\,x) \leq i((\phi \otimes \psi)^{\to y}\,;\,x)$.
6. $y \leq z$ *implies* $i(\phi^{\to y}\,;\,x) \leq i(\phi^{\to z}\,;\,x)$.

As expected, the qualitative measure $\leq$ introduced in Section 7.2 is maintained by the quantitative measure i proposed here. The transport operation adds no information and, hence, does not enlarge the information content. Relative to a fixed question, the combination of two pieces of information is always at least as informative than either of the single pieces of information. The measure of information content relative to a specified question also respects the partial order between pieces of information when both of them are transported to the same, but arbitrary domain. If two pieces of information have been separately transported to the same domain, which is not necessarily the one the measure is relative to, their combination contains less information than the transport of their combination. Finally, the order in the lattice of domains is maintained by the quantitative measure.[10]

The next step of generalization is to drop the initial ignorance we have assumed so far. This is necessary as information comes piecewise, and so one has also to look at the content of a piece of information relative to information previously received.

7.4.3 Information Content Relative to Prior Information

According to the relativity of information introduced in Section 6.1, a piece of information is not only relative to a question (this case has been looked at in Sections 7.4.1 and 7.4.2), but also to prior information. Its measure is based on the following idea: Suppose ϕ is already known, and now we come to know also ψ. Thus the information we dispose of is changed from ϕ to $\phi \otimes \psi$. It is therefore natural to measure the information of ψ by the reduction of uncertainty which results from knowing $\phi \otimes \psi$, when, initially, only ϕ was known. The same procedure as seen above will be applied. The initially given information is ϕ, thus the uncertainty is $\log |At(\phi)|$. When a new piece of information ψ arrives, the total information $\phi \otimes \psi$ is known and

[10] It is time again for a brief excursion into the field of domain-free information algebras. In such a setting, $[\phi]$ is the class of equivalent information elements (see Section 5.1). All elements of this class have the same information content $i(\phi\,;\,x)$ with respect to the domain $x \in D$. It is thus also possible to assign a measure of information content to the elements of domain-free information algebras: $i([\phi]\,;\,x) = i(\phi\,;\,x)$. Similar results to those of Lemma 7.17 are obtained when the partial order in the domain-free algebra is considered.

the new uncertainty is therefore $\log |At(\phi \otimes \psi)|$. Similar to what has been introduced in Definition 7.15 (where total ignorance was changed to the knowledge of ϕ), we will first propose a measure for two pieces of information which got known one after another. Both of them are related to the same domain $x \in D$:

Definition 7.18 (Information Content Relative to Prior Information)

Let $\phi, \psi \in \Phi$ be two pieces of information with domain $x \in D$ in an atomic composed information algebra. The relative information content *of ψ, given that ϕ is already known, is measured by*

$$i(\psi|\phi) \quad = \quad \log |At(\phi)| - \log |At(\phi \otimes \psi)| \quad = \quad -\log \tfrac{|At(\phi\otimes\psi)|}{|At(\phi)|}.$$

Since ϕ and ψ are pieces of information bearing on the same domain x, it holds that $At(\phi \otimes \psi) = At(\phi) \cap At(\psi)$. Therefore,

$$i(\psi|\phi) = -\log p(\psi|\phi) \tag{7.5}$$

holds. Here, $p(\psi|\phi) = \frac{|At(\phi\otimes\psi)|}{|At(\phi)|}$ is the probability that an atom in $At(\phi \otimes \psi)$ is selected out of the atoms of $At(\phi)$, which are already known. The proposed measure is thus the negative logarithm of the conditional probability of ψ given ϕ, under the assumption of a uniform probability distribution over the atoms in $At(e_x)$, see also Section 7.6 for more details.

We will now generalize the measure relative to prior information by extending the definition to arbitrary domains, as we did before when extending Definition 7.15 to Definition 7.16:

Definition 7.19 (Relative Information Content Regarding a Question)

Let $\phi, \psi \in \Phi$ be two pieces of information in an atomic composed information algebra (Φ, D) and $x \in D$ an arbitrary, but fixed domain. The information content of ψ relative to the prior information ϕ and the question x is measured by

$$i(\psi|\phi\,;\,x) \quad = \quad \log |At(\phi^{\to x})| - \log |At((\phi \otimes \psi)^{\to x})| \quad = \quad -\log \tfrac{|At((\phi\otimes\psi)^{\to x})|}{|At(\phi^{\to x})|}.$$

Note that in general $i(\psi|\phi\,;\,x) \neq i(\psi^{\to x}|\phi^{\to x})$, by Property 3 of Lemma 7.3. From the same lemma it is known that $\phi \leq \phi \otimes \psi$ (Property 1), implying that $\phi^{\to x} \leq (\phi \otimes \psi)^{\to x}$ (Property 4). So $|At(\phi^{\to x})| \geq |At((\phi \otimes \psi)^{\to x})|$, $\frac{|At((\phi\otimes\psi)^{\to x})|}{|At(\phi^{\to x})|} \in]0; 1]$ and thus $i(\psi|\phi\,;\,x) \geq 0$. Clearly, $i(\phi|\phi\,;\,x) = 0$. $i(\psi|\phi)$ may be numerically less (as in the murderer example) or greater than $i(\phi)$, but it will always be non-negative. If ϕ and ψ are contradictory (see Equation 4.5), their combination results in the null information which does not contain any atoms. With the same reasoning as for $i(z_x)$ in Section 7.4.1, we obtain $i(\psi|\phi\,;\,x) = +\infty$ if $(\phi \otimes \psi)^{\to x} = z_x$, for $x \in D$. This is only possible if $\phi \otimes \psi = z_{d(\phi)\vee d(\psi)}$. In that case, the two pieces of information are contradictory, as stated by Equation 4.5. Finally, note that $i(\psi|e_x\,;\,x) = i(\psi\,;\,x)$. Based on (Kohlas, 2002, Lemma 18) some further simple results about the measure of relative information $i(\psi|\phi\,;\,x)$ are given below without proof:

Lemma 7.20 *Let (Φ, D) be an atomic composed information algebra with finite sets of atoms $At_x(\Phi), \forall x \in D$. Then, for all $x \in D$ and $\phi, \psi_1, \psi_2 \in \Phi$, it holds that:*

- *If $\phi \leq \psi$, then $i(\psi|\phi\,;\,x) \leq i(\psi\,;\,x)$.*
- *$\psi_1 \leq \psi_2$ implies $i(\psi_1|\phi\,;\,x) \leq i(\psi_2|\phi\,;\,x)$.*
- *$\psi \leq \phi$ implies $i(\psi|\phi\,;\,x) = 0$.*

If some prior information is less informative (in the sense of the qualitative measure $\leq$ introduced in Section 7.2) than some posterior information, then the measure of the posterior information, given the prior information, respects this order, also relative to a specified question. If some piece of information is less informative than another one, this will also be the case when their information content is measured relative to some prior information. Finally, if some piece of information is less informative than another, its knowledge will not change the uncertainty, assuming that the more informative one is already given.

Combined Information

A further important concept is the combined information $\phi \otimes \psi$, relative to some question x. Applying Definitions 7.15 and 7.16, its measure is given by the change of uncertainty caused by learning $(\phi \otimes \psi)^{\to x}$, when, initially, nothing about the domain x is known:

$$i(\phi \otimes \psi\,;\,x) \quad = \quad \log|At(e_x)| - \log|At((\phi \otimes \psi)^{\to x})|. \tag{7.6}$$

It can also be obtained by taking the sum of the information content carried by $\phi^{\to x}$ and the relative information content of ψ regarding x, given that ϕ is already known. This is stated in the so-called Chaining Theorem known from (Kohlas, 2002):

Theorem 7.21 (Chaining Theorem) *Let (Φ, D) be an atomic composed information algebra with finite sets of atoms $At_x(\Phi), \forall x \in D$. Then, for all $x \in D$ and $\phi, \psi \in \Phi$, it holds that*

$$i(\phi \otimes \psi\,;\,x) \quad = \quad i(\phi\,;\,x) + i(\psi|\phi\,;\,x) \quad = \quad i(\psi\,;\,x) + i(\phi|\psi\,;\,x).$$

Proof. Only the first equation is proven; the second follows by symmetry.

$$\begin{aligned} i(\phi \otimes \psi\,;\,x) &= \log|At(e_x)| - \log|At((\phi \otimes \psi)^{\to x})| \\ &= (\log|At(e_x)| - \log|At(\phi^{\to x})|) + \\ &\quad\ (\log|At(\phi^{\to x})| - \log|At((\phi \otimes \psi)^{\to x})|) \\ &= i(\phi\,;\,x) + i(\psi|\phi\,;\,x) \end{aligned}$$

□

The chaining theorem can be generalized to the combination of more than two pieces of information:

$$i(\phi_1 \otimes \cdots \otimes \phi_n \,;\, x) \;=\; i(\phi_1 \,;\, x) + i(\phi_2|\phi_1 \,;\, x) + \cdots + i(\phi_n|\phi_1 \otimes \cdots \otimes \phi_{n-1} \,;\, x).$$

Mutual Information

The measure of the mutual information between two pieces of information ϕ and ψ is another interesting aspect. Strictly speaking, mutual information is not the measure of a piece of information, as for example in the case of the combined information $\phi \otimes \psi$, as seen above. It rather measures the relation between two pieces of information, namely the amount of information we expect to obtain regarding one piece of information from observing the other one. The mutual information is denoted by $i(\phi||\psi \,;\, x)$ and defined as follows:

$$i(\phi||\psi \,;\, x) \;:=\; i(\phi \,;\, x) + i(\psi \,;\, x) - i(\phi \otimes \psi \,;\, x). \tag{7.7}$$

The mutual information between ϕ and ψ equals the mutual information between ψ and ϕ, always relative to question $x \in D$:

$$i(\phi||\psi \,;\, x) \;=\; i(\psi||\phi \,;\, x). \tag{7.8}$$

It is also sometimes considered as a measure of dependence between two pieces of information. Two pieces of information ϕ and ψ are said to be independent relative to the question x if $i(\phi||\psi \,;\, x) = 0$. For such pieces of information, the following additivity property holds:

$$i(\phi \otimes \psi \,;\, x) \;=\; i(\phi \,;\, x) + i(\psi \,;\, x). \tag{7.9}$$

When ϕ and ψ are independent relative to the question x, disposing of one piece of information does not tell anything about the other one. In such a situation, it holds that $i(\phi|\psi \,;\, x) = i(\phi \,;\, x)$, as well as $i(\psi|\phi \,;\, x) = i(\psi \,;\, x)$, according to the chaining theorem and the above additivity property (Equation 7.9). Actually, there exist some further equivalent ways of measuring mutual information. They are summarized in the following theorem:

Theorem 7.22 *Let (Φ, D) be an atomic composed information algebra with finite sets of atoms $At_x(\Phi), \forall x \in D$. Then, for all $x \in D$ and $\phi, \psi \in \Phi$, the following ways of measuring the* mutual information *ϕ and ψ relative to a question x are equivalent:*

1. $i(\phi||\psi \,;\, x) = i(\phi \otimes \psi \,;\, x) - i(\phi|\psi \,;\, x) - i(\psi|\phi \,;\, x)$,
2. $i(\phi||\psi \,;\, x) = i(\phi \,;\, x) - i(\phi|\psi \,;\, x)$,
3. $i(\phi||\psi \,;\, x) = i(\psi \,;\, x) - i(\psi|\phi \,;\, x)$,
4. $i(\phi||\psi \,;\, x) = i(\phi \,;\, x) + i(\psi \,;\, x) - i(\phi \otimes \psi \,;\, x)$.

Proof.

$$
\begin{array}{lrl}
1. \Leftrightarrow 2. & & i(\phi \otimes \psi \,;\, x) - i(\phi|\psi \,;\, x) - i(\psi|\phi \,;\, x) \\
& = & i(\phi \,;\, x) + i(\psi|\phi \,;\, x) - i(\phi|\psi \,;\, x) - i(\psi|\phi \,;\, x) \\
& = & i(\phi \,;\, x) - i(\phi|\psi \,;\, x).
\end{array}
$$

$$
\begin{array}{lrl}
1. \Leftrightarrow 3. & & i(\phi \otimes \psi \,;\, x) - i(\phi|\psi \,;\, x) - i(\psi|\phi \,;\, x) \\
& = & i(\psi \,;\, x) + i(\phi|\psi \,;\, x) - i(\phi|\psi \,;\, x) - i(\psi|\phi \,;\, x) \\
& = & i(\psi \,;\, x) - i(\psi|\phi \,;\, x).
\end{array}
$$

$$
\begin{array}{lrl}
1. \Leftrightarrow 4. & & i(\phi \otimes \psi \,;\, x) - i(\phi|\psi \,;\, x) - i(\psi|\phi \,;\, x) \\
& = & i(\phi \otimes \psi \,;\, x) - i(\phi \otimes \psi \,;\, x) + i(\psi \,;\, x) - i(\phi \otimes \psi \,;\, x) + i(\phi \,;\, x) \\
& = & i(\phi \,;\, x) + i(\psi \,;\, x) - i(\phi \otimes \psi \,;\, x).
\end{array}
$$

□

Example

We conclude with an example illustrating the different measures seen above.

Example 7.4.1 Let (Φ, D) be an atomic composed information algebra. For the sake of simplicity, we will only consider pieces of information on a fixed domain $x \in D$. We are therefore dealing with pieces of information $\phi \in \Phi$ which are composed of the atoms of $At_x(\Phi)$. In the present case, the set of all atoms is $At_x(\Phi) = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}\}$ and thus total ignorance is given by $At(e_x) = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}\}$, too. Figure 7.1 provides an illustration of the measure of information by reduction of uncertainty.

In this figure, all possible atomic pieces of information $\phi \in \Phi$ are given. They are subsets of $At_x(\Phi) = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}\}$. The uncertainty is computed by $\log |At(\phi)|$. When nothing is known, we only dispose of the vacuous information $e_x = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}\}$, and the uncertainty is $\log 4 = 2$ bit. If however $\phi = \{\mathbf{x_2}, \mathbf{x_3}\}$, the uncertainty in such a situation is $\log 2 = 1$ bit.

Information content is measured by change of uncertainty. According to Definition 7.15, the content of the piece of information $\phi = \{\mathbf{x_2}, \mathbf{x_3}\}$ is

$$i(\phi) = \log 4 - \log 2 = 1.$$

Consider also $\psi_1 = \{\mathbf{x_2}, \mathbf{x_4}\}$, $\psi_2 = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}\}$, $\psi_3 = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_4}\}$ and $\psi_4 = \{\mathbf{x_2}\}$. Their respective information content is

$$
\begin{array}{lclclcl}
i(\psi_1) & = & \log |At(e_x)| - \log |\{\mathbf{x_2}, \mathbf{x_4}\}| & = & \log 4 - \log 2 & = & 1 \text{ bit}, \\
i(\psi_2) & = & \log |At(e_x)| - \log |\{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}\}| & = & \log 4 - \log 3 & = & \log \frac{4}{3} \text{ bit}, \\
i(\psi_3) & = & \log |At(e_x)| - \log |\{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_4}\}| & = & \log 4 - \log 3 & = & \log \frac{4}{3} \text{ bit}, \\
i(\psi_4) & = & \log |At(e_x)| - \log |\{\mathbf{x_2}\}| & = & \log 4 - \log 1 & = & 2 \text{ bit}.
\end{array}
$$

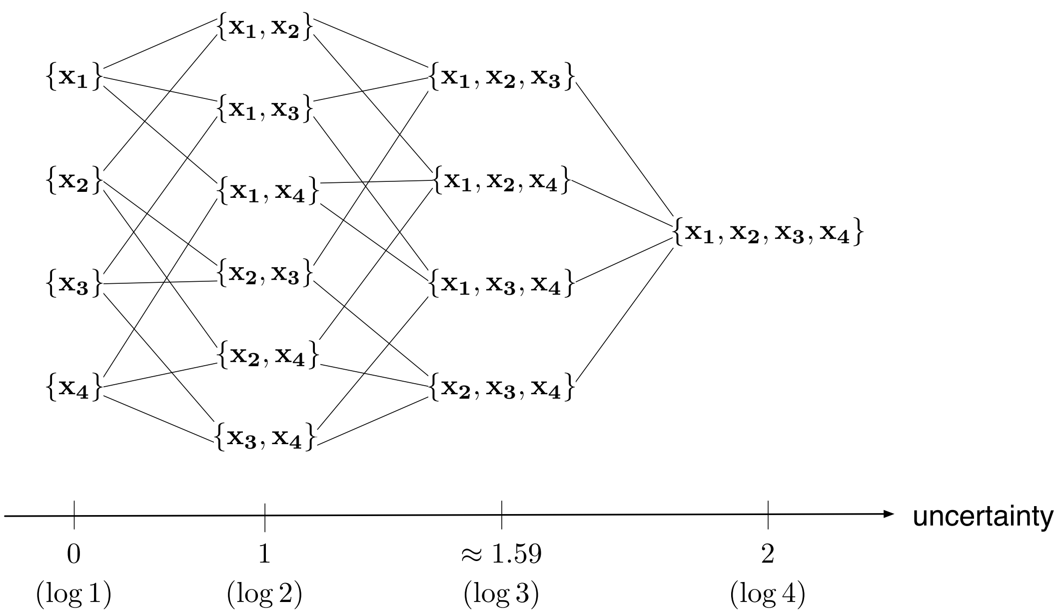


Figure 7.1: Information measure as reduction of uncertainty

When we look at prior information, the combination of two pieces of information comes into play. In Figure 7.1, the links between the pieces of information allow to determine their combination. Given $\phi = \{\mathbf{x_2}, \mathbf{x_3}\}$ (with $i(\phi) = 1$, as seen above), the relative information content of ψ_i, $i \in \{1,2,3,4\}$, is

$$
\begin{array}{rclclcl}
i(\psi_1|\phi) & = & \log|At(\phi)| - \log|\{\mathbf{x_2}\}| & = & \log 2 - \log 1 & = & 1 \text{ bit}, \\
i(\psi_2|\phi) & = & \log|At(\phi)| - \log|\{\mathbf{x_2}, \mathbf{x_3}\}| & = & \log 2 - \log 2 & = & 0 \text{ bit}, \\
i(\psi_3|\phi) & = & \log|At(\phi)| - \log|\{\mathbf{x_2}\}| & = & \log 2 - \log 1 & = & 1 \text{ bit}, \\
i(\psi_4|\phi) & = & \log|At(\phi)| - \log|\{\mathbf{x_2}\}| & = & \log 2 - \log 1 & = & 1 \text{ bit},
\end{array}
$$

The relative information content of ϕ, when ψ_i is already known ($i \in \{1,2,3,4\}$), is

$$
\begin{array}{rclclcl}
i(\phi|\psi_1) & = & \log|At(\psi_1)| - \log|\{\mathbf{x_2}\}| & = & \log 2 - \log 1 & = & 1 \text{ bit}, \\
i(\phi|\psi_2) & = & \log|At(\psi_2)| - \log|\{\mathbf{x_2}, \mathbf{x_3}\}| & = & \log 3 - \log 2 & = & \log \frac{3}{2} \text{ bit}, \\
i(\phi|\psi_3) & = & \log|At(\psi_3)| - \log|\{\mathbf{x_2}\}| & = & \log 3 - \log 1 & = & \log 3 \text{ bit}, \\
i(\phi|\psi_4) & = & \log|At(\psi_4)| - \log|\{\mathbf{x_2}\}| & = & \log 1 - \log 1 & = & 0 \text{ bit}.
\end{array}
$$

The cases where the combination of two pieces of information is the null information (the contradiction z_x) have not been considered so far, as it was stated above that z_x is not really information about a possible atom.

When we look at the combined information $\phi \otimes \psi$, its content may be measured according to Equation 7.6:

$$
\begin{array}{rclclcl}
i(\psi_1 \otimes \phi) & = & \log|At(e_x)| - \log|\{\mathbf{x_2}\}| & = & \log 4 - \log 1 & = & 2 \text{ bit}. \\
i(\psi_2 \otimes \phi) & = & \log|At(e_x)| - \log|\{\mathbf{x_2}, \mathbf{x_3}\}| & = & \log 4 - \log 2 & = & 1 \text{ bit}, \\
i(\psi_3 \otimes \phi) & = & \log|At(e_x)| - \log|\{\mathbf{x_2}\}| & = & \log 4 - \log 1 & = & 2 \text{ bit}. \\
i(\psi_4 \otimes \phi) & = & \log|At(e_x)| - \log|\{\mathbf{x_2}\}| & = & \log 4 - \log 1 & = & 2 \text{ bit}.
\end{array}
$$

It can easily be seen that the application of the Chaining Theorem (Theorem 7.21) provides the same results:

$$\begin{array}{lllllll} i(\psi_1 \otimes \phi) & = & i(\phi) + i(\psi_1|\phi) & = & 1+1 & = & 2 \text{ bit.} \\ i(\psi_2 \otimes \phi) & = & i(\psi_2) + i(\phi|\psi_2) & = & \log \frac{4}{3} - \log \frac{2}{3} & = & 1 \text{ bit,} \\ i(\psi_3 \otimes \phi) & = & i(\phi) + i(\psi_3|\phi) & = & 1+1 & = & 2 \text{ bit.} \\ i(\psi_4 \otimes \phi) & = & i(\psi_4) + i(\phi|\psi_4) & = & 2+0 & = & 2 \text{ bit.} \end{array}$$

Finally, the mutual information, measured by one of the ways stated in Theorem 7.22, is given by

$$\begin{array}{lllllll} i(\phi||\psi_1) & = & i(\phi \otimes \psi_1) - i(\phi|\psi_1) - i(\psi_1|\phi) & = & 2-1-1 & = & 0 \text{ bit,} \\ i(\phi||\psi_2) & = & i(\phi) - i(\phi|\psi_2) & = & \log 2 - \log \frac{3}{2} & = & \log \frac{4}{3} \text{ bit,} \\ i(\phi||\psi_3) & = & i(\psi_3) - i(\psi_3|\phi) & = & \log \frac{4}{3} - \log 2 & = & \log \frac{2}{3} \text{ bit,} \\ i(\phi||\psi_4) & = & i(\phi) + i(\psi_4) - i(\phi \otimes \psi_4) & = & 1+2-2 & = & 1 \text{ bit.} \end{array}$$

$\odot$

7.5 Quantitative Measure in the Dual Algebra

In the previous section, we have seen that disposing of a labeled atomic composed information algebra (Φ, D) allows us to define a quantitative measure. An important special case of the foregoing section comes up when the considered atomic composed information algebra is Boolean in addition; then, a quantitative information measure in the dual algebra can be established. In the case of a atomic composed Boolean labeled information algebra (Φ, D), there is also a dual notion of the concept of an atom. Let $At^d(\Phi)$ denote the set of all dual atoms and $At^d_x(\Phi)$ the set of all dual atoms on x. If $\alpha \in At^d_x(\Phi)$ is a dual atom on x, then α^c is an atom on x. See (Schneuwly, 2007; Kohlas & Schneuwly, 2009) for omitted proofs and further details.

In the following, we will consider the atomic composed Boolean labeled information algebra (Φ, D), which is named dual when it is considered together with the following operators: labeling d, dual combination $\otimes^d$ from Definition 7.7, dual marginalization $\downarrow^d$ from Definition 7.8, dual meet $\wedge^d_x$ from Definition 7.9 and complement ^{c} from Definition 4.14. The proceeding is the same as in Section 7.4: We will step by step generalize the proposed dual measure. Furthermore, we are still assuming that for all domains $x \in D$ the set of all atoms $At_x(\Phi)$, but also that of all dual atoms $At^d_x(\Phi)$, is finite.

When initially, nothing is known, information content is measured by reduction of uncertainty. Initial ignorance is expressed by the vacuous information. According to Theorem 7.10, the vacuous piece of information of domain $x \in D$ in the dual algebra is z_x, the neutral element of Φ_x.

Definition 7.23 (Dual Information Content)

Assuming initial ignorance, the dual information measure *of* $\phi \in \Phi$*, with* $x = d(\phi)$ *arbitrary in* D*, in an atomic composed Boolean information algebra is*

$$i^d(\phi) \quad = \quad \log|At^d(z_x)| - \log|At^d(\phi)| \quad = \quad -\log\frac{|At^d(\phi)|}{|At^d(z_x)|}.$$

The equation $|At^d(\phi)| = |At(\phi^c)|$ *(see Section 4.7), allows another formulation:*

$$i^d(\phi) \quad = \quad \log|At(e_x)| - \log|At(\phi^c)| \quad = \quad i(\phi^c). \tag{7.10}$$

In particular, $i(e_x) = i^d(z_x) = 0$.

From Section 7.4.2, it is known how to generalize this measure, namely by considering it relative to a fixed question $x \in D$. Equation 7.10 gives rise to the following theorem which tells how to measure the dual information content relative to a question:

Theorem 7.24 *Assuming initial ignorance, the measure of the* dual information content relative to a question $x \in D$ *of a piece of information* ϕ *in an atomic composed Boolean information algebra is*

$$i^d(\phi\,;\,x) \quad = \quad i(\phi^c\,;\,x).$$

Proof.

$$\begin{aligned}
i(\phi^c\,;\,x) &= i\left((\phi^c)^{\to x}\right) \\
&= i\left((((\phi^c)^{\to x})^c)^c\right) \\
&\overset{\text{Def. 7.13}}{=} i\left((\phi^{\to^d x})^c\right) \\
&\overset{\text{Eq. 7.10}}{=} i^d(\phi^{\to^d x}) \\
&= i^d(\phi\,;\,x).
\end{aligned}$$

□

The above theorem shows that the measure i^d of a piece of information $\phi \in \Phi$ in the dual algebra can be traced back to the "usual" measure i of Section 7.4. Hence, using Definition 7.19, for $\phi, \psi \in \Phi$ and $x \in D$, the dual information content of ψ relative to the prior information ϕ and the question x is measured by:

$$i^d(\psi|\phi\,;\,x) \quad = \quad i(\psi^c|\phi^c\,;\,x). \tag{7.11}$$

In the same way, we can state the dual chaining theorem, based on Theorem 7.21:

$$i^d(\phi \otimes^d \psi\,;\,x) \quad = \quad i^d(\phi\,;\,x) + i^d(\psi|\phi\,;\,x) \quad = \quad i^d(\psi\,;\,x) + i^d(\phi|\psi\,;\,x). \tag{7.12}$$

This concludes the proposed measures for dual atomic composed Boolean labeled information algebras.

7.6 Taking Probabilities Into Account

Until now, we have not taken any probability distribution into account, or a uniform probability distribution was more or less tacitly assumed. We will now formally introduce probabilistic choice systems, which consist of a choice system and a probability distribution. When additional information arises, the probabilistic choice system may change. This leads to a conditional probability distribution of the prior distribution, given the information which came up. A probability distribution may also be marginalized to some field of interest or question, resulting in the marginal probability distribution. These notions are needed to propose a measure of information based on entropy which takes probabilities into account. We are still assuming to deal with an atomic composed information algebra (Φ, D), see Definition 4.18. So for all $\phi \in \Phi$ it holds that $\phi = \bigwedge At(\phi)$, containing all atoms α relative to ϕ. We will often use ϕ instead of $At(\phi)$, so consider the following conventions:

$$\begin{aligned} \alpha \in \phi &:= \alpha \in At(\phi) \\ \phi \subseteq At_x(\Phi) &:= At(\phi) \subseteq At_x(\Phi) \\ |\phi| &:= |At(\phi)| \end{aligned}$$

Recall that $At(e_x) = At_x(\Phi)$. In particular in Section 7.6.2, we will write e_x instead of $At_x(\Phi)$.

In Section 7.6.1, we will introduce some notions of probability theory, which are needed in the following. This allows to propose in Section 7.6.2 a concrete measure of information which takes a probability distribution over the piece of information into account.

7.6.1 Probabilistic Choice Systems and Their Uncertainty

A choice situation where each possibility has a known probability is formally described by a probabilistic choice system.

Definition 7.25 (Probabilistic Choice System)

Consider a choice system $S = \{e_1, e_2, \ldots, e_n\}$ and a set $P = \{p_1, p_2, \ldots, p_n\}$ of probabilities on S satisfying the following two conditions (7.13) and (7.14):

$$0 \leq p_i \leq 1, \tag{7.13}$$

$$\sum_{i=1}^{n} p_i = 1. \tag{7.14}$$

Clearly, P is a probability distribution*. The pair (S, P) is then termed* probabilistic choice system*.*

A piece of information $\phi \in \Phi$ is a probabilistic choice system: The information algebra (Φ, D) being atomic composed, each piece of information $\phi \in \Phi$ with $d(\phi) =$

x can be given as a subset of the set of all atoms on x: $\phi = \bigwedge At(\phi) \subseteq At_x(\Phi)$. The set $\{\alpha_1, \alpha_2, \ldots, \alpha_{|\phi|}\}$ of atoms α_i contained in $At(\phi)$ corresponds to the choice system S of Definition 7.25. $P = \{p_1, p_2, \ldots, p_{|\phi|}\}$ is the associated probability distribution over the $\alpha_i \in \phi$ such that $p(\alpha_i) = p_i$, for $i \in \{1, 2, \ldots, |\phi|\}$.

The choice system considered in (Hartley, 1928), which is underlying Section 7.4, has always a uniform probability distribution.

Entropy

As already known from Section 7.4, entropy measures the expected number of questions needed to find out which initially selected, but unknown atom α_i of the scheme of choice (ϕ, P) is chosen. Entropy is thus used for measuring the amount of uncertainty regarding a piece of information ϕ.

Definition 7.26 (Entropy)

Let (Φ, D) be an atomic composed information algebra and let $\phi \in \Phi$ be a piece of information. By means of the associated probabilistic choice system (ϕ, P), the amount of uncertainty *regarding ϕ is defined to be the* entropy

$$H(\phi, P) \quad = \quad -\sum_{p_i \in P} p_i \log p_i.$$

We write $H(\phi, P)$, instead of simply $H(\phi)$ or $H(P)$, in order to stress that the piece of information ϕ and its associated probability distribution form a whole and should not be considered separately. Usually, the logarithm is taken to the base 2 and the measuring unit is the bit. In case that some probabilities p_i equal 0, we adopt the convention that $0 \log 0 = 0$, which is justified by $\lim_{x \to 0} x \log x = 0$.

In (Shannon, 1948), this measure has already been called "H" and the author points out that

> "[t]he form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics [...]."

The probability of the piece of information $\phi \subseteq At_x(\Phi)$ is obtained by summing up the probabilities $p_i = p(\alpha_i)$ of the probability distribution P associated with $At_x(\Phi)$, for all $\alpha_i \in \phi$:

$$p(\phi) \quad = \quad \sum_{p_i \in P} p_i. \tag{7.15}$$

If two pieces of information ϕ and ψ are considered simultaneously, they have a common probability distribution $P_{\phi,\psi}$. Its probabilities $p_{\phi,\psi}$ are defined for a pair of atoms α^ϕ, α^ψ; the superscript indicates the choice system the atom originally stems from.

Two pieces of information are called stochastically independent if their common probability $p_{\phi,\psi}$ is obtained by multiplying their respective probabilities:

Definition 7.27 (Independence)

Let ϕ and ψ be two pieces of information in Φ. They are called (stochastically) independent *if, for all atoms $\alpha^\phi \in \phi$ and $\alpha^\psi \in \psi$,*

$$p_{\phi,\psi}(\alpha^\phi, \alpha^\psi) \quad = \quad p(\alpha^\phi) \cdot p(\alpha^\psi).$$

Conditional Entropy

When initially, only the a probability distribution P over the elements of $At_x(\Phi)$, for an arbitrary $x \in D$, is given and if a piece of information ϕ with $d(\phi) = x$ is learned afterwards, P is a prior probability distribution. We thus dispose of a probabilistic choice system $(At_x(\Phi), P)$, with $At_x(\Phi) = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $P = \{p_1, p_2, \ldots, p_n\}$, where $p_i = p(\alpha_i)$, for $i = 1, \ldots, n$. When a piece of information $\phi \subseteq At_x(\Phi)$ arises, then the question is how this affects the uncertainty in $(At_x(\Phi), P)$. Actually, ϕ induces a new situation of uncertainty described by another probabilistic choice system $(\phi, P|\phi)$. $P|\phi$ is the conditional probability distribution and consists of the conditional probabilities

$$p|\phi(\alpha_i) \quad = \quad \frac{p(\alpha_i)}{p(\phi)}, \tag{7.16}$$

for all $\alpha_i \in \phi$, under the assumption that $p(\phi) \neq 0$. The amount of uncertainty in this new probabilistic choice system $(\phi, P|\phi)$ is measured by the conditional entropy:

Definition 7.28 (Conditional Entropy)

Let (Φ, D) be an atomic composed information algebra, x an arbitrary domain in D and $\phi \subseteq At_x(\Phi)$ a piece of information. Then, knowing ϕ changes the probabilistic choice system from $(At_x(\Phi), P)$ to $(\phi, P|\phi)$. The conditional probabilities in $P|\phi$ are given according to Equation 7.16. The amount of uncertainty *in this new situation (after ϕ came up) is defined by the* conditional entropy*:*

$$H(At_x(\Phi), P|\phi) \quad = \quad -\sum_{\alpha_i \in \phi} p|\phi(\alpha_i) \log p|\phi(\alpha_i).$$

Marginal Probability Distribution

Until now, we have always fixed a domain $x \in D$ on which the probability distributions and the pieces of information ϕ were considered. So the atoms α to which a probability is assigned are elements of the frame $\mathfrak{D}_x$. It may however be the case that our field of interest differs from x. In the special case that the question y we are interested in is coarser than x $(y \leq x)$, the pieces of information have to be marginalized to the domain of interest $(\phi^{\downarrow y})$ and a new probability distribution P_y has to be computed, providing probabilities for elements of the frame $\mathfrak{D}_y$. P_y is called marginal probability distribution.

Before going on, some notation has to be introduced. In order to show to which frame an atom α_i belongs, the domain is superscript, so α_i^x is an atom of $\mathfrak{D}_x$. A probability distribution over atoms $\alpha_i^x \in At_x(\Phi)$ will be denoted by P_x, where $p_x(\alpha_i^x) = p_i \in P_x$.

The marginalization of a piece of information $\phi \subseteq At_x(\Phi)$ to a domain $y \leq x$ results in the piece of information $\phi^{\downarrow y} \in \Phi$, where $\phi^{\downarrow y} = \{\alpha_1^y, \alpha_2^y, \ldots, \alpha_n^y\}$ with $n = |\phi^{\downarrow y}|$ and the α_i^y are atoms of $At_y(\Phi)$. With each such atom $\alpha_i^y \in \phi^{\downarrow y}$, a probability $p_y(\alpha_i^y) = p_i$ of the marginal probability distribution P_y is associated, determined in the following way:

$$p_y(\alpha_i^y) \quad = \sum_{(\alpha_j^x)^{\downarrow y} = \alpha_i^y} p_x(\alpha_i^x). \tag{7.17}$$

Each atom $\alpha_j^x \in At_x(\Phi)$ is marginalized to y. If this results in the atom α_i^y we are interested in, the probability $p_x(\alpha_j^x)$, which is known, is summed up, leading to the probability $p_y(\alpha_i^y)$.

The general case, i. e. when the domain of interest $y \not\leq x$, necessitates the introduction of the concept of a hint. However, we will refrain from going into details. Let us nevertheless state the problem arising when such a domain y is considered. The piece of information cannot be marginalized any more, but it has to be transported to y, which implies vacuously extending the atoms of ϕ with the values in the frame $\mathfrak{D}_{y \setminus x}$. In our case, this is not a problem as we presume finite frames, so the piece of information can be vacuously extended to $x \vee y$ and projected afterwards. However, the probability distribution over the atoms of $\phi^{\rightarrow y}$ inconveniences us. Which probability shall we attribute? Due to the extension with the values in $\mathfrak{D}_{y \setminus x}$, additional atoms might be created, or less atoms might constitute $\phi^{\rightarrow y}$. A probability distribution, even the uniform probability distribution, also represents information. As in most of the cases we have no reason for any additional assumptions, the question of the probability distribution is a very tricky one. What we need is an instrument to represent empty information, not only empty pieces of information of Φ, but also empty information in the form of a probability distribution. Hints offer the possibility to represent empty information: probability distributions are only regarded as a special case. See (Kohlas & Monney, 1995) for the theory of hints.

Now we dispose of a sufficient "probabilistic toolbox" for establishing a quantitative measure of information taking probabilities into account.

7.6.2 Measure of Information Content

We have seen above that, for any $x \in D$, a prior probability distribution P over the elements of $At_x(\Phi)$ results in an initial uncertainty of $H(e_x, P)$. Learning a piece of information ϕ with domain x changes the uncertainty from $H(e_x, P)$ to $H(e_x, P|\phi)$. This change of uncertainty gives us a measure of the information content carried by ϕ.

Definition 7.29 (Information Content of ϕ)

Let (Φ, D) be an atomic composed information algebra, x a domain in D, (e_x, P) the initial choice system and $\phi \subseteq At_x(\Phi)$ a piece of information of Φ leading to the choice system $(\phi, P|\phi)$. The information content carried by ϕ is then measured by

$$i(\phi) \quad = \quad H(e_x, P) - H(e_x, P|\phi).$$

The above definition may be easily generalized to arbitrary initial choice systems (ψ, P), with the restriction that $d(\psi) = d(\phi)$. This leads to a measure $i(\phi|\psi)$.

Information Content Relative to a Question

As already known from the previous sections, information is always measured relative to a given question. Until now, we have always assumed an arbitrary, but fixed domain $x \in D$ determining the initial choice system (e_x, P), as well as a piece of information ϕ relative to the same domain x. For measuring the information content of ϕ with $d(\phi) = x$ relative to some other question $y \leq x$[11], both, the initial choice system (e_x, P) and the piece of information ϕ, have to be considered relative to y. As to the choice system (e_x, P), marginalizing e_x results by the stability axiom in e_y; P_y is the marginal probability distribution, obtained according to Equation 7.17 from the prior probability distribution P. The marginalized piece of information $\phi^{\downarrow y}$ changes the uncertainty from $H(e_y, P_y)$ to $H(e_y, P_y|\phi^{\downarrow y})$, as justified below. This leads to a measure of the information content carried by ϕ relative to a question $y \leq d(\phi)$.

Definition 7.30 (Information Content of ϕ Relative to Question y)

Let (Φ, D) be an atomic composed information algebra, x a domain in D, P a probability distribution over $At_x(\Phi)$ and ϕ a piece of information with $d(\phi) = x$. The information content carried by ϕ relative to $y \leq x$ is then measured by

$$i(\phi\,;\,y) \quad = \quad H(e_y, P_y) - H(e_y, P_y|\phi^{\downarrow y}).$$

The marginal probability distributions associated with e_y and $\phi^{\downarrow y}$ are determined using Equation 7.17.

Measuring Information Content

Suppose that we are given the finest possible probabilistic choice system which provides a prior probability distribution P for all atoms on the top domain. This probabilistic choice system is denoted by $(e_\top, P)$ and expresses the most fine-grained information of what is overall known. Now a piece of information $\phi \in \Phi$ comes up with some new specific information. We are interested in $i(\phi\,;\,x)$, the amount of information content that ϕ carries relative some question $x \in D$. Figure 7.2 depicts what has to be done in order to compute $i(\phi\,;\,x)$ in this situation.

[11] In Section 7.6.1 we explain why we restrict ourselves to questions $y \leq x$.

$$\begin{tikzcd}
(e_\top, P) \arrow[r, "\text{marginalize to } x"] \arrow[d, "\text{condition relative to } \phi^{\uparrow\top}"] & (e_x, P_x) \\
(e_\top, \underbrace{P|\phi^{\uparrow\top}}_{P'}) \arrow[d, "\text{marginalize to } x"] & \\
(e_x, P'_x) &
\end{tikzcd}$$

Figure 7.2: How to compute $i(\phi\,;\,x)$

The left part of Figure 7.2 is read as follows: In order to know how ϕ changes the original probabilistic choice system $(e_\top, P)$, ϕ has first to be expressed on the top domain. This is done by vacuous extension, $\phi^{\uparrow\top}$. The new probabilistic choice system, after learning $\phi^{\uparrow\top}$, is $(e_\top, P')$, where $P' = P|\phi^{\uparrow\top}$ is the conditional probability distribution relative to $\phi^{\uparrow\top}$. Finally, we do not consider the whole setting any more, but only what lies in our field of interest x. Therefore, $e_\top$ is marginalized to x and the marginal probability distribution P'_x is computed, leading to the probabilistic choice system (e_x, P'_x), where $P' = P|\phi^{\uparrow\top}$.

Now the right part of Figure 7.2 comes into play: As measuring information means looking at the change of uncertainty, the initial uncertainty about the values of the variables in x has also to be determined. This is done by marginalizing $e_\top$ to x and by computing the marginal probability distribution P_x from the prior probability distribution P of the original probabilistic choice system $(e_\top, P)$.

Summing up, disposing of a prior probability distribution P on $\top$ allows to compute ϕ's information content relative to the question $x \in D$:

$$i(\phi\,;\,x) \quad = \quad H(e_x, P_x) - H(e_x, P'_x), \text{ where } P' = P|\phi^{\uparrow\top}. \tag{7.18}$$

One might wrongly assume that ϕ's information content relative to the question $x \in D$ might be computed simpler, with less computational effort. However, this approach, depicted in Figure 7.3, does not yield the correct result.

A possible (but wrong!) way is to look at what is globally known, given by the original probabilistic choice system $(e_\top, P)$. As we are interested in the domain $x \in D$, the overall knowledge is concentrated on the field of interest x. This leads to the probabilistic choice system (e_x, P_x), consisting of the marginalized vacuous information $e_\top^{\downarrow x} = e_x$ and the marginal probability distribution P_x. Then, the piece of information $\phi \in \Phi$ gets known. It is related to the question x by transporting it to x: $\phi^{\rightarrow x}$. Learning $\phi^{\rightarrow x}$ changes the probabilistic choice system inasmuch as the conditional probability distribution relative to $\phi^{\rightarrow x}$ is considered which finally results in the probabilistic choice system $(e_x, P_x|\phi^{\rightarrow x})$.

$$\begin{tikzcd}
(e_\top, P) \arrow[d, "\text{marginalize to } x"] \\
(e_x, P_x) \arrow[d, "\substack{\text{condition} \\ \text{relative to } \phi^{\to x}}"] \\
(e_x, P_x|\phi^{\to x})
\end{tikzcd}$$

Figure 7.3: How not to compute $i(\phi\,;\,x)$

Summing up, we try to compute $i(\phi\,;\,x)$ by

$$H(e_x, P_x) - H(e_x, P_x|\phi^{\to x}). \tag{7.19}$$

Even if intuitively one may be seduced to suppose that Equations 7.18 and 7.19 yield the same result, this is not true. By an example, we will show that even in the very simple case where $x \leq d(\phi)$ and P being the uniform probability distribution, Equations 7.18 and 7.19 do not lead to the same result, when the information content of ϕ is measured relative to x. Example 7.6.1 also clarifies why Equation 7.18 is the right way to compute $i(\phi\,;\,x)$.

Example 7.6.1 The scenario is as follows: We are given an atomic composed information algebra, denoted by (Φ, D), with a lattice D of sets of variables with top domain $\top = \{X, Y\}$. So there are only two variables considered, X and Y. The sets of possible values are specified for each variable by the corresponding frame: $\mathfrak{D}_X = \{x_1, x_2, x_3, x_4\}$ and $\mathfrak{D}_Y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$. We furthermore assume a uniform probability distribution P over the elements of $At_\top(\Phi) = \mathfrak{D}_X \times \mathfrak{D}_Y$. So $p_{X,Y}(x, y) = \frac{1}{32}$ for all atoms $(x, y) \in At_\top(\Phi)$. Consequently, the original probabilistic choice scheme is $(e_\top, P)$. To simplify matters, the piece of information $\phi \in \Phi$ is given on the top domain, but it may be given on any other domain, too. $\phi = \{(x_1, y_1), (x_2, y_2), (x_3, y_2), (x_4, y_2), (x_5, y_2), (x_6, y_2), (x_7, y_2), (x_8, y_2)\}$. Finally, the field of interest x is $\{X\}$.

Let us first compute $i(\phi\,;\,x)$ according Equation 7.18 and Figure 7.2. The vacuous extension to the top domain does not affect ϕ, as $d(\phi) = \top$:

$$\phi^{\uparrow\top} \;=\; \phi.$$

The conditional probability distribution $P' = P|\phi^{\uparrow\top}$ is given for the $(x, y) \in \phi^{\uparrow\top}$ by

$$\begin{array}{rcccc}
p'(x,y) & = & p_{X,Y}|\phi^{\uparrow\top}(x,y) & = & \frac{p_{X,Y}(x,y)}{p_{X,Y}(\phi^{\uparrow\top})} \\
 & = & \frac{p_{X,Y}(x,y)}{\sum\limits_{(x,y)\in\phi^{\uparrow\top}} p_{X,Y}(x,y)} & = & \frac{1}{8}.
\end{array}$$

Now the marginal probability distribution of P' regarding $\{X\}$ has to be computed. So for all $x \in \mathfrak{D}_X$,

$$p'_X(x) = \sum_{y \in \mathfrak{D}_Y} p'(x,y)$$

has to be determined. This results in $P'_X = \{p'_X(x_1), p'_X(x_2), p'_X(x_3), p'_X(x_4)\}$ with

$$\begin{aligned} p'_X(x_1) &= \tfrac{1}{8} + 0 + \ldots + 0 &&= \tfrac{1}{8}, \\ p'_X(x_2) &= 0 + \tfrac{1}{8} + \ldots + \tfrac{1}{8} &&= \tfrac{7}{8}, \\ p'_X(x_3) &= 0, \\ p'_X(x_4) &= 0. \end{aligned}$$

In order to compute the change of uncertainty regarding the value of X, the marginal probability distribution P_X has to be determined, see the right hand side of Figure 7.2. Every $x \in \mathfrak{D}_X$ has the probability

$$\begin{aligned} p_X(x) &= \sum_{y \in \mathfrak{D}_Y} p(x,y) \\ &= 8 \cdot \frac{1}{32} \\ &= \frac{1}{4}. \end{aligned}$$

The information content of ϕ relative to the question $\{X\}$ may now be computed using Equation 7.18:

$$\begin{aligned} i(\phi\,;\,x) = H(e_{\{X\}}, P_X) - H(e_{\{X\}}, P'_X) &= -4 \cdot \frac{1}{4} \log \frac{1}{4} - \left(- \left(\frac{1}{8} \log \frac{1}{8} + \frac{7}{8} \log \frac{7}{8} \right) \right) \\ &\approx 2 - 0.54 \\ &\approx 1.46\,bit \end{aligned}$$

The procedure is conceptually different when Equation 7.19 and Figure 7.3 are considered, as the marginal probability distribution P_X and the transport $\phi^{\to\{X\}}$ are computed immediately and they are used to determine the conditional probability distribution $P_X|\phi^{\to\{X\}}$. P_X has already been calculated above, so

$$p_X(x) = \frac{1}{4}, \quad \forall x \in \mathfrak{D}_X.$$

The transport of ϕ to the domain $\{X\}$ equals in our case the marginalization, as $d(\phi) = \top \geq \{X\}$:

$$\phi^{\to\{X\}} = \phi^{\downarrow\{X\}} = \{x_1, x_2\}.$$

Now, the conditional probability distribution can be computed. $P_X|\phi^{\to\{X\}}$ is given for the $x \in \phi^{\downarrow\{X\}}$ by

$$\begin{aligned} p_X|\phi^{\downarrow\{X\}}(x) &= \frac{p_X(x)}{p_X(\phi^{\downarrow\{X\}})} \\ &= \frac{p_X(x)}{\sum_{x \in \phi^{\downarrow\{X\}}} p_X(x)} \\ &= \frac{1}{2}. \end{aligned}$$

Finally, using Equation 7.19, leads to the following result:

$$\begin{aligned} H(e_{\{X\}}, P_X) - H(e_{\{X\}}, P_X|\phi^{\rightarrow\{X\}}) &= -4 \cdot \frac{1}{4} \log \frac{1}{4} - \left(- \left(2 \cdot \frac{1}{2} \log \frac{1}{2} \right) \right) \\ &= 2 - 1 \\ &= 1\, bit \end{aligned}$$

So there is a clear difference in the results. The conceptual difference between Equation 7.18 and 7.19 is that in the first case, the computations are done as long as possible on the top domain and the marginalization to the field of interest is realized only in the last step. In the second case, however, we immediately resort to the field of interest and thereby loose information. Marginalizing (or transporting) ϕ at the beginning to $\{X\}$ and conditioning on this domain disregards the fact that for the given ϕ the value x_2 is much more probable than x_1. $\odot$

7.7 Conclusion

Information in an information algebra can be measured regarding two different aspects:

- **Qualitative Approach**:
  A piece of information is compared to another by using a partial order $\leq$, given that the piece of information is interpreted disjunctively (as a scheme of choice). If the information algebra is Boolean in addition, a dual partial order $\leq^d$ can be introduced. In that case, information may be interpreted conjunctively (as an enumeration).

- **Quantitative Approach**:
  Based on the concept of entropy, information, possibly weighted by means of a probability distribution, can be measured numerically. Information is measured by the change of uncertainty its arrival causes. The proposed measure requires an atomic composed information algebra. If it is furthermore Boolean, a measure in the dual algebra can be established, as for the qualitative approach.

Part II

Two Information Algebra Instances

8
Propositional Logic

In the two preceding cases the functions to be developed were equated to 0 *and to* 1 *respectively. In the present case I shall suppose the corresponding function equated to any logical symbol* w. *We are then to endeavour to interpret the equation* $V = w$, V *being a function of the logical symbols* x, y, z, *etc.*

George Boole (1815 – 1864)
An Investigation of the Laws of Thought

Propositional logic is one of the best known formalisms for representing information. Even if it is very popular, only its syntactic level (presented in Section 8.1) is widely used. At the same time, the meaning of the information is neglected; this is why we will attach more importance to the semantic level of propositional logic, as illustrated in Section 8.2. In Section 8.3, propositional logic is shown to be an information algebra instance. It is again presented in a semantic view as working with sets of models of formulae is more intuitive than with the formulae themselves when one is interested in the meaning of the information. However, both aspects of propositional logic are considered and the information algebra of sets of models is given (Section 8.3.1), as well as the information algebra of formulae (Section 8.3.2). Finally, based on Chapter 7, a measure of information is proposed in Section 8.4.

8.1 Syntax

As it is the case with natural languages (like English, French or German), each formal language has an alpahbet. By concatenating the elements of such an alphabet, one can form sequences, which are meaningful (see Section 8.2) if they obey certain rules, corresponding to the grammar of a natural language. Such a meaningful entity is called a formula. As the formal language presented here is propositional logic, we will also refer to such a meaningful entitiy as propositional formula.

The language of propositional logic is built on a set $P = \{p_1, p_2, \ldots\}$ of propositions with indices out of $\omega = \{1, 2, \ldots\}$. Therefore, $\mathcal{L}_P$ denotes the language related to set of propositions P. $\mathcal{L}_P$'s alphabet is defined in the following way:

Definition 8.1 (Alphabet)

The alphabet of the language $\mathcal{L}_P$ of propositional logic consists of

propositions:	$p_i \in P$,
connective symbols:	$\neg, \wedge$,
logical constants:	$\top, \bot$,
auxiliary symbols:	*parentheses.*

The propositions p_i are sometimes also called propositional symbols. The set of all propositions P is countable.

If a finite sequence of elements of the alphabet of $\mathcal{L}_P$ obeys certain rules, it is called a well-formed (propositional) formula. Such well-formed formulae are defined inductively:

Definition 8.2 (Propositional Formula)

A finite sequence f of elements from the alphabet of $\mathcal{L}_P$ is a propositional formula *if either*

1. *f is a proposition,*
2. *f is $\top$ or $\bot$,*
3. *f is $\neg g$, for a propositional formula g,*
4. *f is $g \wedge h$, for propositional formulae g and h.*

Formulae that satisfy rule (1) or rule (2) are called atomic propositional formulae. The set of all formulae of the language $\mathcal{L}_P$ is denoted by $Fml(\mathcal{L}_P)$.

In order to increase the legibility of formulae, we introduce "derived" connectives $\vee, \rightarrow, \leftrightarrow$ for formulae $f, g \in Fml(\mathcal{L}_P)$ into the alphabet of the language of propositional logic. They can be reduced to the known symbols $\neg$ and $\wedge$ by the following substitution rules:

$f \vee g$	for	$\neg(f \wedge g)$,
$f \rightarrow g$	for	$\neg f \vee g$,
$f \leftrightarrow g$	for	$(f \rightarrow g) \wedge (f \rightarrow g)$.

8.2 Semantics

In the preceding section, propositional formulae were introduced, which are constituted of elements of the alphabet, assembled following certain rules. It was already pointed out that formulae are meaningful, i. e. they are used to express information in a formal way. In order to get in touch with the meaning of a formula, a semantics for the language of propositional logic has to be introduced. For that purpose, recall that a formula usually contains one or several propositions. In a first step, we will assign a (truth) value to each proposition in P: The value 0 means that the proposition is false; the value 1 means that the proposition is true. The process of truth value ascription to all propositions is accomplished by a (propositional) valuation. Based on such a valuation, we will show in a second step how to assign a truth value to a formula which results in a sequence of $\{0,1\}$-values. Valuations which make a formula evaluate to 1 (true) are called models of this formula. We will finally point out important properties of sets of models and the relations between them.

Definition 8.3 (Propositional Valuation)

A (propositional) valuation *is a mapping*

$$v : P \to \{0,1\}$$

that assigns to every proposition $p_i \in P$ *one of the values 0 or 1.*

As a valuation attributes a value out of $\{0,1\}$ to every proposition in P, it may also be interpreted as a sequence of $\{0,1\}$-values. A valuation is an element of $\{0,1\}^\omega$, which is the set of all possible sequences. In order to stress that a valuation is a sequence of values, we also write $v = \langle v(1)v(2)\ldots\rangle$, where $v(1)$ is the value out of $\{0,1\}$ that v attributes to $p_1 \in P$ and so on.

Valuations are used to assign a truth value out of $\{0,1\}$ to each formula $\phi \in Fml(\mathcal{L}_P)$. The truth assignment

$$\hat{v} : Fml(\mathcal{L}_P) \to \{0,1\} \tag{8.1}$$

always depends on a fixed propositional valuation v. A formula's truth value is obtained by the following truth assignment:

Definition 8.4 (Truth of a Formula)

Let the propositional valuation $v : P \to \{0,1\}$ *be given. Then, we obtain a mapping* $\hat{v} : Fml(\mathcal{L}_P) \to \{0,1\}$ *for assigning a truth value to a formula by the following inductive definition:*

$\hat{v}(f) = 1$ *if and only if, either*

1. f *is* $\top$,
2. f *is* p_i *and* $v(i) = 1$, *for propositions* $p_i \in P$,
3. f *is* $\neg g$ *and* $\hat{v}(g) = 0$, *for a formula* g,
4. f *is* $g \wedge h$ *and* $\hat{v}(g) = \hat{v}(h) = 1$, *for formulae* g, h.

In all the other cases, $\hat{v}(f) = 0$.

If $\hat{v}(f) = 1$, we say that the associated propositional valuation v satisfies f or is a model of f and we write $v \models f$.

For every formula $f \in Fml(\mathcal{L}_P)$, there exists a subset of $\{0,1\}^\omega$, consisting of all the models of f:

Definition 8.5 (Set of Models)

For each formula $f \in Fml(\mathcal{L}_P)$, one can determine its set of models. *It consists of all those valuations v which* satisfy *f:*

$$\mathcal{M}(f) \quad = \quad \{v \in \{0,1\}^\omega : v \models f\}.$$

Finally, the important definition of logical consequence or the entailment relation is given:

Definition 8.6 (Entailment Relation / Logical Consequence)

Let f and g be two formulae $\in Fml(\mathcal{L}_P)$. Then, a formula f is said to entail *another formula g,*

$$f \models g \quad \textit{iff} \quad \mathcal{M}(f) \subseteq \mathcal{M}(g).$$

In this case, g is called a logical consequence *of f.*

This leads to the definition of logical equivalence: Two formulae are equivalent if they entail each other, i. e. if they have the same set of models:

Definition 8.7 (Logical Equivalence)

Let f and g be two formulae $\in Fml(\mathcal{L}_P)$. The formulae f and g are said to be logically equivalent*,*

$$f \equiv g \quad \textit{iff} \quad f \models g \textit{ and } g \models f.$$

In this case, $\mathcal{M}(f) = \mathcal{M}(g)$.

8.3 Propositional Logic as Domain-Free Information Algebra

We are now going to show that propositional logic is an information algebra instance. For that purpose, we propose in Section 8.3.1 a domain-free information algebra of sets of models. Thereafter, an isomorphism between the information algebra of sets of models and that of propositional formulae of $Fml(\mathcal{L}_P)$ is given (Section 8.3.2).

8.3.1 Information Algebra of Sets of Models

We will first construct a system (Ψ, D) and will prove afterwards that this system is actually a domain-free information algebra, i. e. fulfills the axioms given in Section 5.2. Let D be the lattice of finite subsets of the index set $\omega = \{1, 2, 3, \ldots\}$ of the set P of propositions. In order to define Ψ, which will turn out to consist of specific sets of valuations, some further notions have to be introduced.

For any $x \in D$, one can determine the projection of some valuation $v \in \{0,1\}^\omega$. As x refers to a finite subset of propositions p_i with $i \in x$, only the values $v(i)$ of v are taken into account. The projection of a valuation thus results in a sequence of $\{0,1\}$-values of length $|x|$, which will also be called x-tuple (see Section 2.4).

Definition 8.8 (Valuation Projection)

For $x \in D$ and a valuation $v \in \{0,1\}^\omega$, the valuation projection *is denoted by $v^{\downarrow x}$. It is a sequence of those values $v(i)$ of v with $i \in x$.*

The definition of valuation projection allows to determine whether two valuations agree on the values of x. In that case, they are called x-equivalent, as they provide the same values $v(i)$ for the propositions p_i, $i \in x$:

Definition 8.9 (x-equivalent)

*Given $x \in D$. Two valuations $v, w \in \{0,1\}^\omega$ are x-*equivalent*,*

$$v \equiv_x w \quad \textit{iff} \quad v^{\downarrow x} = w^{\downarrow x}.$$

The corresponding equivalence class is $[v]_x = \{w \in \{0,1\}^\omega : v \equiv_x w\}$.

Obviously, such an x-equivalence class $[v]_x$ is a subset of $\{0,1\}^\omega$. The focusing of any subset M of $\{0,1\}^\omega$ on some $x \in D$ is defined by

$$M^{\Rightarrow x} \quad := \quad \bigcup_{v \in M} [v]_x. \tag{8.2}$$

So $M^{\Rightarrow x}$ is again a subset of $\{0,1\}^\omega$, consisting of every valuation which is x-equivalent to one of the valuations constituting M. Those sets M which do not change under focusing are of special interest:

Definition 8.10 (Cylindric Set)

*Given $x \in D$. A subset $M \subseteq \{0,1\}^\omega$ is called x-*cylindric *or* cylindric over *x iff*

$$M \quad = \quad M^{\Rightarrow x}.$$

The family of x-cylindric subsets of $\{0,1\}^\omega$ is denoted by Ψ_x:

$$\Psi_x \quad := \quad \{M \subseteq \{0,1\}^\omega : M = M^{\Rightarrow x}, x \in D\}. \tag{8.3}$$

Note that every subset M of $\{0,1\}^\omega$ provides valuations which are models of some formula $f \in Fml(\mathcal{L}_P)$.[1] If the set M is not cylindric over some $x \in D$, then M does only contain some of the models of f and is not *the* set of models of f (containing all models), but only a strict subset of $\mathcal{M}(f)$, see Definition 8.5. In the case of M cylindric, M has maximum size and thus equals $\mathcal{M}(f)$.

Now we define the set Ψ of pieces of information as follows:

$$\Psi \quad = \quad \bigcup_{x \in D} \Psi_x. \tag{8.4}$$

The set Ψ of pieces of information is the set of those sets of valuations which are cylindric over some $x \in D$. Recall that D is the lattice of finite subsets of the index set $\omega = \{1,2,3,\ldots\}$ of the set P of propositions. By now, the system (Ψ, D) has been constructed for propositional logic. However, and this might be an unusual approach for logicians, propositional logic is not considered on the syntactic level, i. e. the pieces of information are not represented by formulae. Motivated by the semantic approach of this thesis, we have chosen to identify the pieces of information with sets ψ of valuations. A valuation v is an assignment of truth values to the propositions in P, and is thus a part of the semantics of propositional logic. Sets of valuations are actually sets of models of some propositional formula. In Section 8.3.2, the information algebra of sets of models will be linked to the information algebra of formulae. But first of all, we need to prove that the system (Ψ, D) constructed for propositional logic is in fact an information algebra instance. In order to do that, the operations of combination and focusing have to be identified.

The combination operation applies to two pieces of information $\phi, \psi \in \Psi$, which are sets of valuations. Combination is simply defined as set intersection:

$$\phi \otimes \psi \quad := \quad \phi \cap \psi. \tag{8.5}$$

The focusing operation involves a piece of information $\psi \in \Psi$, which is a set of valuations, and an element $x \in D$. According to Definition 8.9 and Equation 8.2 above, it is defined as the set of all valuations which are x-equivalent to the valuations which constitute ψ:

$$\psi^{\Rightarrow x} \quad := \quad \bigcup_{v \in \psi} [v]_x. \tag{8.6}$$

We now introduce some lemmata which capture important properties of the combination and focusing operations. These lemmata will be used in the proof of Theorem 8.15 below. The first lemma points out the intersection of two x-cylindric sets is again x-cylindric. A consequence of this lemma is that the set Ψ_x of x-cylindric pieces of information is closed under combination (intersection).

[1] For the sake of completeness: Such a formula f exists for all subsets M of $\{0,1\}^\omega$ and can be trivially be found by first constituting for every valuation $v \in M$ a formula g, which contains each proposition $p_i \in P$ exactly once. The propositions of P are connected by conjunction, where p_i occurs negated in g if the ith position of the valuation v equals 0 and p_i occurs not negated in g otherwise. Then, the formula f is built by disjunction of all these formulae g. So we can conclude that such a formula f exists for all possible subsets of $\{0,1\}^\omega$, especially for the cylindric sets.

Lemma 8.11 *Given two pieces of information $\phi, \psi \in \Psi$ which are cylindric over $x \in D$. Their combination $\phi \otimes \psi$ is again cylindric over x.*

Proof.

$$\begin{aligned}
\phi \otimes \psi &= \phi \cap \psi \\
&= \{u \in \{0,1\}^\omega : u \in \phi \text{ and } u \in \psi\} \\
&= \{u \in \{0,1\}^\omega : \text{ there is a } v \in \phi, v' \in \psi \text{ such that } u \equiv_x v \text{ and } u \equiv_x v'\} \\
&= \{u \in \{0,1\}^\omega : \text{ there is a } w \in \phi \cap \psi \text{ such that } u \equiv_x w\} \\
&= \bigcup_{w \in (\phi \cap \psi)} [w]_x \\
&= (\phi \otimes \psi)^{\Rightarrow x}.
\end{aligned}$$

□

In the second line of the above proof, we are making use of the fact that ϕ and ψ are x-cylindric. The valuations of an x-cylindric piece of information assign the same values to the propositions in x.

The next lemma proposes another representation for an equivalence class $[v]_x$. It is used in the proof of Lemma 8.13 below.

Lemma 8.12 *If $x \leq y$, a set $[v]_x$ of x-equivalent valuations may be written as*

$$[v]_x = \bigcup_{u \in [v]_x} [u]_y.$$

Proof. The above equality will be shown in two steps:

- $[v]_x \supseteq \bigcup_{u \in [v]_x} [u]_y$:
  $x \leq y$ means $x \subseteq y$. $u \in [v]_x$ and $x \subseteq y$ imply $[u]_y \subseteq [v]_x$, as every element of $[u]_y$ assigns the same values to the propositions in y, and thus also to those in $x \subseteq y$. The more propositions there are for which the valuations have to agree, the smaller is the equivalence class. As all elements which are unified are a subset of $[v]_x$, the union will not exceed $[v]_x$.

- $[v]_x \subseteq \bigcup_{u \in [v]_x} [u]_y$:
  By definition, $\{u\} \subseteq [u]_y$ and $[v]_x = \bigcup_{u \in [v]_x} u$. The union $\bigcup_{u \in [v]_x} u$ will not exceed the union $\bigcup_{u \in [v]_x} [u]_y$, as the elements of the former are subsets of the elements of the latter.

As $[v]_x \supseteq \bigcup_{u \in [v]_x} [u]_y$ and $[v]_x \subseteq \bigcup_{u \in [v]_x} [u]_y$, both sets are equal, q.e.d. □

Every piece of information $\psi \in \Psi$ is cylindric over some domain $x \in D$. The following lemma states that it is also cylindric over each finer domain.

Lemma 8.13 *A piece of information $\psi \in \Psi$, which is cylindric over $x \in D$, is also cylindric over any $y \in D$ with $y \geq x$.*

Proof.

$$\begin{aligned}
\psi^{\Rightarrow y} &\stackrel{(1)}{=} \bigcup_{w \in \psi} [w]_y \\
&\stackrel{(2)}{=} \bigcup_{w \in \bigcup_{v \in \psi}[v]_x} [w]_y \\
&\stackrel{(3)}{=} \bigcup_{v \in \psi} \left(\bigcup_{w \in [v]_x} [w]_y \right) \\
&\stackrel{(4)}{=} \bigcup_{v \in \psi} [v]_x \\
&\stackrel{(5)}{=} \psi^{\Rightarrow x} = \psi
\end{aligned}$$

As $\psi^{\Rightarrow y} = \psi$, this piece of information is also y-cylindric. □

In the first line, the definition of focusing (Equation 8.6) is applied. In the second line ψ being x-cylindric is used. The third line is obtained by applying the generalized version of the associative law for set union (Halmos, 1974, page 35). Passing to the fourth line is made possible by the above Lemma 8.12. Finally, the fifth line goes once more back to the definition of focusing and the fact that ψ is x-cylindric.

Lemma 8.14 *For a piece of information $\psi \in \Psi$ and for any $x \in D$,*

$$\psi^{\Rightarrow x} = (\psi^{\Rightarrow x})^{\Rightarrow x}.$$

Proof. The underlying idea is that the set $\psi^{\Rightarrow x}$ is already constituted of the maximum set of equivalent valuations. In other words, $\psi^{\Rightarrow x}$ is cylindric over x, so the proof is similar to that of Lemma 8.13:

$$\begin{aligned}
(\psi^{\Rightarrow x})^{\Rightarrow x} &= \bigcup_{w \in \psi^{\Rightarrow x}} [w]_x \\
&= \bigcup_{w \in \bigcup_{v \in \psi}[v]_x} [w]_x \\
&= \bigcup_{v \in \psi} \left(\bigcup_{w \in [v]_x} [w]_x \right) \\
&= \bigcup_{v \in \psi} [v]_x \\
&= \psi^{\Rightarrow x}
\end{aligned}$$

□

Now we dispose of the entire toolbox for proving that propositional logic forms on the semantic level an information algebra.

Theorem 8.15 *The system (Ψ, D), with the set Ψ of pieces of information, given in Equation 8.4 and the lattice D of finite subsets of the index set $\omega = \{1, 2, 3, \ldots\}$ of the set P of propositions, satisfies the five axioms of a domain-free information algebra (see Section 5.2), where the operation of combination is given by Equation 8.5 and the operation of focusing is given by Equation 8.6.*

Proof. The above theorem is proven by making use of the fact that the pieces of information of Ψ are cylindric sets of valuations. The Lemmata 8.11 to 8.14 play an important role in the following proof:

1. The *semigroup* axiom requires that combination is commutative and associative. It is well-known that set intersection fulfills these requirements. Futhermore, $e = \{0,1\}^\omega$ is the neutral element[2] and the null element is $z = \emptyset$.

   It remains to show that Ψ is closed under combination. For that purpose, we have to prove that the intersection of two arbitrary cylindric sets results in a cylindric set: Consider two pieces of information $\phi, \psi \in \Psi$. We assume without loss of generality ϕ to be x-cylindric and ψ to be y-cylindric. As $x, y \in D$ and from Section 3.1, it is known that $x, y \leq x \vee y$. With the above Lemma 8.13 we can therefore justify that ϕ and ψ are both $x \vee y$-cylindric. Finally, by Lemma 8.11 we can conclude that $\phi \otimes \psi$ is also $x \vee y$-cylindric, q.e.d.

2. The proof of the *transitivity of focusing* is a generalization of Lemma 8.13, holding for any $x, y \in D$ and $\psi \in \Psi$.

$$
\begin{aligned}
(\psi^{\Rightarrow x})^{\Rightarrow y} &= \bigcup_{w \in \psi^{\Rightarrow x}} [w]_y = \bigcup_{w \in \bigcup_{v \in \psi} [v]_x} [w]_y \\
&= \bigcup_{v \in \psi} \left(\bigcup_{w \in [v]_x} [w]_y \right) \\
&= \bigcup_{v \in \psi} \left(\bigcup_{w \in [v]_x} \{w' \in \{0,1\}^\omega : w' \equiv_y w\} \right) \\
&= \bigcup_{v \in \psi} \{u \in \{0,1\}^\omega : \text{ there is a } w \in [v]_x \text{ such that } w \equiv_y u\} \\
&= \bigcup_{v \in \psi} \{u \in \{0,1\}^\omega : \text{ there is a } w \in \{0,1\}^\omega, \text{ with } w \equiv_x v, \\
&\qquad\qquad \text{such that } w \equiv_y u\} \\
&= \bigcup_{v \in \psi} [v]_{x \cap y} \\
&= \psi^{\Rightarrow x \cap y}
\end{aligned}
$$

[2] Ψ is only composed of sets which are cylindric over some *finite* $x \in D$. Even if the index set ω of the set P of propositions is not finite, $e = \{0,1\}^\omega$ is cylindric over all finite subsets of ω.

3. The proof of the *combination* axiom is based on the definitions of focusing (Equation 8.6) and combination (Equation 8.5).

$$\begin{aligned}
(\phi^{\Rightarrow x} \otimes \psi)^{\Rightarrow x} &= \bigcup_{v \in (\phi^{\Rightarrow x} \otimes \psi)} [v]_x \\
&= \left(\bigcup_{v \in \phi^{\Rightarrow x}} [v]_x \right) \cap \left(\bigcup_{v \in \psi} [v]_x \right) \\
&= (\phi^{\Rightarrow x})^{\Rightarrow x} \cap \psi^{\Rightarrow x} \\
&\overset{\text{Lemma 8.14}}{=} \phi^{\Rightarrow x} \otimes \psi^{\Rightarrow x}
\end{aligned}$$

4. The *idempotency* axiom is proven using the definition of focusing:

$$\psi^{\Rightarrow x} = \bigcup_{v \in \psi} [v]_x \supseteq \psi, \text{ as } [v]_x \supseteq \{v\}.$$

The intersection of a superset of ψ with ψ itself results in the latter. This shows that $\psi^{\Rightarrow x} \otimes \psi = \psi^{\Rightarrow x} \cap \psi = \psi$.

5. The *support* axiom is satisfied by definition, as every $\psi \in \Psi$ is cylindric over some $x \in D$.

□

8.3.2 Information Algebra of Formulae

In what follows, we will show that propositional logic forms also on the syntactic level an information algebra. Its pieces of information are, however, not formulae, but sets of formulae. That is to say, it can be observed that any propositional formula $f \in Fml(\mathcal{L}_P)$ determines an element ψ of the above information algebra of sets of models (Ψ, D), where $\psi = \mathcal{M}(f)$. So we can state that the set of models of a formula is the information this formula describes. From this point of view, Definition 8.7 is suddenly more meaningful: Two formulae are equivalent if they have the same set of models, which means nothing else than that they describe the same information. For that reason, we have chosen the semantic approach to propositional logic. It allows to understand in which way the propositional language expresses information.

Based on this observation, we will partition the set $Fml(\mathcal{L}_P)$ into sets of equivalent formulae. Thus we get a quotient set $Fml(\mathcal{L}_P)/\mu$, consisting of sets of formulae which are equivalence classes, generated by the equivalence relation μ:

$$f \equiv g \pmod{\mu} \quad \text{iff} \quad \mathcal{M}(f) = \mathcal{M}(g), \text{ for } f, g \in Fml(\mathcal{L}_P). \tag{8.7}$$

The quotient set $Fml(\mathcal{L}_P)/\mu$ will be shown to be an information algebra. In a nutshell, this will be done by showing that there is an isomorphism $i : Fml(\mathcal{L}_P)/\mu \to \Psi$. We will proceed as follows: In a first step, the two-sorted algebra $(Fml(\mathcal{L}_P)/\mu, D)$

is considered, and two operations $\otimes'$ and $\Rightarrow'$ are defined on it. The neutral element e' is also identified. In a second step, we show that $i : Fml(\mathcal{L}_P)/\mu \rightarrow \Psi$ is actually an isomorphism. Then, we can conclude that $(Fml(\mathcal{L}_P)/\mu, D)$ is an information algebra, as (Ψ, D) is known to be one.

In (Kohlas, 2003, Chapter 3.3.2), an isomorphism is defined as a bijective mapping between two information algebras, preserving the combination and the focusing operation, as well as the neutral element. But in the above sketch of our proof, $(Fml(\mathcal{L}_P)/\mu, D)$ is only an algebra, not an information algebra. An isomorphism has therefore to be defined between two algebraic structures, but only the codomain is an information algebra. We thus presuppose the following theorem, whose proof is trivial.

Theorem 8.16

$(Fml(\mathcal{L}_P)/\mu, D)$, $\otimes', \Rightarrow', e'$	&	$i : Fml(\mathcal{L}_P)/\mu \rightarrow \Psi$ *is an isomorphism between an algebraic structure and an information algebra,*	$\Leftrightarrow$	$(Fml(\mathcal{L}_P)/\mu, D)$ *is an information algebra.*

The First Step: Defining $(Fml(\mathcal{L}_P)/\mu, D), \otimes', \Rightarrow', e'$

The quotient set $Fml(\mathcal{L}_P)/\mu$ is constituted of all equivalence classes, generated by the equivalence relation μ of Equation 8.7. For a proof that μ really is an equivalence relation, see the one of Lemma 9.23, which can be taken over one-to-one. Those formulae which have the same set of models are grouped together and form an equivalence class. The equivalence class of a formula $f \in Fml(\mathcal{L}_P)$ is

$$[f]_\mu \quad = \quad \{g \in Fml(\mathcal{L}_P) : g \equiv f \pmod{\mu}\}. \tag{8.8}$$

Now, the quotient set $Fml(\mathcal{L}_P)/\mu$ can be defined by the set of all equivalence classes of formulae.

$$Fml(\mathcal{L}_P)/\mu \quad := \quad \{[f]_\mu : f \in Fml(\mathcal{L}_P)\}. \tag{8.9}$$

D is defined in the same way as in Section 8.3.1 for the information algebra of sets of models. Given the set P of propositions,

$$D \text{ is the lattice of finite subsets} \tag{8.10}$$

of the index set $\omega = \{1, 2, 3, \ldots\}$ of P.

We now come to the definitions of the two operations $\otimes'$ and $\Rightarrow'$. They are defined with reference to the operations $\otimes$ and $\Rightarrow$ of Section 8.3.1, where $\otimes$ is given by set intersection and $\Rightarrow$ according to Equation 8.2.

$\otimes'$ maps any two equivalence classes of formulae $[f]_\mu, [g]_\mu \in Fml(\mathcal{L}_P)/\mu$ to $[f]_\mu \otimes' [g]_\mu$, which is again in $Fml(\mathcal{L}_P)/\mu$:

$$\otimes' : Fml(\mathcal{L}_P)/\mu \times Fml(\mathcal{L}_P)/\mu \rightarrow Fml(\mathcal{L}_P)/\mu \quad \text{and} \quad ([f]_\mu, [g]_\mu) \mapsto [f]_\mu \otimes' [g]_\mu.$$

This operation is defined as follows:

$$[f]_\mu \otimes' [g]_\mu \quad := \quad [h]_\mu, \text{ such that } \mathcal{M}(h) = \mathcal{M}(f) \otimes \mathcal{M}(g). \tag{8.11}$$

$\Rightarrow'$ involves an equivalence class $[f]_\mu \in Fml(\mathcal{L}_P)/\mu$ and an element $x \in D$. The application of $\Rightarrow'$ results in $[f]_\mu^{\Rightarrow' x}$, which is again in $Fml(\mathcal{L}_P)/\mu$:

$$\Rightarrow' \colon Fml(\mathcal{L}_P)/\mu \times D \to Fml(\mathcal{L}_P)/\mu \quad \text{and} \quad ([f]_\mu, x) \mapsto [f]_\mu^{\Rightarrow' x}.$$

This operation is defined as follows:

$$[f]_\mu^{\Rightarrow' x} \quad := \quad [g]_\mu, \text{ such that } \mathcal{M}(g) = (\mathcal{M}(f))^{\Rightarrow x}. \tag{8.12}$$

Finally, the element $e' \in Fml(\mathcal{L}_P)/\mu$ is the equivalence class $[f]_\mu$ of those formulae whose set of models is the neutral element $e = \{0,1\}^\omega$ of Ψ:

$$e' \quad := \quad [f]_\mu, \text{ such that } \mathcal{M}(f) = e. \tag{8.13}$$

The Second Step: Isomorphism $i : Fml(\mathcal{L}_P)/\mu \to \Psi$

From (Kohlas, 2003, Chapter 3.3.2) it is known that in order to count as an isomorphism, the mapping $i : Fml(\mathcal{L}_P)/\mu \to \Psi$ has to be bijective and for any $[f]_\mu, [g]_\mu \in Fml(\mathcal{L}_P)/\mu$ and $x \in D$, the three following conditions have to be met:

(I1) $i([f]_\mu \otimes' [g]_\mu) = i([f]_\mu) \otimes i([g]_\mu)$,

(I2) $i([f]_\mu^{\Rightarrow' x}) = i([f]_\mu)^{\Rightarrow x}$,

(I3) $i(e') = e$.

The mapping $i : Fml(\mathcal{L}_P)/\mu \to \Psi$ we are looking for is defined by making use of $\mathcal{M} : Fml(\mathcal{L}_P) \to \mathfrak{P}(\{0,1\}^\omega)$, given in Definition 8.5. Then, i maps any equivalence class $[f]_\mu$ of formulae to the set $\mathcal{M}(f)$ of models which are shared by all the formulae of the equivalence class:

$$i([f]_\mu) \quad := \quad \mathcal{M}(f). \tag{8.14}$$

One may comment that according to Equation 8.4, Ψ, the codomain of i, consists only of cylindric subsets of $\{0,1\}^\omega$, an not to the whole powerset $\mathfrak{P}(\{0,1\}^\omega)$. This is fortunately guaranteed, as $\mathcal{M}$ maps every formula $f \in Fml(\mathcal{L}_P)$ to its set of models $\mathcal{M}(f)$, which is a cylindric set (see Definition 8.10 and the subsequent explanation).

In order to count as an isomorphism, $i : Fml(\mathcal{L}_P)/\mu \to \Psi$ has not only to satisfy the conditions $(I1) - (I3)$ above (shown in Theorem 8.17 below), but it has also to be a bijective mapping, i. e. one-to-one (injective) and onto (surjective). These two requirements are easily shown:

- The mapping $i : Fml(\mathcal{L}_P)/\mu \rightarrow \Psi$ is injective, if for any $[f]_\mu, [g]_\mu \in Fml(\mathcal{L}_P)/\mu$,

$$i([f]_\mu) = i([g]_\mu) \quad \text{implies} \quad [f]_\mu = [g]_\mu.$$

  This is warranted, as any two formulae which share the same set of models also determine the same equivalence class:

$$\mathcal{M}(f) = \mathcal{M}(g) \quad \text{implies} \quad [f]_\mu = [g]_\mu.$$

- The mapping $i : Fml(\mathcal{L}_P)/\mu \rightarrow \Psi$ is surjective, if for every $\psi \in \Psi$, there is an $[f]_\mu \in Fml(\mathcal{L}_P)/\mu$, such that

$$i([f]_\mu) \quad = \quad \psi.$$

  This is guaranteed, as already explained on page 110, after Definition 8.10. Ψ only contains cylindric sets ψ, which are the sets of models of the formulae of $Fml(\mathcal{L}_P)$:

$$\mathcal{M}(f) \quad = \quad \psi.$$

  As in the quotient set $Fml(\mathcal{L}_P)/\mu$, every formula $f \in Fml(\mathcal{L}_P)$ belongs to some equivalence class, there corresponds to every $\psi \in \Psi$ at least one[3] $[f]_\mu \in Fml(\mathcal{L}_P)/\mu$.

Now that the mapping i has been shown to be bijective, we can take advantage of its properties. An important consequence of the bijectivity of a mapping i is the existence of an inverse mapping i^{-1}. As $i : Fml(\mathcal{L}_P)/\mu \rightarrow \Psi$ is a bijective mapping, defined in Equation 8.14, there exists an inverse mapping $i^{-1} : \Psi \rightarrow Fml(\mathcal{L}_P)/\mu$, obviously defined as follows:

$$\mathcal{M}(f) \quad := \quad i([f]_\mu). \tag{8.15}$$

This inverse mapping i^{-1} plays an important in the proof of the following theorem, where i is shown to be an isomorphism.

Theorem 8.17 *The bijective mapping $i : Fml(\mathcal{L}_P)/\mu \rightarrow \Psi$ of Equation 8.14 satisfies the conditions $(I1) - (I3)$ above. Therefore, i is an isomorphism.*

[3] actually, exactly one, to be precise,

Proof.

$$
\begin{array}{llcl}
(I1) & i([f]_\mu \otimes' [g]_\mu) & \overset{\text{Eq. 8.11}}{=} & i([h]_\mu), \quad \text{with } \mathcal{M}(h) = \mathcal{M}(f) \otimes \mathcal{M}(g), \\
 & & \overset{\text{Eq. 8.14}}{=} & \mathcal{M}(h) \\
 & & = & \mathcal{M}(f) \otimes \mathcal{M}(g) \\
 & & \overset{\text{Eq. 8.15}}{=} & i([f]_\mu) \otimes i([g]_\mu).
\end{array}
$$

$$
\begin{array}{llcl}
(I2) & i([f]_\mu^{\Rightarrow' x}) & \overset{\text{Eq. 8.12}}{=} & i([g]_\mu), \quad \text{with } \mathcal{M}(g) = (\mathcal{M}(f))^{\Rightarrow x}, \\
 & & \overset{\text{Eq. 8.14}}{=} & \mathcal{M}(g) \\
 & & = & (\mathcal{M}(f))^{\Rightarrow x} \\
 & & \overset{\text{Eq. 8.15}}{=} & i([f]_\mu)^{\Rightarrow x}.
\end{array}
$$

$$
\begin{array}{llcl}
(I3) & i(e') & \overset{Eq. 8.13}{=} & i([f]_\mu), \quad \text{with } \mathcal{M}(f) = e, \\
 & & \overset{\text{Eq. 8.14}}{=} & \mathcal{M}(f) \\
 & & = & e.
\end{array}
$$

As the mapping i was already proven to be bijective, we can now justifiably call it an isomorphism. □

Information Algebra of Equivalence Classes of Formulae

Recall that our goal was to show that the quotient set $Fml(\mathcal{L}_P)/\mu$, containing equivalence classes of formulae, forms an information algebra. Therefore, the following schematic overview of the different steps of the proof was given above:

$(Fml(\mathcal{L}_P)/\mu, D), \otimes', \Rightarrow', e'$	&	$i : Fml(\mathcal{L}_P)/\mu \to \Psi$ is an isomorphism,	$\Leftrightarrow$	$(Fml(\mathcal{L}_P)/\mu, D)$ is an information algebra.

In Equations 8.9 to 8.13, the two-sorted algebra $(Fml(\mathcal{L}_P)/\mu, D)$ of equivalence classes of formulae, the two operations $\otimes'$ and $\Rightarrow'$, as well as the neutral element e' are defined. Theorem 8.17 states that there exists an isomorphism between $Fml(\mathcal{L}_P)/\mu$ and Ψ. As in Section 8.3.1, (Ψ, D) has been shown to form an information algebra, the same holds by (Kohlas, 2003, Chapter 3.3.2) for the algebra $(Fml(\mathcal{L}_P)/\mu, D)$ with $\otimes', \Rightarrow'$ and e'. Thus, equivalence classes of formulae also form an information algebra.

8.4 Measure

As explained in Chapter 7, there are different views on information and thus different interpretations of the measure. Furthermore, there is the possibility to measure information content qualitatively or quantitatively. In the following, we propose a measure for propositional logic, based on Chapter 7.

8.4.1 Qualitative Measure

In Sections 7.2 and 7.3, the foundations of two qualitative measures have been laid. Both scenarios apply, as it is well-known that the elements of the domain-free information algebra we are dealing with form not only a semi-lattice, but even a Boolean lattice. In the case of propositional logic, the information algebra of sets of models as well as the information algebra of (equivalence classes of) formulae are Boolean.[4]

Disjunctive View - Partial Order

According to Section 7.2, the qualitative measure in the disjunctive view on pieces of information is given by the partial order. Even if its definition is there only given in terms of labeled information algebras, it applies in the same manner to domain-free information algebras, as explained in Section 6.2 and in the footnotes of Section 7.2. Consider the information algebra (Ψ, D), whose pieces of information are sets of models of some propositional formula of $Fml(\mathcal{L}_P)$. The set Ψ of pieces of information is given by Equation 8.4. Applying Definition 7.2 to $\phi, \psi \in \Psi$, where combination is set intersection, we say that the piece of information ϕ is less informative than the piece of information ψ and write

$$\phi \leq \psi \quad \text{iff} \quad \phi \cap \psi = \psi. \tag{8.16}$$

The above equation is a partial order between sets of models. It expresses the meaning of the partial order between two formulae f and g of $Fml(\mathcal{L}_P)$, if we fix without loss of generality $\phi = \mathcal{M}(g)$ and $\psi = \mathcal{M}(f)$. $\phi \leq \psi$ can now be further developed as follows:

$$\phi \cap \psi = \psi \Leftrightarrow \phi \supseteq \psi \Leftrightarrow \mathcal{M}(g) \supseteq \mathcal{M}(f) \overset{\text{Definition 8.6}}{\Leftrightarrow} f \models g. \tag{8.17}$$

As to the information algebra of (equivalence classes of) formulae $(Fml(\mathcal{L}_P)/\mu, D)$, whose pieces of information are isomorphic to those of (Ψ, D), the partial order can be expressed by means of the above considerations as

$$[g]_\mu \leq [f]_\mu \quad \text{iff} \quad f \models g. \tag{8.18}$$

Regarding a fixed question $x \in D$, a preorder is proposed in Section 7.2.2. This order is in the case of two sets of models $\phi, \psi \in \Psi$ given by

$$\phi \leq_x \psi \quad \text{iff} \quad \phi^{\Rightarrow x} \subseteq \psi^{\Rightarrow x}, \tag{8.19}$$

and in the case of two equivalence classes of formulae $[f]_\mu, [g]_\mu \in Fml(\mathcal{L}_P)/\mu$ by

$$[g]_\mu \leq_x [f]_\mu \quad \text{iff} \quad [g]_\mu^{\Rightarrow' x} \supseteq [f]_\mu^{\Rightarrow' x}. \tag{8.20}$$

[4] See (Kohlas, 2003, Chapter 6.5.2) for the definition of domain-free Boolean information algebras. In the Boolean algebra of sets of models, join is set intersection, meet is set union and the complement operator is the set complement. In the Boolean algebra of (equivalence classes of) formulae, join is conjunction, meet is disjunction and the complement operator is negation.

Conjunctive View - Dual Partial Order

The dual partial order $\leq^d$, as proposed in Section 7.3, is based on the dual combination $\otimes^d$. As we are dealing with a Boolean lattice, dual combination is dual join, which is actually the meet. This corresponds to set union in the propositional logic case, as combination is set intersection. The qualitative measure in the conjunctive view on pieces of information is given by the dual partial order. For two sets of models $\phi, \psi \in \Psi$ we have

$$\phi \leq^d \psi \quad \text{iff} \quad \phi \cup \psi = \psi. \tag{8.21}$$

By the same reasoning as above, this can be further developed as:

$$\phi \cup \psi = \psi \Leftrightarrow \phi \subseteq \psi \Leftrightarrow \mathcal{M}(g) \subseteq \mathcal{M}(f) \overset{\text{Definition 8.6}}{\Leftrightarrow} g \models f. \tag{8.22}$$

So for two equivalence classes of formulae $[f]_\mu, [g]_\mu \in Fml(\mathcal{L}_P)/\mu$ we get

$$[g]_\mu \leq [f]_\mu \quad \text{iff} \quad g \models f. \tag{8.23}$$

As the information algebras (Ψ, D) and $(Fml(\mathcal{L}_P)\mu, D)$ are both Boolean, it holds for the sets of models $\phi, \psi \in \Psi$ and for the equivalence classes of formulae $[f]_\mu, [g]_\mu \in Fml(\mathcal{L}_P)/\mu$ that

$$\phi \leq^d \psi \text{ iff } \psi \leq \phi \quad \text{and} \quad [f]_\mu \leq^d [g]_\mu \text{ iff } [g]_\mu \leq [f]_\mu.$$

Similarly, $\leq^d_x$, the dual partial order regarding a question $x \in D$ can also be formulated in terms of the partial order $\leq$, see Section 7.3.

8.4.2 Quantitative Measure

As pointed out in Section 7.4, the quantitative measure of information content requires the pieces of information to be part of an atomic composed labeled information algebra, which will be denoted by (Φ, D). Furthermore, we assume that for all domains $x \in D$, the set $At_x(\Phi)$ of all atoms of the domain x is finite. Even if we have looked at propositional logic in this chapter only from a domain-free point of view, we now consider the labeled information algebra for the measure of information. This is completely unproblematic, as both approaches are equivalent. In Section 5.3, it has been shown how to construct a labeled information algebra from a domain-free one.[5]

The components and operations of the atomic composed information algebra (Φ, D) are listed below. Following the semantic approach of this thesis, the pieces of information and the operations are provided on the semantic level. For that purpose, we use tuples, which have already been introduced in Chapter 2.

[5] Alternatively, see (Langel & Kohlas, 2005) for a description of propositional logic as labeled information algebra.

- D: lattice of finite subsets of the set $P = \{p_1, p_2, p_3 \ldots\}$
- atom: x-tuple $f : x \to \{0,1\}^{|x|}$, where $x \in D$
  (To every proposition $p_i \in x \subseteq P$, the value 0 or 1 is attributed.)
- set of pieces of information: $\Phi = \bigcup_{x \in D} \Phi_x$, where $\Phi_x = \mathfrak{P}(\{0,1\}^{|x|})$
- piece of information: $\phi \in \Phi$, such that $\phi = \wedge At(\phi)$
- neutral element: $e_x = \{0,1\}^{|x|} = At_x(\Phi)$
- null element: $z_x = \emptyset_x$
  (By convention, every $At_x(\Phi)$ has its own empty subset $\emptyset_x$.)
- labeling: $d(\phi) = x$, if $\phi \in \Phi_x$
- marginalization: $\phi^{\downarrow x} = \{f[x] : f \in \phi\}$, where
  - $x \leq d(\phi)$ and
  - $f[x] = g$, such that $g(p_i) = f(p_i)$, for all $p_i \in x$
- combination: $\phi \otimes \psi = \{f \in \{0,1\}^{|x \cup y|} : f[x] \in \phi, f[y] \in \psi\}$
- vacuous extension: $\phi^{\uparrow x} = \phi \otimes e_x$, where $x \geq d(\phi)$
- transport: $\phi^{\to x} = (\phi^{\uparrow d(\phi) \cup x})^{\downarrow x}$, for all $x \in D$

Using the above definitions and operations, the measures of information content proposed in Section 7.4 can directly be applied to propositional logic. The same holds for the measures of information content in the dual algebra, seen in Section 7.5. When probabilities are taken into account (Section 7.6), we dispose in addition of a probability distribution over the tuples. This leads to a probabilistic choice system and the proposed measure is based on entropy.

8.5 Conclusion

Two aspects of propositional logic have been presented:

1. On a syntactic level, propositional logic was looked at as a formal language expressing some information by formulae.
2. The information described by a formula is given by its set of models. This is the semantic level of propositional logic.

Thereafter, it has been shown that propositional logic is an information algebra instance. The proof has been carried out on the semantic level, leading to an information algebra of sets of models. Based on this result, (equivalence classes of) formulae were shown to form as well an information algebra, as both algebras are isomorphic. Finally, we have explained how to apply the measures of information content from Chapter 7 to propositional logic, by giving the specific definitions and operations.

9
Predicate Logic

Cylindric algebras are algebraic structures arising by abstraction from two sources: algebras of formulas determined by deductive systems of logic, and algebras of subsets of some Cartesian product space, among whose operations are those of cylindrification parallel to each axis.

Henkin, Monk, Tarski
Cylindric Algebras

Information cannot only described by propositional logic, but also by predicate logic, which is more powerful. In this chapter, the language of predicate logic will be briefly introduced. In predicate logic, one can assert a property of an individual by stating the property and giving the individual's name as an argument. This is a clear advantage over propositional logic, since the individual concerned by a proposition, i. e. the argument, is distinguished from the property asserted. Furthermore, one can ascribe one property to different individuals or formulate generalizations concerning several individuals.

After an informal introduction to those elements of predicate logic, which are a clear extension of propositional logic (Section 9.1), a formalization of their syntactic characteristics is given in Section 9.2. Note that from the syntactic point of view, predicate logic is a language for information representation. However, when meaning is attached to the language, one considers the semantics of predicate logic. The semantic details are explained in Section 9.3. In Section 9.4, it turns out that it is possible to switch between both levels, the level of language for representing information (syntactic aspects) and the level of meaning, for touching the information content (semantic aspects), as there is a one-to-one correspondence between both points of view. This correspondence is established by the concept of quantifier algebras, which points out the algebraic properties of predicate logic. In Section 9.5, the formalism of quantifier algebras is taken up and it is shown that it satisfies the information algebra axioms. Therefore, predicate logic also forms an information algebra, on the semantic, as well on the syntactic level. Finally, based on Chapter 7, a measure of predicate logic information is proposed in Section 9.6.

9.1 Extending Elements

Objects or individuals are labeled or named by so-called (individual) constants. As an example, take Tim and Tom, two constants that act as labels for two people. Another example are the first and the second natural number, which are named by 1 and 2, respectively.

Furthermore, there are different sorts of individuals, belonging to different classes. For instance, Tim and Tom are of the sort "human being"; 1 and 2 belong to the class "natural number". Since variables stand for a certain individual, the type or sort of each variable is important and has to be determined.

The properties or relationships one can attribute to individuals are predicate symbols, or predicates, for short. Predicates return a Boolean value, which is either true or false, meaning that the individuals do fulfill the property or not.

Each predicate provides information about a fixed set of individuals. For each place in the predicate, it is specified of which sort the argument (individual) has to be. As an example, consider a predicate $Taller$, which has arity two, so it takes exactly two arguments. Furthermore, both places are of the sort "human being". So, $Taller(Tim, Tom)$ tells us that Tim is taller than Tom.

Generalizations can be formulated by means of so-called quantifiers. The existential quantifier $\exists x$, read "exists x", tells that a certain fact holds for at least one value of a variable. The universal quantifier $\forall x$, read "for all x", states something for every value of a variable. Typically, $\forall x Taller(x, Tim)$ says that everyone is taller than Tim and $\exists x Taller(x, Tom)$ signalizes that there is at least someone who is taller than Tom.

In the following, we will only use the existential quantifier $\exists x$, since the universal quantifier $\forall$ can be rewritten by means of $\exists$ and $\neg$, as shown at the end of Section 9.2. Furthermore, note that we are not considering any function symbols. That is why we are not talking of first order logic, but of predicate logic, only dealing with predicates.

9.2 Syntax

Predicate logic is a formal language to represent information. The syntax describes how to constitute well-formed statements. In contrast to propositional logic, presented in Chapter 8, there are different predicate logic languages. The difference lies in the arity of predicates of the language, as well as in the sorts of its variables and constants. However, all predicate logic languages share the same alphabet and the same rules of assembling the symbols of the alphabet.

Statements, given in the language of predicate logic, are formed by dint of the following symbols:

Definition 9.1 (Alphabet)

An alphabet $\mathcal{A}$ *consists of*

connective symbols	$\neg, \wedge$,
a quantifier	$\exists$,
predicates	$P_i, i \in I \subseteq \mathbb{N}$,
variables	$v_j, j \in J \subseteq \mathbb{N}$,
constants	$c_k, k \in K \subseteq \mathbb{N}$,
logical constants	$\top, \bot$,
auxiliary symbols	*parentheses, commas.*

The index sets I (for predicates), J (for variables) and K (for constants) have to be of the right size to name the predicates, variables and constants needed. These sets can be finite or countably infinite. The set of all variables of a predicate language $\mathcal{L}$ (see Definition 9.2 below) is denoted by $Vbl(\mathcal{L})$.

Each variable or constant is of a specific sort. $St = \{st_s : s \in \mathbb{N}\}$ is the set of all sorts st_s, the variables or constants of a predicate language $\mathcal{L}$ can belong to. St and $J \cup K$ may have the same size, but often they do not, since several variables or constants may belong to the same sort. In order to determine the sort of a variable or a constant, one has to dispose of two functions $\sigma_v : J \to St$ and $\sigma_c : K \to St$, determining a partition of J or K, respectively, into subsets of the same sort. In other words, σ_v specifies for every variable v_j of the language $\mathcal{L}$ its sort $\sigma_v(j) \in St$; σ_c attributes to each constant c_k its sort $\sigma_c(k) \in St$.

A predicate, finally, has a fixed arity, denoting the number of its arguments, given by the function $\lambda : I \to \mathbb{N}$, where $\lambda(i)$ is the arity of the i-th predicate symbol P_i. For every predicate P_i, the sorts of its arguments have to be specified. This is done by the function $\mu : I \times \mathbb{N} \to St$, where $\mu(i, n)$ gives the type of the n-th argument of P_i and $n \leq \lambda(i)$. Often, predicates are also denoted by upper-case letters P, Q, R or by words starting with capital letters.

The symbols mentioned above in Definition 9.1 are part of the alphabet of every predicate language. The predicate languages differ in their predicate symbols and the sorts of their variables and constants. This leads to the following definition:

Definition 9.2 (Language)

A predicate language $\mathcal{L}$ *is characterized by*

- *the alphabet* $\mathcal{A}$,
- *a set of sorts* $St = \{st_s : s \in \mathbb{N}\}$,
- *an arity function* $\lambda : I \to \mathbb{N}$,
- *three typing functions* $\sigma_v : J \to St$, $\sigma_c : K \to St$ *and* $\mu : I \times \mathbb{N} \to St$,

and is denoted and completely determined by $\mathcal{L} = (\lambda, \mu, \sigma_v, \sigma_c, St)$.

Note that the above definition does not take the sets I, J and K into account. They are the domains of the functions λ, σ_v, σ_c and can therefore be retrieved from them.

Now it is time to determine what different kinds of statements there are and how well-formed ones are built. In a predicate language $\mathcal{L}$, one talks about individuals and their properties. Therefore, strings of symbols from the alphabet, which refer to individuals or which express their properties, are needed. Such a sequence of symbols is called formula.

Definition 9.3 (Formula)

Let $\mathcal{L} = (\lambda, \mu, \sigma_v, \sigma_c, St)$ be a predicate language. Then a finite string ϕ of symbols from $\mathcal{L}$ is a formula *if either*

1. *ϕ is $\top$ or $\bot$,*
2. *ϕ is $P_i(a_1, a_2, \ldots, a_{\lambda(i)})$, for variables a_n of type $\mu(i,n)$, $1 \leq n \leq \lambda(i)$, $i \in I$,*
3. *ϕ is $\neg\alpha$, for a formula α,*
4. *ϕ is $\alpha \wedge \beta$, for formulae α and β,*
5. *ϕ is $\exists v_j \alpha$, for a formula α and a variable $v_j, j \in J$.*

Formulae that satisfy rule (1) or rule (2) are called atoms or atomic formulae. Those formulae which satisfy rule (5) are called quantified formulae. Sets of formulae are denoted by Greek upper-case letters, such as Φ, Ψ. The set of all formulae of a language $\mathcal{L}$ is denoted by $Fml(\mathcal{L})$.

The variables in quantified formulae are of special interest. They can have different properties:

Definition 9.4 (Free Variable, Bound Variable)

Let v_j be a variable and ϕ a formula. Then v_j is called free in *ϕ if v_j is among the symbols constituting ϕ, but $\exists v_j$ is not among them. Those variables which are not free in ϕ are called* bound*.*

The set of all variables free in ϕ is denoted by $Fr(\phi) := \{v : v \text{ is free in } \phi\}$.

There are formulae which contain only bound variables:

Definition 9.5 (Sentence)

A formula is called a sentence *if it does not contain any free variables.*

Note that logicians are mainly interested in sentences. Techniques like theorem proving or consequence finding are based on the fact that sentences do not contain any free variables. However, sentences are not of special importance for our information theoretical purposes, since our interest lies in the free variables and the values which can be assigned to them, in order to fulfill some formula (see Section 9.3).

For increasing the legibility of formulae, we introduce "derived" connectives $\vee, \rightarrow, \leftrightarrow$ and the quantifier $\forall$ for formulae ϕ, ψ into the alphabet of predicate language. They can be put down to the known symbols $\neg, \wedge$ and $\exists$ by the following substitution rules:

$\phi \vee \psi$	for	$\neg(\phi \wedge \psi)$,
$\phi \rightarrow \psi$	for	$\neg\phi \vee \psi$,
$\phi \leftrightarrow \psi$	for	$(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$,
$\forall v_j \phi$	for	$\neg(\exists v_j \neg\phi)$.

9.3 Semantics

We are dealing with statements, i. e. with formulae that can be either true or false. The value "false" will be represented by 0 and the value "true" by 1. Clearly, the assignment of a truth value out of $\{0, 1\}$ depends on the objects the variables in a formula refer to. These objects are the constants $c_k, k \in K$, of the alphabet $\mathcal{A}$ and include all individuals one wants to talk about. Every such individual is of a certain sort, i. e. belongs to a class of objects, which will also be called a frame. Furthermore, for every predicate $P_i, i \in I$, it has to be specified which individuals fulfill which properties. This leads to the formal definition of a structure:

Definition 9.6 (Structure)

Let $\mathcal{L} = (\lambda, \mu, \sigma_v, \sigma_c, St)$ be a predicate language. A structure *$\Sigma = \{F, R\}$ over $\mathcal{L}$ consists of*

1. *a set F of frames. For every sort $st_s \in St$, one can determine a frame, i. e. a set $\mathfrak{D}_{st_s} = \{c_k : \sigma_c(k) = st_s,\ \text{with } k \in K\}$. $F := \{\mathfrak{D}_{st_s} : st_s \in St\}$,*

2. *a set R of relations. For every $i \in I$, one disposes of a $\lambda(i)$-ary relation R_i, according to the typing function μ which gives the type of the n-th argument of P_i: $R_i \subseteq \mathfrak{D}_{\mu(i,1)} \times \ldots \times \mathfrak{D}_{\mu(i,\lambda(i))}$. R_i is called the* interpretation *of P_i. $R := \{R_i : i \in I\}$.*

Note that for a variable $v_j, j \in J$, or a constant $c_k, k \in K$, one can retrieve its sort by $\sigma_v(j)$, or $\sigma_c(k)$, respectively, resulting in some specific $st_s \in St$. It is therefore convenient to write $\mathfrak{D}_{v_j}$ or $\mathfrak{D}_{c_k}$, instead of $\mathfrak{D}_{\sigma_v(j)} = \mathfrak{D}_{st_s}$ or $\mathfrak{D}_{\sigma_c(k)} = \mathfrak{D}_{st_s}$.

What still needs to be determined is the value of the variables in a formula. For that purpose, a function h_Σ is provided. h_Σ is a function from the set of variables to the set of elements of all frames. This function assigns an element of the variable's frame to a variable.

Definition 9.7 (Variable Assignment Function)

Let $\mathcal{L}$ be a predicate language and Σ a structure over $\mathcal{L}$. A function

$$h_\Sigma : Vbl(\mathcal{L}) \to \bigcup_{st_s \in St} \mathfrak{D}_{st_s}$$

is called a variable assignment function *into Σ if h_Σ assigns to a variable v_j an element $h_\Sigma(v_j) \in \mathfrak{D}_{v_j} = \mathfrak{D}_{\sigma_v(j)}$.*

Denote by $\mathfrak{D}$ the Cartesian product of the frames of all variables $v_j \in Vbl(\mathcal{L})$ considered: $\mathfrak{D} = \mathfrak{D}_{v_1} \times \mathfrak{D}_{v_2} \times \ldots$.

When ascribing a value to each variable $v_j \in Vbl(\mathcal{L})$, this sequence of values is also referred to as a valuation:

Definition 9.8 (Valuation)

Let $\mathcal{L}$ be a predicate language, Σ a structure and h_Σ a variable assignment function into Σ for each variable $v_j \in Vbl(\mathcal{L})$. Then, a sequence ω of values $\langle h_\Sigma(v_1)\, h_\Sigma(v_2) \ldots \rangle$ is called a valuation in Σ. *Clearly, $\omega \in \mathfrak{D}$, the set of all valuations. The values $h_\Sigma(v_1), h_\Sigma(v_2), \ldots$ are also denoted as $\omega_1, \omega_2, \ldots$ and therefore $\omega = \langle \omega_1\, \omega_2 \ldots \rangle$.*

It will be interesting to change the value of such a valuation ω in a single argument. The resulting valuation assigns the value a to v_i and is the same as ω for all variables v_j different from v_i:

Definition 9.9 (i-Modification)

Let Σ be a structure, ω a valuation in Σ, $v_i \in Vbl(\mathcal{L})$ and $a \in \mathfrak{D}_{v_i}$. Then, the valuation

$$mod\binom{v_i}{a}(\omega) \quad = \quad \langle \omega_1\, \omega_2 \ldots \omega_{i-1}\, a\, \omega_{i+1} \ldots \rangle$$

*is called an i-*modification *of the valuation ω.*

There are cases where one only wants to consider some of the values of the variables in $Vbl(\mathcal{L})$. The values of interest have to be extracted from a valuation ω. The result is a valuation, projected on the set $x \subset Vbl(\mathcal{L})$ of interesting variables, denoted by $\omega^{\Rightarrow x}$:

Definition 9.10 (Valuation Projection)

Let $x \subset Vbl(\mathcal{L})$ be a finite set of variables, where $x = \{v_{x_1}, v_{x_2}, \ldots, v_{x_n}\}, x_j \in J$, and ω a valuation of $\mathfrak{D}$. The valuation projected *on x is*

$$\omega^{\Rightarrow x} \quad = \quad \langle \omega_{x_1}, \omega_{x_2}, \ldots, \omega_{x_n} \rangle.$$

From a theoretical point of view, the above projection set x may also be infinite (e. g. x containing all variables of $Vbl(\mathcal{L})$ which have an odd index j, where $Vbl(\mathcal{L})$ is countable). But in practice, this is not feasible, since projected valuations are mainly used for verifying whether the values specified fulfill a certain predicate (see Definition 9.11 below), which is impossible for an infinite sequence of values. This is why we define the projection of a valuation only for finite sets.

Valuations are used to assign a truth value $\tilde{h}_\Sigma(\omega, \phi) \in \{0, 1\}$ to each formula $\phi \in Fml(\mathcal{L})$. This truth value is inductively defined as follows:

Definition 9.11 (Truth of a Formula)

Let $\mathcal{L}$ be a predicate language, ϕ a formula, Σ a structure and ω a valuation in Σ. Then, $\tilde{h}_\Sigma(\omega, \phi) = 1$ if, and only if,

1. $\phi = \top$,
2. $\phi = P_i(a_1, \ldots, a_{\lambda(i)})$, *for variables* $a_n, 1 \leq n \leq \lambda(i)$, *and* $\omega^{\Rightarrow\{a_1,\ldots,a_{\lambda(i)}\}} \in R_i$,
3. ϕ *is* $\neg\alpha$ *and* $\tilde{h}_\Sigma(\omega, \alpha) = 0$, *for a formula* α,
4. ϕ *is* $\alpha \wedge \beta$ *and* $\tilde{h}_\Sigma(\omega, \alpha) = \tilde{h}_\Sigma(\omega, \beta) = 1$, *for formulae* α, β,
5. ϕ *is* $\exists v_i\alpha$ *and one of the following cases is given:*
   (a) $\tilde{h}_\Sigma(\omega, \alpha) = 1$ *for a formula* α, *or*
   (b) *there is a constant* $c \in \mathfrak{D}_{v_i}$ *and a valuation* $\theta \in \mathfrak{D}$ *such that* $\theta = mod\binom{v_i}{c}(\omega)$, *where* ω *satisfies (a).*

In all the other cases, $\tilde{h}_\Sigma(\omega, \phi) = 0$.

The above definition shows that the truth of a formula depends on the actual valuation.

Definition 9.12 (Model)

Let $\mathcal{L}$ be a predicate language, ϕ a formula, Σ a structure and ω a valuation in Σ. Then, ω is called a model *of ϕ in Σ if*

$$\tilde{h}_\Sigma(\omega, \phi) = 1.$$

In this case, the valuation ω is said to satisfy *the formula ϕ in Σ, denoted by $\omega \models_\Sigma \phi$.*

For every formula $\phi \in Fml(\mathcal{L})$, there exists a subset of $\mathfrak{D}$ consisting of all the models of ϕ:

Definition 9.13 (Set of Models)

Let $\mathcal{L}$ be a predicate language, ϕ a formula, Σ a structure and ω a valuation in Σ. For each formula, one can determine its set of models*. It consists of all those valuations ω which satisfy ϕ:*

$$\hat{h}_\Sigma(\phi) \quad = \quad \{\omega \in \mathfrak{D} : \omega \models_\Sigma \phi\}.$$

Valuations satisfying a formula ϕ assign values, in particular to its free variables. A valuation $\omega \in \mathfrak{D}, \omega \models_\Sigma \phi$, determines the values of the free variables of ϕ (the variables in $Fr(\phi)$). Every variable which is not in $Fr(\phi)$ can have any value, as long as ϕ is still satisfied. Therefore, we consider a set of models $\hat{h}_\Sigma(\phi)$ as the **information relative to the unknown values of the variables of** $Fr(\phi)$ – the information expressed by the formula ϕ. Each model describes a possible state the world could be in (a "possible world", for short).

Finally, the important definition of logical consequence or the entailment relation, with regard to structures, is given:

Definition 9.14 (Entailment Relation / Logical Consequence)

Let $\mathcal{L}$ be a predicate language, Σ a structure, ϕ and ψ two formulae $\in Fml(\mathcal{L})$. Then, a formula ϕ is said to entail *another formula ψ in a structure Σ, denoted by $\phi \models_\Sigma \psi$ iff $\hat{h}_\Sigma(\phi) \subseteq \hat{h}_\Sigma(\psi)$. In this case, ψ is called a* logical consequence *of ϕ.*

This leads to the definition of logical equivalence: Two formulae are equivalent if they entail each other, i. e. if they have the same set of models:

Definition 9.15 (Logical Equivalence)

Let $\mathcal{L}$ be a predicate language, Σ a structure, ϕ and ψ two formulae $\in Fml(\mathcal{L})$. Then, the formulae ϕ and ψ are said to be logically equivalent *in a structure Σ, denoted by $\phi \equiv_\Sigma \psi$ iff $\phi \models_\Sigma \psi$ and $\psi \models_\Sigma \phi$. In this case, $\hat{h}_\Sigma(\psi) = \hat{h}_\Sigma(\phi)$.*

In computer science, usually, specific situations are considered. One is not so much interested to see whether a formula evaluates to 1 in every possible structure Σ. Instead, a fixed structure is given and one will only work with this one. This is why in the following, the lowered $_\Sigma$ will be omitted, since it is obvious which structure we refer to – there is only one.

9.4 Quantifier Algebras: Algebraization of Predicate Logic

Cylindric (set) algebras (Henkin *et al.*, 1971), quantifier algebras, polyadic or Halmos algebras (Halmos, 1962; Plotkin, 1994) can be associated with the predicate calculus, depicted in the previous sections. These algebras have been introduced for the algebraic study of predicate logic.

The goal of this section is to outline the algebraic properties of predicate logic (Section 9.4.1) and to show that sets of models and formulae both form a so-called quantifier algebra and that these algebras are isomorphic (Sections 9.4.2 and 9.4.3). For each formula $\phi \in Fml(\mathcal{L})$, one can determine the set of its models $\hat{h}(\phi)$. However, note that it is not possible to go back from a set of models to the corresponding formula, since there are many equivalent formulae with the same set of models.

9.4.1 Quantifier Algebras

The concept of a quantifier algebra necessitates the definition of an existential quantifier, see (Halmos, 1962; Plotkin, 1994).[1]

Definition 9.16 (Existential Quantifier)

Suppose that $\mathcal{B}$ is a Boolean algebra and $\phi, \psi \in \mathcal{B}$. An existential quantifier *on $\mathcal{B}$ is a mapping $\exists : \mathcal{B} \rightarrow \mathcal{B}$ subject to the following conditions:*

$$\exists 0 = 0 \tag{9.1}$$

$$\exists \phi \leq \phi \tag{9.2}$$

$$\exists(\phi \vee \exists \psi) = \exists \phi \vee \exists \psi \tag{9.3}$$

For moving on to quantifier algebras, we consider an existential quantifier $\exists(Y)$ for every set $Y \in D$, where D is a lattice of subsets of some set X.

Definition 9.17 (Quantifier Algebra)

Let $\mathcal{B}$ be a Boolean algebra and D a lattice of subsets of some set X. Then, for all $\phi \in \mathcal{B}$ and for all $Y \in D$, there is an existential quantifier $\exists(Y)$, such that

$$\exists(\emptyset)\phi = \phi \tag{9.4}$$

$$\exists(Y_1)(\exists(Y_2)\phi) = \exists(Y_1 \cup Y_2)\phi, \tag{9.5}$$

for $Y_1, Y_2 \in D$. Then, $(\mathcal{B}, D)$ is termed a quantifier algebra *over X.*

See (Henkin *et al.*, 1971; Plotkin, 1994) for more details on quantifier algebras.[2]

Furthermore, a dimension set is associated with each element of such an algebra:

Definition 9.18 (Dimension Set)

The dimension set *of $\phi \in \mathcal{B}$, written $\Delta(\phi)$, is the set of all variables $x \in X$ for which the following inequality holds:*

$$\exists(\{x\})\phi \neq \phi.$$

[1] The original definition has slightly been modified regarding the order. Now, it reflects the order used in this thesis, introduced in Section 3.1. Furthermore, it is in accordance with the Boolean algebra known from Section 4.6, which will be important later in this chapter.

[2] For our purposes, it would actually be sufficient to consider a semilattice in Definitions 9.16 and 9.17, instead of a Boolean algebra. But in (Halmos, 1962; Henkin *et al.*, 1971; Plotkin, 1994), always a Boolean algebra is considered, as the goal is the algebraization of predicate logic.

Performing an existential quantification with one of the variables $x \in \Delta(\phi)$ will change ϕ, whereas an existential quantification with an $x \notin \Delta(\phi)$ will not do this. The below corollary follows directly from Definition 9.18:

Corollary 9.19 *For all $\phi \in \mathcal{B}$, $\exists(Y)\phi = \phi$ if $Y \subseteq (X \setminus \Delta(\phi))$.*

Quantifier algebras containing only elements which can be described by a finite set of variables, i. e. elements with a finite dimension set, will be referred to as locally finite quantifier algebras:

Definition 9.20 (Locally Finite Quantifier Algebra)

A quantifier algebra $(\mathcal{B}, D)$ is called locally finite-dimensional*, or simply* locally finite*, if the dimension set $\Delta(\phi)$ is finite for every $\phi \in \mathcal{B}$.*

For locally finite quantifier algebras, one may assume that D is a lattice of finite subsets of X.

9.4.2 The Quantifier Algebra of Sets of Models

We will now show that predicate logic, considered on the semantic level, forms a quantifier algebra. As seen before, a quantifier algebra consists of a Boolean algebra $\mathcal{B}$ and a lattice D of subsets of some set X. After having specified the lattice D and the Boolean algebra for the case of predicate logic, we will extend the Boolean algebra with an existential quantifier, according to Definition 9.16. Then, we can conclude that the semantics of predicate logic forms a quantifier algebra, as it satisfies Definition 9.17.

The Lattice D

D is the lattice of finite subsets of $Vbl(\mathcal{L})$. As D contains only of finite sets, the resulting quantifier algebra will be locally finite (see Definition 9.20).

The Boolean Algebra $\mathfrak{M}$

It is well known that predicate logic is a Boolean algebra. Take $Fml(\mathcal{L})$, the set of all formulae of a predicate language $\mathcal{L}$, and consider the corresponding sets of models for all the predicate logic formulae $\phi \in Fml(\mathcal{L})$. This gives rise to the family $\mathfrak{M}$ of sets of models $\hat{h}(\phi)$:

$$\mathfrak{M} \quad = \quad \{\hat{h}(\phi) \subseteq \mathfrak{D} : \phi \in Fml(\mathcal{L})\}. \tag{9.6}$$

Recall that $\mathfrak{D}$ is the Cartesian product $\mathfrak{D}_{v_1} \times \mathfrak{D}_{v_2} \times \ldots$ of the frames $\mathfrak{D}_{v_j}$ of all variables. $\mathfrak{M}$ is constituted of sets of models, according to Definition 9.13. So far we have specified the elements of the Boolean algebra, they are sets of models. Furthermore, the following definitions are needed:

- The bottom element, $0 \in \mathfrak{M}$, is given by the empty set $\emptyset$.
- The top element, $1 \in \mathfrak{M}$, is given by the whole set $\mathfrak{D}$.
- The join operator $\vee$ is given by set intersection $\cap$.
- The meet operator $\wedge$ is given by set union $\cup$.
- The complement operator ^{c} is given by the set complement ^{c}.

Note that we use the reverse order of that usually considered in Boolean algebra, as mentioned in Footnote 1 on page 131. This is why the definitions of the bottom and the top element are also inverted. As to the set complement,

$$(\hat{h}(\phi))^c \quad = \quad \mathfrak{D} \setminus \hat{h}(\phi) \tag{9.7}$$

is possible as by the nature of D, we are only considering finite sets of variables, even if the set of all variables may be countable. From (Henkin *et al.*, 1971) it is known that the family $\mathfrak{M}$ of sets of models of all formulae of $Fml(\mathcal{L})$ is a Boolean algebra. The join (meet) of two sets of models corresponds to the conjunction (disjunction) of the formulae which gave rise to these two sets of models. In the same way, the complement of a set of models corresponds to the negation of the formula which gave rise to the set of models. See also Figure 9.2 on page 141. Therefore, an application of the above operations results again in an element of $\mathfrak{M}$.

Extension with Existential Quantifier

$\mathfrak{M}$ is now to be extended with a further operation, namely the existential quantifier $\exists$, given for sets $\hat{h}(\phi) \in \mathfrak{M}$ by an operation which is called cylindrification:

Definition 9.21 (Cylindrification)

For a predicate logic set of models $\hat{h}(\phi) \in \mathfrak{M}$ and a predicate logic variable $v_i \in Vbl(\mathcal{L})$, the operation of cylindrification *is denoted by $\exists$ and defined by i-modifying the valuations which constitute $\hat{h}(\phi)$:*

$$\exists v_i \hat{h}(\phi) \quad := \quad \{mod\binom{v_i}{a}(\omega) : \omega \in \hat{h}(\phi), a \in \mathfrak{D}_{v_i}\}.$$

Figure 9.1 below illustrates cylindrification, as described by the foregoing definition. The set $\hat{h}(\phi)$ is depicted by the set having the shape of a boomerang, the constituting valuations $\omega \in \hat{h}(\phi)$ are given by the three shaded nodes. Cylindrification, or existential quantification, relative to v_2 consists of the union of all those valuations which have the same v_1-value as a $\omega \in \hat{h}(\phi)$, and an arbitrary v_2-value. They form the cylinder representing the set $\hat{h}(\exists v_2 \phi)$.

According to (Plotkin, 1994), the quantifier of Definition 9.21 is an existential quantifier on the Boolean algebra $\mathfrak{M}$, in the sense of quantifier algebras. However, in Definition 9.17, $\exists$ is applied to sets, not to variables. An extension to finite sets $L \subseteq Vbl(\mathcal{L})$ has to be defined.

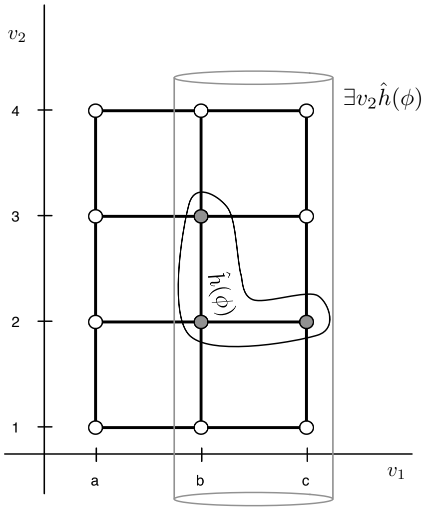


Figure 9.1: Existential quantification is realized by cylindrification

Definition 9.22 (Cylindrification, relative to a set)

For a predicate logic set of models $\hat{h}(\phi) \in \mathfrak{M}$ and a finite set $L \subseteq Vbl(\mathcal{L})$ of predicate logic variables, cylindrification *is denoted by $\exists(L)$, and defined by i-modifying the valuations which constitute $\hat{h}(\phi)$:*

$$\exists(L)\hat{h}(\phi) \quad := \quad \{mod\binom{v_i}{a}(\omega) : \omega \in \hat{h}(\phi), v_i \in L, a \in \mathfrak{D}_{v_i}\}.$$

The above definition is actually an extension of Definition 9.21, as the i-modification of the valuations in $\hat{h}(\phi)$ is carried out on all variables in L, not only on a single one. For each $v_i \in L$, all possible values $a \in \mathfrak{D}_{v_i}$ have to be considered.

The Quantifier Algebra of Semantic Predicate Logic

The Boolean algebra $\mathfrak{M}$ could be extended with an existential quantifier satisfying Definition 9.16. This allows us to state that, according to Definition 9.17, $(\mathfrak{M}, D)$ is a quantifier algebra, where

- $\mathfrak{M}$ is a Boolean algebra
  - whose elements $\hat{h}(\phi) \subseteq \mathfrak{D}$ are sets of models for formulae $\phi \in Fml(\mathcal{L})$,
  - with bottom element $0 = \emptyset$ and top element $1 = \mathfrak{D}$,
  - with join being the set intersection $\cap$, meet being the set union $\cup$ and complement being the set complement ^{c},
- D is the lattice of finite subsets of $Vbl(\mathcal{L})$, and, finally,
- the operation of existential quantification is given by cylindrification as specified in Definition 9.22.

9.4.3 The Quantifier Algebra of Formulae

Now that we know that the sets of models of predicate logic indeed form a quantifier algebra, we link them to formulae. This allows us to show that predicate logic forms a quantifier algebra on the syntactic level as well. As already seen in the case of propositional logic, syntactically different formulae can have the same meaning, i. e. share the same set of models. Again, a set of models does not correspond to a single formula $\phi \in Fml(\mathcal{L})$, but to a set of equivalent formulae. This gives rise to a congruence relation ϑ on the Boolean algebra $Fml(\mathcal{L})$, compatible with the Boolean algebra operations and with $\exists$. Then, a new algebra, the quotient algebra $Fml(\mathcal{L})/\vartheta$, is established, with the Boolean algebra operations and $\exists$ between equivalence classes. This leads to a quantifier algebra of sets of equivalent formulae.

Equivalent Formulae

The process sketched above will be realized by referring to the semantics of predicate logic. We are interested in the meaning of the information conveyed by a formula, not in its syntactic form. Therefore, it is most natural to link formulae to their models and thereby determine under which circumstances two formulae are equivalent.

Lemma 9.23 *Two formulae ϕ and ψ of $Fml(\mathcal{L})$ are* equivalent *if, and only if, they have the same set of models:*

$$\phi \equiv \psi \pmod{\vartheta} \quad \textit{iff} \quad \hat{h}(\phi) = \hat{h}(\psi).$$

Proof. As the following properties are satisfied for any $\phi, \psi, \zeta \in Fml(\mathcal{L})$, ϑ is actually an equivalence relation:

1. *Reflexivity:* $\phi \equiv \phi \pmod{\vartheta}$, since $\hat{h}(\phi) = \hat{h}(\phi)$.
2. *Symmetry:* $\phi \equiv \psi \pmod{\vartheta}$ implies $\psi \equiv \phi \pmod{\vartheta}$, since $\hat{h}(\phi) = \hat{h}(\psi)$.
3. *Transitivity:* $\phi \equiv \psi \pmod{\vartheta}$ and $\psi \equiv \zeta \pmod{\vartheta}$ imply $\phi \equiv \zeta \pmod{\vartheta}$, since $\hat{h}(\phi) = \hat{h}(\psi) = \hat{h}(\zeta)$.

□

Operations in the Semantic Point of View

Based on the above equivalence relation ϑ, we want to establish a congruence relation which is compatible with the Boolean algebra operations and with existential quantification. Before doing this, we first need to introduce some Lemmata on the operations which are involved in the congruence proof below. As the equivalence relation ϑ is expressed on the semantic level of predicate logic, we have to show how the operations on $Fml(\mathcal{L})$ are expressed on the semantic side. Each of the following lemmata is concerned with one of the following operations: $\wedge, \neg, \vee, \exists$. These operations are defined between formulae of $Fml(\mathcal{L}_P)$. The first three of them correspond in the following way to the operations of a Boolean algebra $\mathfrak{B}$:

$Fml(\mathcal{L}_P)$	conjunction $\wedge$	disjunction $\vee$	negation $\neg$
$\mathfrak{B}$	join $\vee$	meet $\wedge$	complement ^{c}

Depending on the arity of the operation, it is applied to one or two formulae $\phi, \psi \in Fml(\mathcal{L})$. Then, the set of models of the resulting formula is taken. By application of some of the definitions given in Sections 9.3 and 9.4.2 and of the de Morgan laws, the syntactic operations are linked with set operations on the semantic level of predicate logic.

The conjunction of two formulae corresponds to the intersection of their sets of models:

Lemma 9.24 *The set of models of the conjunction of two formulae $\phi, \psi \in Fml(\mathcal{L})$ is*

$$\hat{h}(\phi \wedge \psi) \quad = \quad \hat{h}(\phi) \cap \hat{h}(\psi).$$

Proof: The following lines are based on the definitions of Section 9.3.

$$\begin{aligned} \hat{h}(\phi \wedge \psi) &\overset{\text{Def. 9.11(4)}}{=} \{\omega \in \mathfrak{D} : \tilde{h}(\omega, \phi) = \tilde{h}(\omega, \psi) = 1\} \\ &\overset{\text{Def. 9.12}}{=} \{\omega \in \mathfrak{D} : \omega \models \phi\} \cap \{\omega \in \mathfrak{D} : \omega \models \psi\} \\ &\overset{\text{Def. 9.13}}{=} \hat{h}(\phi) \cap \hat{h}(\psi). \end{aligned}$$

□

The negation of a formula corresponds to the complement of the set of its models:

Lemma 9.25 *The set of models of the negation of a formula $\phi \in Fml(\mathcal{L})$ is*

$$\hat{h}(\neg\phi) \quad = \quad (\hat{h}(\phi))^c,$$

where ^{c} is the set complement operator.

Proof: The following lines are based on the definitions of Sections 9.3 and 9.4.2.

$$\begin{aligned} \hat{h}(\neg\phi) &\overset{\text{Def. 9.11(3)}}{=} \{\omega \in \mathfrak{D} : \tilde{h}(\omega, \phi) = 0\} \\ &= \mathfrak{D} \setminus \{\omega \in \mathfrak{D} : \tilde{h}(\omega, \phi) = 1\} \\ &\overset{\text{Def. 9.13}}{=} \mathfrak{D} \setminus \hat{h}(\phi) \\ &\overset{\text{Eq. 9.7}}{=} (\hat{h}(\phi))^c. \end{aligned}$$

□

The disjunction of two formulae corresponds to the union of their sets of models:

Lemma 9.26 *The set of models of the disjunction of two formulae $\phi, \psi \in Fml(\mathcal{L})$ is*

$$\hat{h}(\phi \vee \psi) \quad = \quad \hat{h}(\phi) \cup \hat{h}(\psi).$$

Proof: The following lines are based on the de Morgan laws.

$$\begin{aligned}
\hat{h}(\phi \vee \psi) &= \hat{h}(\neg(\neg\phi \wedge \neg\psi)) \\
&= (\hat{h}(\neg\phi \wedge \neg\psi))^c \\
&= (\hat{h}(\neg\phi) \cap \hat{h}(\neg\psi))^c \\
&= ((\hat{h}(\phi))^c \cap (\hat{h}(\psi))^c)^c \\
&= \hat{h}(\phi) \cup \hat{h}(\psi).
\end{aligned}$$

□

The existential quantification of a formula over one variable corresponds to the cylindrification of the set of its models:

Lemma 9.27 *The set of models of an existential quantified formula $\phi \in Fml(\mathcal{L})$, over some variable $v_i \in Vbl(\mathcal{L})$, is*

$$\hat{h}(\exists v_i \phi) \quad = \quad \exists v_i \hat{h}(\phi).$$

Proof: The following lines are based on the definitions of Sections 9.3 and 9.4.2.

$$\begin{aligned}
\hat{h}(\exists v_i \phi) &\overset{\text{Def. 9.11(5)}}{=} \left\{mod\binom{v_i}{a}(\omega) : a \in \mathfrak{D}_{v_i}, \omega \in \mathfrak{D}, \omega \models \phi\right\} \\
&= \left\{mod\binom{v_i}{a}(\omega) : \omega \in \hat{h}(\phi), a \in \mathfrak{D}_{v_i}\right\} \\
&\overset{\text{Def. 9.21}}{=} \exists v_i \hat{h}(\phi).
\end{aligned}$$

□

The existential quantification of a formula over a set of variables corresponds to the cylindrification of the set of its models:

Lemma 9.28 *The set of models of an existentially quantified formula $\phi \in Fml(\mathcal{L})$, over a set $L \subseteq Vbl(\mathcal{L})$ of variables is*

$$\hat{h}(\exists(L)\phi) \quad = \quad \exists(L)\hat{h}(\phi).$$

Proof: This lemma is proven by Lemma 9.27 above and the fact that $\exists(L)\phi$ is independent of the order of quantification, where

$$\exists(L)\phi \quad \equiv \quad \exists v_1(\ldots(\exists v_l \phi)\ldots), \quad \text{for } L = \{v_1, \ldots, v_l\} \subseteq Vbl(\mathcal{L}).$$

□

Note furthermore that the two formulae $\top$ (tautology) and $\perp$ (contradiction) correspond to the two extreme sets of models:

Lemma 9.29 *The set of models of* $\top \in Fml(\mathcal{L})$ *and* $\bot \in Fml(\mathcal{L})$ *are*

$$\begin{aligned} \hat{h}(\top) &= \mathfrak{D} \textit{ and} \\ \hat{h}(\bot) &= \emptyset, \textit{ respectively.} \end{aligned}$$

Proof: Based on the definitions of Section 9.3 and on the de Morgan laws,

$$\begin{aligned} \hat{h}(\top) &\overset{\text{Def. 9.11(1)}}{=} \{\omega \in \mathfrak{D} : \tilde{h}(\omega, \top) = 1\} \\ &= \mathfrak{D}, \end{aligned}$$

$$\begin{aligned} \hat{h}(\bot) &= \hat{h}(\neg\top) \\ &\overset{\text{Def. 9.11(3)}}{=} \{\omega \in \mathfrak{D} : \tilde{h}(\omega, \top) = 0\} \\ &= \emptyset. \end{aligned}$$

□

Congruence Relation and Quotient Algebra

A congruence relation is an equivalence relation that is compatible with some algebraic operations. Here, we are considering the equivalence relation ϑ (see Lemma 9.23) on the Boolean algebra $Fml(\mathcal{L})$ with the operations $\wedge, \vee, \neg, \exists$. The congruence relation gives rise to a quotient algebra $Fml(\mathcal{L})/\vartheta$ with the operations $\wedge, \vee, \neg, \exists$ between equivalence classes of the quotient algebra.

Based on the lemmata introduced before, it is shown that the equivalence relation ϑ on $Fml(\mathcal{L}_P)$ respects $\wedge, \vee, \neg$ and $\exists$.

Lemma 9.30 *The equivalence relation* ϑ *of Lemma 9.23 is a congruence relation, as it is compatible with the operations* $\wedge, \vee, \neg$ *and* $\exists$.

Proof. For any formulae $\phi_1, \phi_2, \psi_1, \psi_2 \in Fml(\mathcal{L})$ and any set of variables $L \subseteq Vbl(\mathcal{L})$, $\phi_1 \equiv \phi_2 \pmod{\vartheta}$ and $\psi_1 \equiv \psi_2 \pmod{\vartheta}$ imply

- $\phi_1 \wedge \psi_1 \equiv \phi_2 \wedge \psi_2 \pmod{\vartheta}$, as by Lemmata 9.24 and 9.23 it holds that
  $\hat{h}(\phi_1 \wedge \psi_1) = \hat{h}(\phi_1) \cap \hat{h}(\psi_1) = \hat{h}(\phi_2) \cap \hat{h}(\psi_2) = \hat{h}(\phi_2 \wedge \psi_2)$.
- $\phi_1 \vee \psi_1 \equiv \phi_2 \vee \psi_2 \pmod{\vartheta}$, as by Lemmata 9.26 and 9.23 it holds that
  $\hat{h}(\phi_1 \vee \psi_1) = \hat{h}(\phi_1) \cup \hat{h}(\psi_1) = \hat{h}(\phi_2) \cup \hat{h}(\psi_2) = \hat{h}(\phi_2 \vee \psi_2)$.
- $\neg\phi_1 \equiv \neg\phi_2 \pmod{\vartheta}$, as by Lemmata 9.25 and 9.23 it holds that
  $\hat{h}(\neg\phi_1) = (\hat{h}(\phi_1))^c = (\hat{h}(\phi_2))^c = \hat{h}(\neg\phi_2)$.
- $\exists(L)\phi_1 \equiv \exists(L)\phi_2 \pmod{\vartheta}$, as by Lemmata 9.28 and 9.23 it holds that
  $\hat{h}(\exists(L)\phi_1) = \exists(L)\hat{h}(\phi_1) = \exists(L)\hat{h}(\phi_2) = \hat{h}(\exists(L)\phi_2)$. □

The congruence relation ϑ gives rise to a partition of $Fml(\mathcal{L})$ into non-empty disjoint subsets, called equivalence classes. The equivalence class of a formula $\phi \in Fml(\mathcal{L})$ is of the form

$$[\phi]_\vartheta \quad = \quad \{\psi \in Fml(\mathcal{L}) : \psi \equiv \phi \pmod{\vartheta}\}, \tag{9.8}$$

i. e. it contains all formulae which have the same set of models as ϕ.

Since ϑ has been shown to be a congruence relation on the Boolean algebra $Fml(\mathcal{L})$, compatible with the Boolean operations $\wedge, \vee$ and $\neg$, and also $\exists$, we can now establish a new algebra $Fml(\mathcal{L})/\vartheta$. Its elements are the equivalence classes mentioned in Equation 9.8 above, and it is called quotient algebra of $Fml(\mathcal{L})$ over ϑ. For any n-ary operation on $Fml(\mathcal{L})$, for which ϑ has been shown to be compatible, a well-defined n-ary operation on $Fml(\mathcal{L})/\vartheta$ is given by construction.

The quotient algebra of $Fml(\mathcal{L})$ over ϑ is given by the set

$$Fml(\mathcal{L})/\vartheta \quad = \quad \{[\phi]_\vartheta : \phi \in Fml(\mathcal{L})\}. \tag{9.9}$$

The operations of $\wedge, \vee, \neg$ and $\exists$ on $Fml(\mathcal{L})/\vartheta$ are defined for the equivalence classes. For $[\phi]_\vartheta, [\psi]_\vartheta \in Fml(\mathcal{L})/\vartheta$ and $L \subseteq Vbl(\mathcal{L})$, the following definitions are given:

$$[\phi]_\vartheta \wedge [\psi]_\vartheta \quad := \quad [\phi \wedge \psi]_\vartheta, \tag{9.10}$$

$$[\phi]_\vartheta \vee [\psi]_\vartheta \quad := \quad [\phi \vee \psi]_\vartheta, \tag{9.11}$$

$$\neg[\phi]_\vartheta \quad := \quad [\neg\phi]_\vartheta, \tag{9.12}$$

$$\exists(L)[\phi]_\vartheta \quad := \quad [\exists(L)\phi]_\vartheta. \tag{9.13}$$

When such an operation is applied, the result does not depend on the formula which has been selected as the representative of the equivalence class. It can be shown that $Fml(\mathcal{L})/\vartheta$ and the three operations of Equations 9.10, 9.11 and 9.12 define a Boolean algebra, where Equation 9.10 corresponds to the join, Equation 9.11 to the meet and Equation 9.12 to the complement operation of the Boolean algebra.

The Quantifier Algebra of Syntactic Predicate Logic

Once again with (Plotkin, 1994), we have a Boolean algebra $Fml(\mathcal{L})/\vartheta$. Further, the existential quantifier $\exists$ satisfies Definition 9.16 by construction. Therefore, $(Fml(\mathcal{L})/\vartheta, D)$ is indeed a quantifier algebra, according to Definition 9.17, where

- $Fml(\mathcal{L})/\vartheta$ is a Boolean algebra
  - whose elements $[\phi]_\vartheta$, $\phi \in Fml(\mathcal{L})$, are equivalence classes of formulae,
  - with bottom element $[\bot]_\vartheta$ and top element $[\top]_\vartheta$,
  - with join being the conjunction $\wedge$ of Equation 9.10, meet being the disjunction $\vee$ of Equation 9.11 and complement being the negation $\neg$ of Equation 9.12,
- D is the lattice of finite subsets of $Vbl(\mathcal{L})$, and, finally,
- the operation of existential quantification is given by the quantifier applied to equivalence classes, defined in Equation 9.13.

9.4.4 Linking Syntax and Semantics

The congruence relation ϑ on $Fml(\mathcal{L})$ of Lemma 9.30 and the resulting quotient algebra $Fml(\mathcal{L})/\vartheta$ of Equation 9.9 have their origin in the semantics of predicate logic. An equivalence class of formulae is linked to the set of models shared by all the formulae of the equivalence class. The elements of the quantifier algebra $(\mathfrak{M}, D)$ of Section 9.4.2 are precisely these sets of models, which are related with the equivalence classes of the quantifier algebra $(Fml(\mathcal{L})/\vartheta, D)$ of Section 9.4.3. The quantifier algebra $(\mathfrak{M}, D)$ of sets of models is thus reflected in the corresponding quantifier algebra $(Fml(\mathcal{L})/\vartheta, D)$ of equivalence classes of formulae. This allows to treat both quantifier algebras in the same manner and to switch between the syntax and the semantics of predicate logic without any problems.

9.5 Predicate Logic as a Domain-Free Information Algebra

In the following, it will be shown that the formalism of information algebras covers predicate logic. The content of this section is based on (Langel & Kohlas, 2008). From the previous section it is known that predicate logic forms a quantifier algebra on the syntactical, as well as on the semantic level. What needs to be done is to show that a quantifier algebra $(\mathcal{B}, D)$, as given by Definition 9.17, is an information algebra (Ψ, D), as given in Definition 5.4. In order to do this, let the Boolean algebra $\mathcal{B}$ correspond to the set Ψ of the information algebra framework. Every $\phi \in \mathcal{B}$ is therefore a piece of information in the information algebraic view. Furthermore, a lattice of subsets of variables is required in order to form an information algebra. It is provided by the quantifier algebra's lattice D of subsets of some set X. This set X corresponds to the countable set Vbl of all variables considered in an information algebra. Moreover, two operations are required in Definition 5.4. Take the join operation of the Boolean algebra $\mathcal{B}$ as the combination operator of information algebras:

$$\phi \otimes \psi = \phi \vee \psi, \text{ for any } \phi, \psi \in \mathcal{B}. \tag{9.14}$$

Focusing will be expressed by variable elimination of Equation 5.6, which corresponds to existential quantification. Recall that we want to show that quantifier algebras $(\mathcal{B}, D)$, and therefore predicate logic, actually form an information algebra. As a predicate logic formula of $Fml(\mathcal{L})$ is usually finite[3], it is only constituted of a finite set of variables $Y \in D$. Therefore, we will assume a locally finite quantifier algebra $(\mathcal{B}, D)$ with D being constituted of finite subsets of X. This warrants a finite dimension set[4] for all pieces of information and allows us to use the operation of variable elimination instead of the operation of focusing. For $\psi \in \mathcal{B}$ with a finite dimension set $\Delta(\psi)$ and $Z \in D$, focusing can be expressed as existential quantification (i. e. variable elimination):

[3] At least, this is always the case in situations related to computer science.

[4] Called "least support" in an information algebra context and denoted by Δ.

$$\psi^{\Rightarrow Z} = \psi^{-\Delta(\psi)\setminus Z} = \exists(\Delta(\psi) \setminus Z)\psi. \tag{9.15}$$

This is an important equation which will be used again and again in this section. Note that one cannot eliminate an infinite sequence of variables, whereas it is possible to focus on a countably infinite set of variables. Therefore, it is important keep in mind that variable elimination can only be performed on pieces of information ψ that have a finite support (dimension set) $\Delta(\psi) \in D$. [5]

Summing up, Definition 5.4 requires a lattice D of subsets of some set of variables Vbl, a set of pieces of information Ψ and two operations, namely combination (denoted by $\otimes$) and focusing (denoted by an upper $\Rightarrow$). The latter may be replaced by the operation of variable elimination (denoted by an upper $-$). The table given in Figure 9.2 shows the constituting elements of information algebras (IA), and their correspondences in the formalism of quantifier algebras (QA), predicate logic on a semantic level (PL sem) and predicate logic on a syntactic level (PL syn):

<table>
<tr><th>IA</th><th>QA</th><th>PL sem</th><th>PL syn</th></tr>
<tr><td>Vbl</td><td>X</td><td colspan="2">$Vbl(\mathcal{L})$</td></tr>
<tr><td>D
lattice of subsets of Vbl</td><td>D
lattice of subsets of X</td><td colspan="2">[not named]
lattice of subsets of $Vbl(\mathcal{L})$</td></tr>
<tr><td>Ψ</td><td>$\mathcal{B}$</td><td>$\mathfrak{M}$</td><td>$Fml(\mathcal{L})/\vartheta$</td></tr>
<tr><td>$\otimes$
combination</td><td>$\vee$
join of the
Boolean algebra</td><td>$\cap$
set intersection</td><td>$\wedge$
conjunction</td></tr>
<tr><td>$-$ $(\Rightarrow)$
variable elimination
(focusing)</td><td>$\exists$
existential
quantification</td><td>$\exists$
cylindrification
as given in Def. 9.22</td><td>$\exists$
existential
quantification</td></tr>
</table>

Figure 9.2: Information algebra instances related to predicate logic

For the proof of Theorem 9.32, a lemma has to be introduced for quantifier algebras. Its proof can be found in (Halmos, 1962).

Lemma 9.31 *Let $(\mathcal{B}, D)$ be a quantifier algebra. If $\phi \in \mathcal{B}$ and $Z \in D$, then $\Delta(\exists(Z)\phi) = \Delta(\phi) \setminus Z$.*

Now the important theorem that every quantifier algebra is also an information algebra can be stated and proven, making use of Equation 9.15. This theorem implies that predicate logic is an information algebra instance.

Theorem 9.32 *Consider a Boolean algebra $\mathcal{B}$ and a lattice D of finite subsets of a set X of variables. The quantifier algebra $(\mathcal{B}, D)$ forms an information algebra, i. e.*

[5] Theoretically, the set Z we are focusing on can be infinite, as long as we dispose of a finite support of ψ: $\psi^{\Rightarrow Z} = (\psi^{\Rightarrow \Delta(\psi)})^{\Rightarrow Z} = \psi^{\Rightarrow \Delta(\psi) \cap Z}$. Variable elimination is possible since $\Delta(\psi) \cap Z$ is a finite set.

it satisfies the five information algebra axioms of Definition 5.4. Combination and focusing are given by Equations 9.14 and 9.15, respectively.

Proof:

1. The requirements of the *semigroup axiom* are satisfied, as it is well known that the join of any Boolean algebra is a commutative and associative operation. The neutral element is the Boolean algebra's top element 1. Consequently, the null element is the bottom element 0.

2. In order to prove the second axiom (*transitivity of focusing*), Lemma 9.31 is needed. Furthermore, Equation 9.5 is applied.

$$\begin{aligned}
\left(\phi^{\Rightarrow Y}\right)^{\Rightarrow Z} &= \left(\phi^{-(\Delta(\phi)\setminus Y)}\right)^{-\left(\Delta(\phi^{-(\Delta(\phi)\setminus Y)})\setminus Z\right)} \\
&= \exists\left(\Delta(\phi^{-(\Delta(\phi)\setminus Y)}) \setminus Z\right)\left(\exists(\Delta(\phi) \setminus Y)\phi\right) \\
&= \exists\Big(\Delta\big(\exists(\Delta(\phi) \setminus Y)\phi\big) \setminus Z\Big)\Big(\exists(\Delta(\phi) \setminus Y)\phi\Big) \\
&= \exists\Big(\big(\Delta(\exists(\Delta(\phi) \setminus Y)\phi) \setminus Z\big) \cup \big(\Delta(\phi) \setminus Y\big)\Big)\phi \\
&= \exists\Big(\big((\Delta(\phi) \setminus (\Delta(\phi) \setminus Y)) \setminus Z\big) \cup \big(\Delta(\phi) \setminus Y\big)\Big)\phi \\
&= \exists\Big(\big((\Delta(\phi) \cap Y) \setminus Z\big) \cup \big(\Delta(\phi) \setminus Y\big)\Big)\phi \\
&= \exists\big(\Delta(\phi) \setminus Y \cap Z\big)\phi \\
&= \phi^{-(\Delta(\phi)\setminus Y\cap Z)} \\
&= \phi^{\Rightarrow Y\cap Z}.
\end{aligned}$$

3. The proof of the *combination axiom* is based on some important properties of support, collected in (Kohlas, 2003, Lemma 3.6). For proving $(\psi^{\Rightarrow Z} \otimes \phi)^{\Rightarrow Z} = \psi^{\Rightarrow Z} \otimes \phi^{\Rightarrow Z}$, let $A = \Delta(\phi) \cup \Delta(\psi) \cup Z$. Now, the following statements about A can be made:

   (i) A is a support of $\psi^{\Rightarrow Z}$, by (Kohlas, 2003, Lemma 3.6 (6) & (2)).
   (ii) A is a support of ϕ, by (Kohlas, 2003, Lemma 3.6 (6)).
   (iii) A is a support of $\psi^{\Rightarrow Z} \otimes \phi$, by (i), (ii), (Kohlas, 2003, Lemma 3.6 (7)).

   The combination axiom is now expressed by means of existential quantification. In contrast to Equation 9.15, where the least support is involved ($\phi^{\Rightarrow Z} = \exists(\Delta(\phi) \setminus Z)$), we will use the set A, as defined and motivated above. The important step in the third line is justified by Equation 9.3 of Definition 9.16.

$$\begin{aligned}
(\psi^{\Rightarrow Z} \otimes \phi)^{\Rightarrow Z} &= (\psi^{-(A\setminus Z)} \vee \phi)^{-(A\setminus Z)} \\
&= \exists(A \setminus Z)(\phi \vee \exists(A \setminus Z)\psi) \\
&= \exists(A \setminus Z)\phi \vee \exists(A \setminus Z)\psi \\
&= \phi^{\Rightarrow Z} \otimes \psi^{\Rightarrow Z} \\
&= \psi^{\Rightarrow Z} \otimes \phi^{\Rightarrow Z}.
\end{aligned}$$

4. For the proof of the *idempotency axiom*, the focusing operation is rewritten by means of Equation 9.15, Inequality 9.2 of Definition 9.16 is used and the combination operation is substituted by the join of the Boolean algebra, according to Equation 9.14.

$$\begin{aligned}\psi &= \psi \otimes \psi^{\Rightarrow Z} \\ &= \psi \otimes \underbrace{\exists(\Delta(\psi) \setminus Z)\psi}_{\leq \psi} \\ &= \psi \vee \phi, \text{ where } \phi \leq \psi \\ &= \psi.\end{aligned}$$

5. The *support* axiom also holds: By Definition 9.18, there is a dimension set $\Delta(\phi)$ for every $\phi \in \mathcal{B}$. The dimension set is actually the least support of ϕ. So using Equation 9.4 of Definition 9.17, we dispose of a least support $Z \in D$ for any piece of information.

$$\begin{aligned}\phi^{\Rightarrow Z} &= \phi^{-(\Delta(\phi)\setminus Z)} \\ &= \exists(\Delta(\phi) \setminus Z)\phi \\ &= \exists(\Delta(\phi) \setminus \Delta(\phi))\psi \\ &= \exists(\emptyset)\phi \\ &= \phi\end{aligned}$$

Hence, $(\mathcal{B}, D)$ is a domain-free information algebra. □

Quantifier algebras come along with existential quantification, i. e. variable elimination, which is restricted to finite subsets of variables. Therefore, information algebras (defined by focusing) are more general than quantifier algebras. However, when using predicate logic in applications from computer science, e. g. looking at the process of querying a data base, the possibility of focusing on an infinity of variables, which is only provided by information algebras, will never be used.

Now that it has been shown that quantifier algebras are information algebras, we also know that the formalism of information algebras covers predicate logic. Considering predicate logic on a semantic level as an information algebra means that the set Ψ of pieces of information is given by $\mathfrak{M} \subset \mathfrak{P}(\mathfrak{D})$. Recall that $\mathfrak{D}$ is the Cartesian product of the frames of all variables $v_j \in Vbl(\mathcal{L})$ or the set of all valuations. The pieces of information $\phi, \psi \in \mathfrak{M}$ are sets of models. The neutral element e is given by $\mathfrak{D}$ itself, it corresponds to the tautology $(\top)$. The null element z is represented by the empty set $\emptyset$. It is the piece of information resulting from the combination of two contradictory pieces of information and is therefore itself a contradiction $(\perp)$. The support (or dimension set) of a piece of information ψ is constituted by all those variables v_j such that ψ does not form a cylinder parallel to the j^{th} axis. When considering (equivalence classes of) formulae, those variables v_j form a support for ψ which are free in ψ, as well as in every formula belonging to the equivalence class of ψ. If no variable occurs free in ψ, which is the case for $\top$, $\bot$ and sentences, ψ has least support $\emptyset$.

From Section 9.4 it is known that the set Ψ of pieces of information forms a Boolean algebra. So it is obvious that predicate logic even forms a Boolean information algebra. But in Section 4.6, Boolean information algebras have only been considered in the labeled version, as this form is needed for the measure of information. However, the above proof is carried out for the domain-free version. In (Kohlas, 2003, Chapter 6.5.2), a domain-free formulation for Boolean information algebras is provided. This definition requires essentially that Ψ is a Boolean lattice. From Section 9.4.2 we know that this holds, with join being combination (set intersection), meet being set union and the complement operator being the set complement.

9.6 Measure

From Chapter 7 it is known that information can be measured qualitatively and quantitatively. Section 9.6.1 is about the qualitative measure of information, expressed by predicate logic. In Section 9.6.2, a quantitative measure of predicate logic information is proposed.

9.6.1 Qualitative Measure

The qualitative measure of predicate logic corresponds one-to-one to that of propositional logic, provided in Section 8.4.1. The partial order $\leq$ between two pieces of information is given on the syntactic level by entailment of formulae and on the semantic level by set inclusion between sets of models.

9.6.2 Quantitative Measure

The quantitative measure of information, expressed by predicate logic, is also very similar to that of propositional logic information, given in Section 8.4.2. The measurements proposed in Sections 7.4, 7.5 and 7.6 can be applied, as predicate logic forms an atomic composed information algebra. It will be considered in its labeled form (see Section 4.7) and is denoted by (Φ, D). We assume that, for all domains $x \in D$, the set of all atoms $At_x(\Phi)$ of the domain x is finite. Even if we have looked at predicate logic in this chapter only from a domain-free point of view, we now consider the labeled information algebra for the measure of information. This is completely unproblematic, as both approaches are equivalent. In Section 5.3, it has been shown how to construct a labeled information algebra from a domain-free one.

The components and operations of the atomic composed information algebra (Φ, D) are listed below. Following the semantic approach of this thesis, the pieces of information and the operations are provided on the semantic level. For that purpose, we use tuples, which have already been introduced in Chapter 2.

- D: lattice of finite subsets of the set $Vbl(\mathcal{L}) = \{v_1, v_2, v_3, \ldots\}$
- atom: x-tuple $f : x \to \mathfrak{D}_x$, where $x \in D$
  (To each variable $v_j \in x$, a value of its frame $\mathfrak{D}_{v_j}$ is attributed.)
- set of pieces of information: $\Phi = \bigcup_{x \in D} \Phi_x$, where $\Phi_x = \mathfrak{P}(\mathfrak{D}_x)$
- piece of information: $\phi \in \Phi$, such that $\phi = \wedge At(\phi)$
- neutral element: $e_x = \mathfrak{D}_x = At_x(\Phi)$
- null element: $z_x = \emptyset_x$
  (By convention, every $\mathfrak{D}_x$ has its own empty subset $\emptyset_x$.)
- labeling: $d(\phi) = x$ if $\phi \in \Phi_x$
- marginalization: $\phi^{\downarrow x} = \{f[x] : f \in \phi\}$, where
  - $x \subseteq d(\phi)$ and
  - $f[x] = g$, such that $g(v_j) = f(v_j)$, for all $v_j \in x$
- combination: $\phi \otimes \psi = \{f \in \mathfrak{D}_{x \cup y} : f[x] \in \phi, f[y] \in \psi\}$
- vacuous extension: $\phi^{\uparrow x} = \phi \otimes e_x$, where $x \supseteq d(\phi)$
- transport: $\phi^{\to x} = (\phi^{\uparrow d(\phi) \cup x})^{\downarrow x}$, for all $x \in D$

Using the above definitions and operations, the measures of information content proposed in Section 7.4 can directly be applied to predicate logic. The same holds for the measures of information content in the dual algebra, seen in Section 7.5. When probabilities are taken into account (Section 7.6), we also dispose of a probability distribution over the tuples. This leads to a probabilistic choice system and the proposed measure is based on entropy.

9.7 Conclusion

In this chapter, predicate logic has been looked at from different points of view. Predicate logic was presented as a formal language, which allows to express information by means of formulae. The semantic level of predicate logic provides the meaning of the information by a set of models. It could be shown that both aspects of predicate logic give rise to a quantifier algebra, i. e. a Boolean algebra with cylindrification (semantic level) or existential quantification (syntactic level) as further operation. In order to prove that predicate logic forms an information algebra, it could be shown that each quantifier algebra is also an information algebra. As predicate logic forms a quantifier algebra, it also forms an information algebra. Finally, we have explained how to apply the different measures of information content, known from Chapter 7, by giving the specific definitions and operations for predicate logic.

Part III

Semantic Information Theories

10 Carnap and Bar-Hillel's Theory of Semantic Information

Information ist Selektion aus der Alternativmenge eines Möglichkeitsraums.

Helmut F. Spinner
Die Wissensordnung: Ein Leitkonzept für die dritte Grundordnung des Informationszeitalters

One of the first comprehensive semantic information theories was proposed by Rudolf Carnap and Yehoshua Bar-Hillel in the early 1950s. Based on the theory of inductive logic and inductive probability developed by Carnap, he and his postdoc Bar-Hillel published "An Outline of a Theory of Semantic Information", an MIT technical report (Bar-Hillel & Carnap, 1952). In the following years, several journal articles were published (Bar-Hillel, 1952; Bar-Hillel & Carnap, 1953; Bar-Hillel, 1955), pointing out different aspects of this semantic theory of information. Some of them were republished one decade later in (Bar-Hillel, 1964). The authors' theory of semantic information stems from inductive logic (see Section 10.1). It is important to note that their theory of information is not founded on the syntactic form of a logical formula, but on the semantic form of the formula. For probably the first time. the valuations are taken into account. In order to illustrate the theory, Bar-Hillel and Carnap introduce a sample language with a finite number of individual constants and primitive one-place predicates. This simple first-order functional language, serving for explanation purposes, is given in Section 10.2. Thereafter, in Section 10.3, Carnap and Bar-Hillel's view of the concept of information is presented, as well as its properties. An instantiation for the sample language is also given. The concept of information is followed by the concept of amount of information in Section 10.4. Bar-Hillel and Carnap point out that its properties justify different instantiations. The following table gives an overview of the elements of Carnap and Bar-Hillel's semantic theory of information and in which section of this chapter the different elements are presented:

conceptual level	instance ("explicatum")
presystematic concept of semantic information (Section 10.3.1)	*Cont* (Section 10.3.2)
presystematic concept of amount of information (Section 10.4.1)	*cont* (Section 10.4.3) *inf* (Section 10.4.4)

After this introduction in Carnap and Bar-Hillel's semantic theory of information, it is compared to our information algebraic point of view, first regarding the concept of information (Section 10.5) and afterwards regarding the measurement (Section 10.6).

10.1 Inductive Logic

Based on (Hacking, 2001) and the entries in the online Stanford Encyclopedia of Philosophy on inductive logic and interpretations of probability (Hawthorne, Winter 2008; Hájek, Winter 2008), we will give a very brief introduction to inductive logic and inductive probability, as studied by Carnap. The encyclopedia provides a good characterization of inductive logic which is conform to Carnap's view of this subject, as developed in his foundational works (Carnap, 1950; Carnap, 1952):

> "An inductive logic is a system of reasoning that extends deductive logic to less-than-certain inferences. In a valid deductive argument the premises logically entail the conclusion, where such entailment means that the truth of the premises provides a guarantee of the truth of the conclusion. Similarly, in a good inductive argument the premises should provide some degree of support for the conclusion, where such support means that the truth of the premises indicates with some degree of strength that the conclusion is true." (Hawthorne, Winter 2008)

In the late 19th and the early 20th century, important results were achieved in the field of deductive logic. With predicate logic, a rigorous formal system for deductive logic arose which allowed to represent all valid deductive arguments in mathematics and the sciences. This was for the first time a logic in which one can determine the validity of deductive arguments by only considering the syntactic structure of the sentences involved. Inspired by such an important result, some logicians aimed at applying a similar approach to inductive reasoning by extending the deductive entailment relation with the notion of probability. The deductive paradigm was followed very closely and thus attempts were made to specify the inductive probabilities only in terms of the syntactic structure of premise and conclusion. Carnap provided the most systematic study of this kind of inductive or logical probabilities. In his works he points out that the logical interpretation provides a framework for induction, i. e. he veers away from the syntactic structure of the sentences. The foundation for a theory of semantic information was laid.

In (Carnap, 1950), the author considers a monadic predicate logic language which is simple but also powerful enough to illustrate every of his concepts by an example. This language is the topic of Section 10.2. Roughly speaking, Carnap's idea was to look only at a finite number of predicates and a finite number of individuals, so that, for every individual, a finite statement can be made describing its properties in as much detail as the expressive power of the language allows. These descriptions of each individual can be concatenated to the strongest but still consistent statements that can be made of all the individuals of the language. Typically, a strongest description is perceived from a semantic point of view, namely as being a valuation (see Sections 8.2 and 9.3). However, it may also be expressed on the syntactic level by a formula. As such a strongest statement can be made for each state the world can be in, Carnap calls it state-description. An inductive probability is now attributed to every state-description. As each sentence is equivalent to a disjunction of state-descriptions (from a syntactic point of view) or a set of state-descriptions (namely the set of models of the sentence, from a semantic perspective), the probability measure over the set of state-descriptions extends automatically to a measure over all sentences.

10.1.1 Inductive Probability

Probability is involved in Bar-Hillel and Carnap's measure of information, see Section 10.4. It is therefore important to point out that there are different kinds of probabilities[1] and which is the one used in (Bar-Hillel & Carnap, 1952). In his foregoing work, Carnap proposes two ideas of probability, called "probability$_1$" and "probability$_2$". We prefer to call them belief-type probability and frequency-type probability respectively, following Hacking's approach. Belief-type probability is defined in (Carnap, 1950) as follows:

Probability$_1$: the logical concept of probability, degree of confirmation

Degree of confirmation: a quantitative concept representing the degree to which the assumption of the hypothesis h is supported by the evidence e

The definition of the frequency-type probability is given as follows, again cited from (Carnap, 1950):

Probability$_2$: the statistical concept of probability, relative frequency in the long run

Relative frequency of the property M in the class K: the absolute frequency of M in K divided by the cardinal number of K

Let us start with frequency-type probability, Carnap's second concept of probability. Its meaning is easier to capture as "statistical" and "frequency" are part of our

[1] A good overview can be found in (Fine, 1970).

everyday vocabulary and most people are more familiar with this more scholastic definition. Frequency-type probability statements express how the world is. As an example, consider some proposition stating a physical property about some object, such as a coin ("This coin is biased. With a probability of about 0.6 the result will be head.") or a die ("This is a perfect die. The probability of getting an arbitrary number is $\frac{1}{6}$."). The truth of such a proposition has nothing to do with the personal belief of the person who utters the proposition. Moreover, we can do experiments to test a frequency-type probability statement, by repeatedly tossing the coin or dicing. Thereby, reasons are collected which prove or disprove the statement.

But this is not the kind of probability used in (Bar-Hillel & Carnap, 1952). For their theory of semantic information, the authors use what they call inductive probability or logical probability. In (Carnap, 1950) it is called probability$_1$ and belief-type probability in (Hacking, 2001, Chapter 11). When the words "probability" or "probable" are used in a belief-type sense, they are related to the ideas of evidence, confidence, belief and credibility. Typical belief-type statements are about how credible or believable a certain fact is. The degree of confirmation from the above definition of probability$_1$ is a level of how much someone believes or should believe a proposition. Hacking's example is the following statement: "It is probable that the dinosaurs were made extinct by a giant asteroid hitting the earth." Then, he lists more details relative to this statement, namely facts based on new discoveries in physics, geology, climatology and so on. It thus turns out that the given statement is only short for "Relative to the available evidence, the probability that the dinosaurs were made extinct by a giant asteroid hitting the Earth is very high - about 0.9." Clearly, belief-type probability statements are about how confident one can or should be, in the light of acquired evidence. Note that it is possible to do experiments to test the claims of the natural sciences (the evidence), the results may or may not lead to assert the above statement about dinosaurs. But it makes no sense at all to talk about repeatedly testing the statement itself. So belief-type probability statements express how much confidence a person has in a belief. Belief-type probability states the credibility of a proposition on the strength of the available evidence.

10.2 Language System

Bar-Hillel and Carnap develop their theory of information relative to a language system which has been introduced in (Carnap, 1950). The proposed language system is constituted of

- n different individual constants, standing for individuals like things, events, positions, etc., and
- π primitive one-place predicates, designating primitive properties of the individuals.

Such a language system gives rise to all kind of languages $\mathcal{L}_n^\pi$ with n and π being natural numbers. However, the authors point out that under certain assumptions one can easily extend their results to a language system with a denumerably infinite number of individual constants. This restriction has several reasons, as explained in (Bar-Hillel & Carnap, 1952). On the one hand, the stage of development of inductive logic, as given in (Carnap, 1950), does not cover languages richer than the ones supported by the chosen language system. On the other hand, the languages $\mathcal{L}_n^\pi$, with n and π being relatively small, are sufficiently powerful to give significant examples, without loosing the simplicity of presentation. Individual constants are denoted by lower case letter, such as a, b, c, predicates by capitals, like P, Q, R, sometimes with indices in $\mathbb{N}$. We will first see how sentences in $\mathcal{L}_n^\pi$ are formed and point out some special ones. Thereafter the focus falls on state-descriptions which are one of the main components of the semantics of the proposed language system. Finally, an instance of the language system is given. This sample language will accompany us through the rest of this chapter. Note that this whole section about the language system used by Bar-Hillel and Carnap will be rather informal, as we merely want to communicate their concepts. A very formal presentation would be cumbersome.

10.2.1 Sentences

Sentences can be formed using the n individual constants and the π predicates of $\mathcal{L}_n^\pi$, as well as the five customary connectives, namely $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (implication) and $\leftrightarrow$ (equivalence).[2] The following table gives a survey of the different kinds of sentences defined:

name	description	example
atomic sentence	one property attributed to one individual	Pa (individual a has property P)
basic sentence	atomic sentence or its negation	$\neg Pa$
Q-sentence	conjunction of π basic sentences covering all predicates, always relative to the same individual constant	$P_1a \wedge \neg P_2a \wedge \ldots \wedge P_{\pi-1}a \wedge \neg P_\pi a$

Obviously, the number of atomic sentences is $\pi \cdot n$. Note that Q-sentences are special cases of sentences. Sentences in general are just defined to be basic sentences, concatenated by the above five connectives, but no special requirements, neither regarding the number of predicates nor the occurrence of the individual constants, as in the case of the Q-sentence, exist. There is a simplified notation for sentences, emphasizing that one individual can have several properties. Instead of connecting a certain number of basic sentences of the same individual constant by one of the above binary connectives, only the predicates are concatenated and put into brackets,

[2] Note that this is not the notation used by Bar-Hillel and Carnap, but the one which is already known from Chapters 8 and 9. The authors use . instead of $\wedge$, $\sim$ instead of $\neg$, $\supset$ instead of $\rightarrow$ and $\equiv$ instead of $\leftrightarrow$.

followed by one occurrence of the individual constant. $P_1a \wedge \neg P_2a \wedge \ldots \wedge P_{\pi-1}a \wedge \neg P_\pi a$ can thus be rewritten as $[P_1 \wedge \neg P_2 \wedge \ldots \wedge P_{\pi-1} \wedge \neg P_\pi]a$.

10.2.2 State-Descriptions

One of the central notions in Bar-Hillel and Carnap's theory of semantic information is the state-description, which is a special kind of sentence, but may also be seen from a semantic point of view.

Definition 10.1 (State-Description)

A state-description is a conjunction of n Q-sentences, one for each individual constant. The set containing all state-descriptions is denoted by V.

A language $\mathcal{L}_n^\pi$ makes n individuals and π properties or predicates available. Each individual has a certain set of these properties. For every individual, a state-description determines which of the π properties it has, and which not. For example, given a language $\mathcal{L}_2^3$, where the set of individual constants consists of a and b and the set of predicates is given by P, Q and R, a possible state-description is $[P \wedge \neg Q \wedge \neg R]a \wedge [P \wedge \neg Q \wedge R]b$. This means that the individual a has the property P, but not Q nor R and the individual b has properties P and R, but not Q. It is important to point out that also the predicates which are not fulfilled by an individual constant are mentioned in the state-description, though in the negated form. The language $\mathcal{L}_n^\pi$ is seen as fixing a universe of discourse, so a state-description exhaustively describes a possible state of the world. A state-description therefore says the most that can be said in a given universe, short of self-contradiction. Clearly, the number of possible state-descriptions is $2^{\pi \cdot n}$.

It is quite evident that state-descriptions, even if they are looked at from a syntactic point of view, namely as sentences, play the same role as valuations (see Definition 9.8). Thus all the concepts of Section 9.3, in particular Definitions 9.12 and 9.13, can be taken over. Bar-Hillel and Carnap do not speak of some valuation being a model of a formula, but they say that some state-description holds in a sentence. Consequently, in their terminology, a formula does not have a corresponding set of models, but for any sentence j which can be expressed by means of $\mathcal{L}_n^\pi$, there is a set of state-descriptions in which j holds. This (maximum) set of state-descriptions is called the range of sentence j:

Definition 10.2 (Range of a Sentence)

The range of some sentence j consists of all state-descripitons in which j holds. It is denoted by $R(j)$.

In terms of Chapter 9, the range of a sentence corresponds to its set of models. The language $\mathcal{L}_n^\pi$ fixes a universe of discourse, which can be in $2^{\pi \cdot n}$ possible states. A sentence j thus limits the possible states from V (the set of all state descriptions) to

$R(j)$ by saying that the state of the universe is one of the state-descriptions in $R(j)$. Bar-Hillel and Carnap subdivide sentences into three semantic classes. They are listed and exemplified in the table given in Figure 10.1.

name	description	example
true	logically true tautological	$Pa \vee \neg Pa$ $\top$
false	logically false self-contradictory	$Pa \wedge \neg Pa$ $\perp$
factual	logically indeterminate, synthetic	$Pa \vee [Q \wedge \neg R]b$

Figure 10.1: Sentences are either true, false or factual

So if, and only if, the sentence j is false, its range is the empty set. On the other hand, the range of a true sentence contains all possible $2^{\pi \cdot n}$ state-descriptions. Nowadays, a sentence that is neither a tautology nor a contradiction is not any more called factual, but is said to be contingent.

Furthermore, Carnap and Bar-Hillel define four relations between sentences i and j, called logical relations by the authors:

$$\begin{array}{llll} i \text{ implies } j & := & i \rightarrow j & \text{is true} \\ i \text{ is equivalent to } j & := & i \leftrightarrow j & \text{is true} \\ i \text{ is disjunct with } j & := & i \vee j & \text{is true} \\ i \text{ is exclusive of } j & := & i \wedge j & \text{is false} \end{array}$$

Every state-description in the range of some sentence j may be seen as a sentence i which implies j. If j is not false, it is equivalent to the disjunction of the state-descriptions in its range.

10.2.3 Example Language

Bar-Hillel and Carnap propose in their work an instance of the language system $\mathcal{L}_n^\pi$, used to illustrate the concepts of their theory. This sample language is very easy, it consists of three individual constants a, b, c and two properties M and Y, so it is a $\mathcal{L}_3^2$ language. The authors even invent a story behind their example language: A census is taken in a small community of only three inhabitants (a, b, c). One is only interested whether the inhabitants are male (M) or non-male ($\neg M$, i. e. female) and young (Y) or non-young ($\neg Y$, i. e. old). In (Bar-Hillel & Carnap, 1952), the authors give the list of all possible 64 state-descriptions, in an abbreviated form. Figure 10.2 shows this table: It is a copy from the original paper. This table being a central element of the theory and the examples given in the sequel, it will be reused later on, in a different interpretation. As $\neg M$ means female, it is rewritten by F. In the same way, $\neg Y$ is replaced by O for old.

	M, Y	M, O	F, Y	F, O		M, Y	M, O	F, Y	F, O
1.	a, b, c	-	-	-	33.	b	-	-	a, c
2.	-	a, b, c	-	-	34.	a	-	-	b, c
3.	-	-	a, b, c	-	35.	-	c	-	a, b
4.	-	-	-	a, b, c	36.	-	b	-	a, c
5.	a, b	c	-	-	37.	-	a	-	b, c
6.	a, c	b	-	-	38.	-	-	c	a, b
7.	b, c	a	-	-	39.	-	-	b	a, c
8.	a, b	-	c	-	40.	-	-	a	b, c
9.	a, c	-	b	-	41.	a	b	c	-
10.	b, c	-	a	-	42.	a	c	b	-
11.	a, b	-	-	c	43.	b	a	c	-
12.	a, c	-	-	b	44.	b	c	a	-
13.	b, c	-	-	a	45.	c	a	b	-
14.	c	a, b	-	-	46.	c	b	a	-
15.	b	a, c	-	-	47.	a	b	-	c
16.	a	b, c	-	-	48.	a	c	-	b
17.	-	a, b	c	-	49.	b	a	-	c
18.	-	a, c	b	-	50.	b	c	-	a
19.	-	b, c	a	-	51.	c	a	-	b
20.	-	a, b	-	c	52.	c	b	-	a
21.	-	a, c	-	b	53.	a	-	b	c
22.	-	b, c	-	a	54.	a	-	c	b
23.	c	-	a, b	-	55.	b	-	a	c
24.	b	-	a, c	-	56.	b	-	c	a
25.	a	-	b, c	-	57.	c	-	a	b
26.	-	c	a, b	-	58.	c	-	b	a
27.	-	b	a, c	-	59.	-	a	b	c
28.	-	a	b, c	-	60.	-	a	c	b
29.	-	-	a, b	c	61.	-	b	a	c
30.	-	-	a, c	b	62.	-	b	c	a
31.	-	-	b, c	a	63.	-	c	a	b
32.	c	-	-	a, b	64.	-	c	b	a

Figure 10.2: The 64 state-descriptions of the sample language

Read the table given in Figure 10.2 as follows: State-description 1 corresponds to the sentence $[M \wedge Y]a \wedge [M \wedge Y]b \wedge [M \wedge Y]c$. A more sophisticated example of a state-description is number 42, given by the sentence $[M \wedge Y]a \wedge [F \wedge Y]b \wedge [M \wedge O]c$, which is the rewritten form of $[M \wedge Y]a \wedge [\neg M \wedge Y]b \wedge [M \wedge \neg Y]c$. As seen above, a sentence is equivalent to the disjunction of the state-descriptions in its range. Some examples given by Carnap and Bar-Hillel are now quoted:

- The range of the sentence $j = [M \wedge Y]a \wedge [F \wedge Y]b$ contains four state-descriptions, $R(j) = \{9, 25, 42, 53\}$.
- The range of the sentence Fa contains 32 state-descriptions, namely those in which a occurs either in the column F,Y or F,O.
- The range of the sentence $Ma \vee Ya \vee Fb \vee Yb \vee Fc \vee Oc$ contains 63 state-descriptions, namely all except 52.[3]

[3]This sentence is equivalent to $\neg(Fa \wedge Oa \wedge Mb \wedge Ob \wedge Mc \wedge Yc)$, hence to the negation of state-description 52. This is an important hint concerning the following sections, especially related to Definition 10.7.

10.3 Semantic Information

One important part of Bar-Hillel and Carnap's theory consists of a description of the nature of semantic information in an abstract way. The authors give a framework for semantic information by describing its properties in an axiomatic way. They call this framework the presystematic concept of semantic information. It provides requirements and derived properties that each formalism for semantic information should fulfill, see Section 10.3.1. A formalism which meets the characteristics of the framework is thereafter presented in Section 10.3.2. The authors call such an instance an explicatum for the presystematic concept. Carnap and Bar-Hillel show that the formalism presented as an instance is in fact one, by verifying that its properties agree with what is asked by the concept (Section 10.3.3). Finally, this section is concluded with a graphical representation of *Cont* and its properties.

10.3.1 The Presystematic Concept

Bar-Hillel and Carnap distinguish between information, seen in an absolute sense, and information, relative to other information which came up previously.

Absolute Information

(Bar-Hillel & Carnap, 1952) propose only one requirement and derive a couple of theorems from it. Together they form a set of properties characterizing semantic information, as perceived by Carnap and Bar-Hillel. For simplifying the notation, the authors introduce the abbreviation

$$In(i) \quad \text{for} \quad \text{“the information carried by the sentence } i\text{”}. \tag{10.1}$$

Based on the idea that whenever a sentence i implies another sentence j, the former tells everything that the latter tells, and even possibly more, the requirement is formulated:

Requirement 10.3 $In(i) \supseteq In(j)$ *iff* $i \rightarrow j$.

This shows that Bar-Hillel and Carnap treat semantic information as a set, and they give further properties of the semantic information In without proofs, as they are well-known theorems of the theory of sets[4], see (Halmos, 1974):

Theorem 10.4 *Consider two sentences i and j. The semantic information carried by them, denoted by $In(i)$ and $In(j)$, respectively, has the following properties:*

[4] Note that Carnap and Bar-Hillel always use the term “class” when they talk of what is called “set” nowadays. In our description of their theory, we have chosen the term “set”, as it is more easily readable in today's speech.

1. $In(i) = In(j)$ *iff* $i \leftrightarrow j$.
2. $In(i)$ *is the minimum set* min *of all* In*-sets iff* i *is true.*
3. $In(i)$ *is the maximum set* max *of all* In*-sets iff* i *is false.*
4. $In(i) \supset In(j)$ *iff* $i \rightarrow j$, *and* $j \not\rightarrow i$.
5. $\max \supset In(i) \supset \min$ *iff* i *is factual.*
6. $In(i \wedge j) \supseteq In(i) \supseteq In(i \vee j)$.

The first property says that two sentences which are equivalent provide also the same semantic information. For the second and the third property, some further vocabulary is needed. Consider a set K of sets $k_i, i \in \mathbb{N}$. A set $k' \in K$ which is included in every $k_i \in K$ is called the minimum set of K. The maximum set of K is defined to be the member of K which includes every $k_i \in K$. By determining for every sentence i its corresponding set $In(i)$, one can build the set of all In-sets whose minimum set is related to a true sentence and whose maximum set to a false sentence. The authors point out that it might be strange to attribute to a contradiction the most inclusive information, but they give a very good answer to this point of criticism in (Bar-Hillel & Carnap, 1952):

> "A self-contradictory sentence asserts too much; it is too informative to be true."

Property four asserts that a proper inclusion relation between two In-sets is only possible if the corresponding sentences are not equivalent. Another proper inclusion, this time regarding the minimum and the maximum set, is given by the fifth property. Recall from the table given in Figure 10.1 that a factual (contingent) sentence is neither tautological nor contradictory. The min and max sets represent a tautology and a contradiction, respectively. So max properly includes the In-set related to a factual sentence and min is properly included by such an In-set. Finally, the sixth property states that the semantic information supplied by sentence i is more inclusive than the one by i or ($\vee$) some other sentence j. At the same time, it is less inclusive than the semantic information of i and ($\wedge$) another sentence j.

Relative Information

After having presented their view on the information carried by a sentence, Bar-Hillel and Carnap direct the readers' attention to the fact that the information carried by a sentences often refers to that carried by some other sentence(s). This is why there must not only be a presystematic concept of absolute information, but also a presystematic concept of relative information, $In(j|i)$. It captures the characteristics of semantic information $In(j)$, which is given in addition to some already available semantic information $In(i)$:

Definition 10.5 (Relative Information)

Consider two sentences i and j. The additional information carried by j with respect to i is captured by the concept of relative information:

$$\begin{aligned} In(j|i) &:= In(i \wedge j) \setminus In(i) \\ &= In(i \wedge j) \cap (In(i))^c \,. \end{aligned}$$

Relative information is thus either defined as the set-theoretical difference of $In(i \wedge j)$ and $In(i)$ or as the intersection of $In(i \wedge j)$ and the complement of $In(i)$. The result, $In(j|i)$, is obviously again a set whose members belong to the same type as the members of $In(i)$. Bar-Hillel and Carnap present some properties of the concept of relative information $In(j|i)$ in a couple of theorems, however without proof, as they follow directly from Definition 10.5, Requirement 10.3 and Theorem 10.4. We have united them under one theorem:

Theorem 10.6 *Consider four sentences i, j, k, l. The relative semantic information has the following properties:*

1. $\max \supseteq In(j|i) \supseteq \min$
2. *If $i \leftrightarrow j$, then $In(k|i) = In(k|j)$ and $In(i|l) = In(j|l)$.*
3. *If $i \rightarrow j$, then $In(j|i) = \min$.*
4. *If j is true, then $In(j|i) = \min$.*
5. *$In(j|i) \supset \min$ iff $i \not\rightarrow j$.*
6. *If i is a true sentence, then $In(j|i) = In(j|\top) = In(j)$.*

The first property tells that $In(j|i)$, the semantic information carried by the sentence j in addition to that carried by another sentence i, can be of any kind: factual (proper inclusion), true (minimum set) or false (maximum set). Property two states that, if two sentences are equivalent, equality holds for the corresponding relative information, no matter whether the equivalent sentences provide the additional or the original information. The third property deals with the case where some sentence i implies another sentence j. In an implication situation, i asserts everything that is asserted by j and therefore no supplementary information is available. Learning nothing new means learning a tautology (a true sentence) whose semantic information is expressed by the minimum set. A similar setting is captured by property 4: Here, originally an arbitrary sentence i is received and thereafter a true one, thus the initial information does not matter any more, $In(j|i)$ is the minimum set. Property five states that $In(j|i)$ expresses more than a tautology (as it properly includes the minimum set) if and only if the sentence i does not assert all that is asserted by j. The statement made by the sixth property is that $In(j|i)$ equals $In(j)$ if i, the sentence learned first, is a tautology, not carrying any information. Bar-Hillel and Carnap point out that this last property allows to define absolute information by means of relative information, as knowing a tautology means referring to no previous knowledge.

10.3.2 The Explicatum

Carnap and Bar-Hillel propose several formalisms which fulfill Requirement 10.3 and could therefore serve as explicatum for the concept In. They eventually choose the formalism $Cont$, which is based on the old philosophical principle "omnis determinatio est negatio", to be the adequate one. $Cont$ is defined to be a set of so-called content-elements.

Content-Elements

Content-elements are closely related to state-descriptions. Recall that the range $R(j)$ of some sentence j is the set of all state-descriptions in which j holds, see Definition 10.2. So j says that the universe of discourse is in one of the states provided by $R(j)$. In other words, it says that the universe is in none of the states of $V \setminus R(j)$, where V is the set of all possible $2^{\pi \cdot n}$ state-descriptions. From Section 10.2 we know that the sentence j is implied by every state-description in its range $R(j)$. But by the above consideration, j also implies the negation of each state-description in $V \setminus R(j)$. These negated state-descriptions are called content-elements and are denoted by E by Bar-Hillel and Carnap:

Definition 10.7 (Content-Element)

A content-element E is a negated state-description.

Such a scenario is already known from the algebraic theory of semantic information introduced in the first part of this thesis. Content-elements describe what is excluded by the information (conjunctive view). However, the information itself is given by state-descriptions, in a disjunctive fashion. See Section 10.5.2 for more details.

The table shown in Figure 10.2 provides all $2^{2 \cdot 3} = 64$ state-descriptions of the example language $\mathcal{L}_3^2$ in a condensed form. In order to obtain the state-description in a syntactically correct form, conjunctions are introduced. The properties for each individual are connected by $\wedge$, thereafter the resulting Q-sentences (see Section 10.2.1) are combined using once again the connective $\wedge$ and thus form a state-description. Content-elements are constructed in a very similar way, but the whole state-description is negated afterwards. The content-element 1, i. e. the negation of state-description 1, corresponds to the sentence $\neg([M \wedge Y]a \wedge [M \wedge Y]b \wedge [M \wedge Y]c)$, which is by the de Morgan laws $[F \vee O]a \vee [F \vee O]b \vee [F \vee O]c$. Let us consider as another example content-element 42, $\neg([M \wedge Y]a \wedge [F \wedge Y]b \wedge [M \wedge O]c)$, which is equivalent to $[F \vee O]a \vee [M \vee O]b \vee [F \vee Y]c$. Clearly, when starting with a state-description, one can obtain the corresponding content-element by replacing the conjunctions by disjunctions and by changing the sign of each predicate.

Bar-Hillel and Carnap give a list of theorems which can be deduced from some theorems in (Carnap, 1950), but which can also be verified with the knowledge about state-descriptions and content-elements acquired so far. Recall from Section 10.2 that a factual sentence is neither a tautology nor a contradiction and that sentence i being disjunct with sentence j means that their disjunction is true.

Theorem 10.8 *For any two content-elements E and E', the following holds:*

1. *E is factual.*
2. *If E' is distinct with E, then E and E' are disjunct.*
3. *The conjunction of all E is false.*
4. *If E implies the sentence j, then j is either true or equivalent to E; in other words, E is the weakest factual sentence.*

Property one says that a content-element can never be a tautology or a contradiction. The disjunction of two different content-elements is always true, as stated by property two. From property three we know that the conjunction of all $2^{\pi \cdot n}$ content-elements results in a contradiction. Finally, property four states that a content-element says the least that can be said in a given universe of discourse, beyond a tautology.[5]

Cont

Just as before, in the case of state-descriptions, we can define a set similar to the range $R(j)$ of a sentence j (see Definition 10.2). The sentence j states that the universe is not in one of the states described by $V \setminus R(j)$, so j implies the negation of every state-description in this set, as seen above. In other words, j implies the content-elements of $V \setminus R(j)$. Bar-Hillel and Carnap call this the set $Cont(j)$, or the content of j.

Definition 10.9 (Content of a Sentence)

The content of a sentence j consists of all content-elements which are implied by j. It is denoted by $Cont(j)$. Alternatively formulated, $Cont(j)$ contains the negation of all state-descriptions which are not in the range of j.

Graphically speaking, $Cont(j)$ is the complement of $R(j)$. In Figure 10.3, $R(j)$ is the white circle. Therefore, the dashed part is $Cont(j)$, i. e. everything which is not in $R(j)$.

We will now give some examples for the content of sentences. The examples are already known from Section 10.2.3 in the context of state-descriptions and range. They refer to the table given in Figure 10.2.

[5] Recall that a state-description says the most that can be said in a universe.

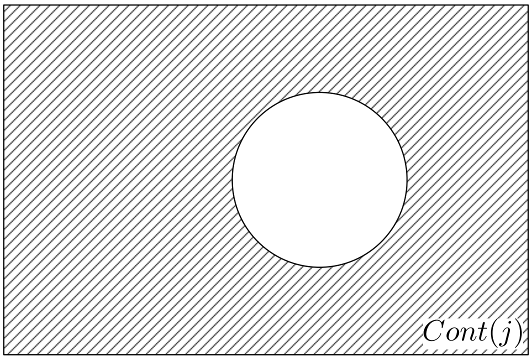


Figure 10.3: *Cont* as the complement of the range of a sentence

- The content of the sentence $j = [M \wedge Y]a \wedge [F \wedge Y]b$ contains 60 content-elements, namely the negation of all state-descriptions in V, except the state-descriptions in $R(j) = \{9, 25, 42, 53\}$.
- The content of the sentence Fa contains 32 content-elements, namely those in which a occurs either in the column M,Y or M,O.
- The content of the sentence $Ma \vee Ya \vee Fb \vee Yb \vee Fc \vee Oc$ contains only one content-element, namely the one given by line 52. This is due to the fact that this sentence is equivalent to $\neg(Fa \wedge Oa \wedge Mb \wedge Ob \wedge Mc \wedge Yc)$, i. e. the negation of state-description 52.

In (Bar-Hillel & Carnap, 1952), the authors list a lot of properties of *Cont* which are unified in the following theorem. Recall from Section 10.2 that a factual sentence is neither a tautology nor a contradiction and that sentence i being disjunct with sentence j means that their disjunction is true. We refer to Figure 10.6 for an illustration.

Theorem 10.10 *Consider any two sentences i and j. Their content, denoted by $Cont(i)$ and $Cont(j)$, respectively, has the following properties:*

1. *$Cont(i)$ is the minimum set* min *of all Cont-sets iff i is true.*
2. *$Cont(i)$ is the maximum set* max *of all Cont-sets iff i is false.*
3. *$Cont(i) \neq \min$, $Cont(i) \neq \max$ iff i is factual.*
4. *$Cont(i) \supseteq Cont(j)$ iff $i \rightarrow j$.*
5. *$Cont(i) = Cont(j)$ iff $i \leftrightarrow j$.*
6. *$Cont(i) \cap Cont(j) = \emptyset$ iff i and j are disjunct.*
7. *$Cont(\neg i) = (Cont(i))^c$.*

8. $Cont(i \vee j) = Cont(i) \cap Cont(j)$.

9. $Cont(i \wedge j) = Cont(i) \cup Cont(j)$.

The first three properties are related to the set of all $Cont$-sets. This set contains all possible sets of content-elements, the minimum set is the empty set (property one) and the maximum set is composed of all possible content-elements, i. e. all negations of all state-descriptions in V (property two). The content of a true sentence is thus the empty set, the content of a false sentence is the set of all content-elements and the content of every factual sentence lies between those two extreme points (property three). Property four states that whenever a sentence i implies a sentence j, the content of the former includes the content of the latter. The fifth property tells that two sentences which are equivalent have exactly the same content, meaning that their $Cont$-sets are constituted of the same content-elements. The contrary is the case for two disjunct sentences (i. e. their disjunction is a tautology), as they do not share any content-element, as indicated by the sixth property. Sentences fulfilling this property are also called content-exclusive or disjoint. This is an important quality in Bar-Hillel and Carnap's theory of semantic information and is illustrated by Figure 10.4.

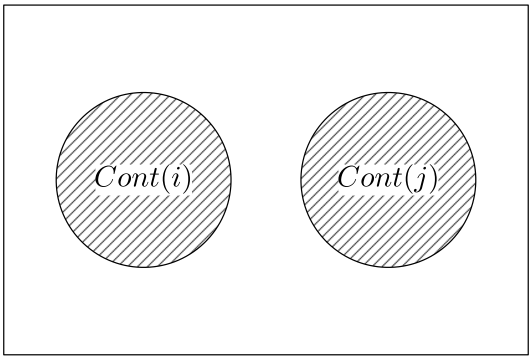


Figure 10.4: Content-exclusiveness: $Cont(i)$ and $Cont(j)$ do not have any content-elements in common

Property seven proposes another writing for the complement of $Cont(i)$. $Cont(\neg i)$ contains all the content-elements which are implied by $\neg i$, or alternatively, which are not implied by i, as the closed world assumption (see page 74) holds. The last two properties show how to trace back the disjunction and the conjunction of two sentences to their respective contents. $Cont(i \vee j)$ corresponds to the intersection of the contents of the two sentences i and j. $Cont(i \wedge j)$ has the same result as the union of the contents of these two sentences.

10.3.3 Verification of the Explicatum

The properties of $Cont$ given in Theorem 10.10 are in accord with Theorem 10.4. Even more important is the fact that the fourth point of Theorem 10.10 shows that *Cont* fulfills Requirement 10.3. In the beginning of this section, we presented this requirement to be the essential condition for Bar-Hillel and Carnap that every formalism for semantic information has to meet. The requirement states that the information carried by some sentence j is included in the information of some other sentence i if, and only if, i implies j. In other words, the information supplied by i is more inclusive than the information provided by j. Point four of Theorem 10.10 and Requirement 10.3 show that *Cont* is an instance of In.

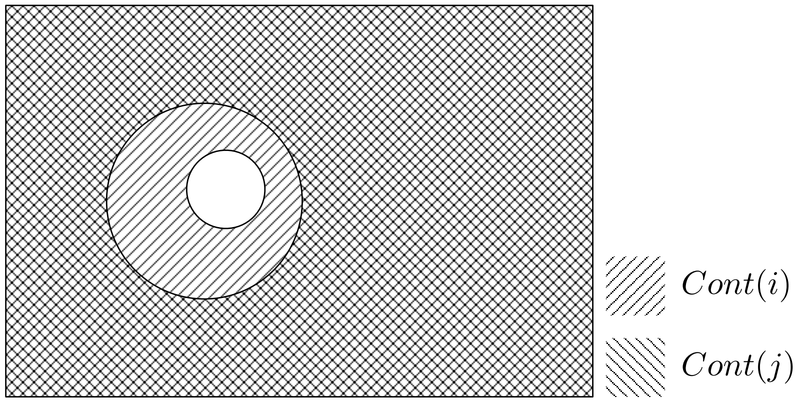


Figure 10.5: Illustration of Requirement 10.3 by means of *Cont*

Figure 10.5 gives a graphical representation of the situation. $Cont(i)$ covers more space than $Cont(j)$, namely everything but the white circle. $Cont(j)$ spans only the squared part, resulting from the superimposed shadings. So $Cont(i)$ contains more content-elements than $Cont(j)$, i.e. $Cont(i) \supseteq Cont(j)$. Looked at from the point of view of state-descriptions, $R(i) \subseteq R(j)$, where $R(i)$ corresponds to the state-descriptions of the smaller circle (white part), and $R(j)$ is constituted of the state-descriptions in the bigger circle (shaded part and white part). From the previous chapters it is known that whenever a sentence i implies a sentence j, the set of models (or the range, in terms of Carnap and Bar-Hillel) of the former is included in the set of models of the latter.

In (Bar-Hillel & Carnap, 1952), the authors justify the fact of having discarded other explicata priorly proposed and having chosen *Cont* as explicatum as follows:

> "The explication of the information carried by a sentence j, as the set of the negations of all those state-descriptions which are excluded by j, is intuitively plausible and in accordance with the old philosophical principle ≪omnis determinatio est negatio≫. Our main reason, however, for giving it preference over the two explicata [...], lies in the fact that an explication of amount of information will turn out to be rather simple if based on *Cont*, [...]."

Before going on with the announced amount of information (see Section 10.4), Carnap and Bar-Hillel provide a the definition of the relative content of a sentence j with respect to a sentence i:

Definition 10.11 (Relative Content of a Sentence)

Consider two sentences i and j. The additional content of j with respect to i is the set-theoretical difference of $Cont(i \wedge j)$ and $Cont(i)$. It is called relative content of j, given that i is already known and denoted by

$$Cont(j|i) \quad := \quad Cont(i \wedge j) \setminus Cont(i).$$

Obviously, the properties of the relative information stated in Theorem 10.6 also hold for the relative content *Cont*. In the third row of Figure 10.6, on the right hand side, an illustration of the relative content is provided.

10.3.4 Graphical Representation of *Cont*

Figure 10.6 illustrates how *Cont* represents semantic information. It helps to understand the properties of *Cont* given in Theorem 10.10, as well as Definition 10.11.

The closed world assumption allows to represent the universe of discourse by a rectangle, containing the $2^{\pi \cdot n}$ content-elements, which are however not depicted. Note that the shaded parts correspond to the respective *Cont*-set.

The first row shows the content of sentence i and the content of sentence j. The second row provides the content of $i \vee j$ (left hand side, intersection of $Cont(i)$ and $Cont(j)$) and the content of $i \wedge j$ (right hand side, union of $Cont(i)$ and $Cont(j)$). As easily seen, the complement of $Cont(i)$ is on the left of row three, whereas on the right the relative content of j with respect to i is shown. By comparing this picture to the first row, it is obvious that $Cont(j|i)$ describes a situation, where $Cont(i)$ is already known, and all that knowing $Cont(j)$ can tell in addition is the shaded part, the rest is already known thanks to $Cont(i)$. Another way to look at the relative content is to apply Definition 10.11, where $Cont(i)$ is subtracted from $Cont(i \wedge j)$. Two special cases are given in the last row. On the left, the minimum set (empty set) presents a true sentence. A false sentence is depicted on the right; it is the maximum set, containing all content-elements.

10.4 Amount of Information

In (Bar-Hillel & Carnap, 1952), the authors do not leave their theory of semantic information at a description of the nature of semantic information[6] for which they have provided the instance "*Cont*" - a formalism fulfilling the requirement for semantic information. They furthermore turn to the measure of semantic information and propose again a framework for the conceptual level (Section 10.4.1) where the conditions for a formalism measuring the amount of information are formulated. Two ways of fulfilling the requirements are outlined in Sections 10.4.3 and 10.4.4. As both of them rely on the same measure-function, these two sections are preceded by

[6]This description has also been called the framework or presystematic concept of information.

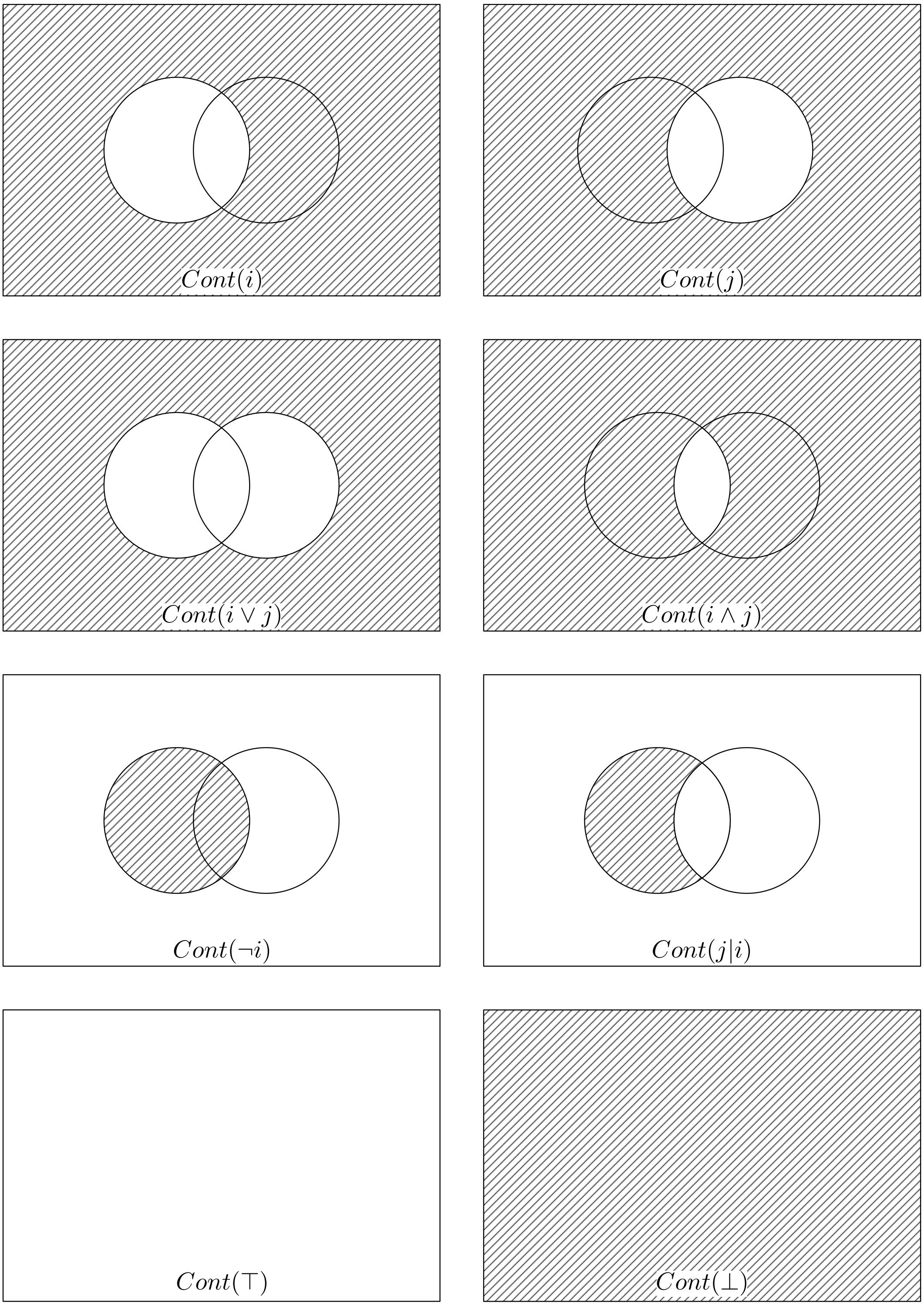


Figure 10.6: Illustration of $Cont$

a very brief introduction to m-functions, see Section 10.4.2. Finally, it is explained why the authors speak for two explicata for amount of information, instead of only one (Section 10.4.5). Note that the part about amount of information is much more detailed in (Bar-Hillel & Carnap, 1952). The authors attach greater importance to the measure of information than to the semantic information itself. Therefore, this section about amount of information outlines in a very condensed way the authors' ideas, stressing the parts which are important for a later comparison to the algebraic theory of semantic information in Section 10.6.

10.4.1 The Presystematic Concept

Carnap and Bar-Hillel want to measure the amount of semantic information of a sentence by a numerical value. They list some requirements that such a measure has to fulfill in their opinion. The requirements are formulated using an abstract entity expressing the amount of information, called the presystematic concept of amount of information and denoted by the symbol in:

$in(i)$ stands for the absolute amount of information of a sentence i and
$in(j|i)$ stands for the relative amount of information of a sentence j with respect to sentence i.

As already seen in Section 10.3.1, the relative information may be expressed by means of absolute information. The same holds for the relative amount of information, which can be reduced to the absolute amount:

Definition 10.12 (Relative Amount of Information)

Consider two sentences i and j. The amount of information that j provides with respect to i is determined by the relative amount of information:

$$in(j|i) \quad := \quad in(i \wedge j) - in(i).$$

Note that in contrast to Definition 10.5, where the relative information was defined using the set difference, this time we are dealing with measurements and thus, "$-$" is the numerical difference and the result will be a numerical value. Now that relative amount of information has been traced back to the absolute amount, it is sufficient to state only the requirements with respect to the latter. i and j are sentences, $in(i)$ and $in(j)$ their respective amounts of information:

Requirement 10.13 $in(i) \geq in(j)$ *if (but not necessarily only if)* $i \rightarrow j$.

Requirement 10.14 $in(j) = 0$ *if* j *is true.*

Requirement 10.15 $in(j) > 0$ *if* j *is not true.*

It seems plausible to Bar-Hillel and Carnap to require that the amount of information of i should not be less than the amount of information of j, if i implies j. In terms of the information In carried by a sentence (In being the "presystematic concept of semantic information", see Section 10.3.1), if the amount of information of the sentence i is greater than or equal to the amount of information of the sentence j, then the information of i includes the information of j:

$$in(i) \geq in(j) \text{ if } In(i) \supseteq In(j).$$

Furthermore, the authors claim that the amount of information of a true sentence should be zero, so a true sentence does not carry any information. Consequently, the amount of information of every sentence which is not true (factual or false) should be greater than zero. There are further properties of in which are derived from the three requirements above:

Theorem 10.16 *Consider four sentences i, j, k and l. Their amount of information, denoted by $in(i)$, $in(j)$, $in(k)$ and $in(l)$, respectively, has the following properties:*

1. *If $i \leftrightarrow j$, then $in(i) = in(j)$.*
2. *If j is false, then $in(j)$ has the maximum in-value.*
3. *$0 < in(j) <$ the maximum in-value iff j is factual.*
4. *$in(i \vee j) \leq in(j) \leq in(i \wedge j)$.*
5. *$in(i \wedge j) \geq \max(in(i), in(j))$.*
6. *The maximum in-value $\geq in(j|i) \geq 0$.*
7. *If $i \leftrightarrow j$, then $in(k|i) = in(k|j)$ and $in(i|l) = in(i|l)$.*
8. *If $i \rightarrow j$, then $in(j|i) = 0$.*
9. *If j is true, then $in(j|i) = 0$.*
10. *$in(j|i) > 0$ iff $i \not\rightarrow j$.*
11. *If i is a true sentence, then $in(j|i) = in(j|\top) = in(j)$.*

Obviously, two equivalent sentences, which have the same content, provide also the same amount of information (property one). In Section 10.3.1, we have seen that a false sentence provides the most inclusive information. Property two therefore attributes the highest amount of information to such sentences. As a consequence, the next property states that the amount of information of a factual sentence, which is neither true nor false, is located between two extreme points, determined by Requirement 10.14 and property two. The fourth and the fifth property are dealing with the amount of information of a conjunction and a disjunction of sentences. The amount of information of some sentence j is always superior to the amount of information of its disjunction with another sentence, but it is inferior to its conjunction with another sentence. A conjunction of two sentences provides a higher amount of information than the maximum of its two elements. Equality may hold. The last six points of this theorem correspond to the properties of relative information given in Theorem 10.6 which can be consulted for further explications.

Carnap and Bar-Hillel admit that Requirements 10.13 to 10.15 and the Theorem 10.16 are rather weak and point out that one might require additivity of amount of information. They state that this requirement has though not been treated deliberately. The two explicata presented in the following are indeed additive, but not in the same sense, due to different conditions of independence. All that the authors presuppose so far is the lower limit for $in(i \wedge j)$ (property five of Theorem 10.16) and they state furthermore that in general $in(i \wedge j) \neq in(i) + in(j)$.

10.4.2 Measure-Functions

In Bar-Hillel and Carnap's opinion, there is not only one formalism fulfilling the above conditions of the presystematic concept of amount of information, but it makes sense to consider at least two such formalisms. Both of them are based on belief-type probability measure-functions ranging over state-descriptions, which can be extended to sentences. These measure-functions, originally introduced in (Carnap, 1950), are called m-functions. They are meant to be an adequate explicatum for the concept of belief-type probability (see Section 10.1).

Let us start with some basic properties defining an m-function, which offers the characteristics required for representing a formalism for belief-type probability. The following definition shows that an m-function is a positive probability measure over state-descriptions.

Definition 10.17 (m-Function)

Let Z be a state-description out of V (the set of all state-descriptions), j a sentence and $R(j)$ its range, containing all state-descriptions in which j holds. m is said to be an m-function, if the following conditions are met:

1. $m(Z) > 0, \forall Z \in V$.
2. $\sum_{Z \in V} m(Z) = 1$.
3. *If j is a false sentence, then* $m(j) := 0$.
4. *If j is not a false sentence, then* $m(j) = \sum_{Z \in R(j)} m(Z)$.
5. *For any sentence j,* $m(\neg j) = 1 - m(j)$.

By definition, the belief-type probability of every state-description is greater than zero. The belief-type probabilities of all possible $2^{\pi \cdot n}$ state-descriptions sum up to one. The belief-type probability of sentence j is denoted by $m(j)$. This probability is computed by adding the respective probabilities of the state-descriptions in j's range. Whenever a sentence is false, its range contains no state-descriptions, it belief-type probability is thus 0. One may obtain the belief-type probability of a negated sentence $\neg j$ by deducing $m(j)$ (the belief-type probability of the unnegated sentence) from 1. The above conditions are not independent, as the fourth implies the fifth.

10.4.3 The First Explicatum: Content-Measure (*cont*)

The first measure which is proposed by Bar-Hillel and Carnap is called content-measure and denoted by *cont* with a lower-case "c", measuring the content $Cont(j)$, with a capital "C", of a sentence j. From Definition 10.9 it is known that $Cont(j)$ contains the negation of all state-descriptions which are not in $R(j)$. Thus the following definition of $cont(j)$ is very straightforward:

Definition 10.18 (Content-Measure *cont*)

Let $m(j)$ be the belief-type probability of the sentence j. Then,

$$cont(j) \quad := \quad 1 - m(j)$$

is a measure for $Cont(j)$, the content of sentence j, or any other adequate explicatum for $In(j)$, the semantic information carried by sentence j. This measure is called the content-measure of j.

Before listing the properties of *cont* that are pointed out by Bar-Hillel and Carnap, we want to attract attention to the fact that $cont(j)$ is simply the probability of $Cont(j)$. Even if Carnap and Bar-Hillel do not consider this fact, $cont(j) = p(Cont(j))$ holds, since $m(j) = p(R(j))$, and therefore,

$$\begin{aligned} cont(j) &= 1 - m(j) \\ &= 1 - p(R(j)) \\ &= p(R(j)^c) \\ &= p(V \setminus R(j)) \\ &= p(Cont(j)). \end{aligned}$$

The properties of *cont* which are pointed out by Bar-Hillel and Carnap are listed in the next theorem, without proof. Most of them are based on *cont*'s definition by means of m and are commonly known properties of probability. Recall from Section 10.2 that a factual sentence is neither a tautology nor a contradiction, that a sentence i being exclusive of a sentence j means that their disjunction is false, whereas i being disjunct with j means that their disjunction is true.

Theorem 10.19 *Consider two sentences i and j. Their content-measures $cont(i)$ and $cont(j)$ have the following properties:*

1. $0 \leq cont(i) \leq 1$.
2. $cont(i) = 0$ *iff i is true.*
3. $cont(i) = 1$ *iff i is false.*
4. $0 < cont(i) < 1$ *if i is factual.*
5. *If $i \rightarrow j$, then $cont(i) \geq cont(j)$.*
6. $cont(i \wedge j) \geq cont(i) \geq cont(i \vee j)$.
7. $cont(i \vee j) = cont(i) + cont(j) - cont(i \wedge j)$.
8. $cont(i \vee j) = cont(i) + cont(j) - 1$ *iff i and j are exclusive.*
9. $cont(i \wedge j) \begin{cases} = & cont(i) + cont(j) \quad \textit{if } i \textit{ and } j \textit{ are disjunct,} \\ < & cont(i) + cont(j) \quad \textit{else.} \end{cases}$

The first property states that measuring the content of a sentence always results in a numerical value in the interval $[0, 1]$. The two extreme points are reached when the sentence is true or false (properties 2 and 3). In any other case, the content-measure returns a value in $]0, 1[$, as stated by property 4. The fifth property asserts that if sentence i implies sentence j (which means nothing else than $Cont(i) \supseteq Cont(j)$), then i's content measure is greater than the one of j. It is known from (Carnap, 1950) that if $i \rightarrow j$, then $m(i) \leq m(j)$. Thus it holds that the greater the belief-type probability of a sentence is, the smaller is its content measure. The last four properties are all about the content-measure of conjunctions and disjunctions of sentences. The content-measure of the conjunction of two sentences always results in a greater value than the content-measure of one of both sentences, which is again greater than the content-measure of their disjunction (property 6). A generic way of computing the content-measure of the disjunction or the conjunction of two sentences is given in property 7. Property 8 states that whenever it holds that $i \wedge j$ is false, then the content-measure of their disjunction equals the sum of the single content-measures, minus one. But if i and j are disjunct (first line of property 9), i.e. content-exclusive ($Cont(i) \cap Cont(j) = \emptyset$, see Theorem 10.10, property 6 and Figure 10.4), the content-measure of the conjunction of both sentences equals simply the sum of the single content-measures. Otherwise, the *cont*-value of a conjunction of sentences is less than the sum of the *cont*-values of the single sentences. It can be easily seen that property 5 fulfills Requirement 10.13, property 2 corresponds to Requirement 10.14 and Requirement 10.15 is met by property 3 and 4. Thus *cont*'s adequacy has been verified and it is actually a formalism fulfilling the necessary requirements for measuring the amount of information.

In (Bar-Hillel & Carnap, 1952), and especially in (Bar-Hillel, 1964), it is pointed out that *cont* does not fulfill the usual additivity condition based on independence (see below). From property 9 of Theorem 10.19 we know that

$$cont(i) + cont(j) = cont(i \wedge j) \text{ if } Cont(i) \cap Cont(j) = \emptyset.$$

By Theorem 10.10, property 9, $Cont(i) \cap Cont(j) = Cont(i \vee j)$. In terms of state-descripitons and range, this may be rewritten as $V \setminus R(i \vee j)$. For the above additivity requirement, this means that V, the set of all state-descriptions, has to equal $R(i \vee j)$. This is only possible if $i \vee j$ is a tautology, as stated in the first property of Theorem 10.10. So, Bar-Hillel and Carnap's additivity condition for *cont* may be rewritten as

$$cont(i) + cont(j) = cont(i \wedge j) \text{ if } i \vee j = \top.$$

This property does not correspond to the usual additivity described in Equation 7.9. The latter says that the information content of the combination of two pieces of information equals the sum of the information contents of the single pieces of information, if their mutual information is 0.

$$i(\phi) + i(\psi) = i(\phi \otimes \psi) = i(\phi \otimes \psi) \text{ if } i(\phi || \psi) = 0.$$

Recall that mutual information is a measure of the relation between two pieces of information. It measures the amount of information we expect to obtain regarding one piece of information from observing the other one. Sometimes, the mutual information is also considered as a measure of (in)dependence between two pieces of information.

Bar-Hillel and Carnap define two sentences to be (inductively) independent in the standard way. The following definition corresponds to our Definition 7.27 of (stochastic) independence:

Definition 10.20 (Inductively Independent)

Two sentences i and j are said to be (inductively) independent if it holds for their belief-type probabilities that

$$m(i \wedge j) \quad = \quad m(i) \cdot m(j).$$

It would be desirable to deduce the usual requirement of additivity, $cont(i \wedge j) = cont(i) + cont(j)$, from the above definition. But unfortunately, this does not hold in general, as easily seen:

$$\begin{array}{rrcl} & m(i \wedge j) & = & m(i) \cdot m(j) \\ \Leftrightarrow & 1 - cont(i \wedge j) & = & (1 - cont(i)) \cdot (1 - cont(j)) \\ \Leftrightarrow & 1 - cont(i \wedge j) & = & 1 - cont(j) - cont(i) + cont(i) \cdot cont(j) \\ \Leftrightarrow & cont(i \wedge j) & = & cont(i) + cont(j) - cont(i) \cdot cont(j). \end{array}$$

Relative Content-Measure

Relative content-measure will now be determined. In (Bar-Hillel & Carnap, 1952), the relative content-measure of the sentence j with respect to the sentence i is defined as follows:

Definition 10.21 (Relative Content-Measure)

Consider two sentences i and j. The relative content-measure *of j with respect to i is the content-measure that j adds to the value already provided by the content-measure of i:*

$$cont(j|i) \quad := \quad cont(i \wedge j) - cont(i).$$

A reformulation of this definition by means of Definition 10.18 leads to

$$\begin{aligned} cont(j|i) &= cont(i \wedge j) - cont(i) \\ &= 1 - m(i \wedge j) - 1 + m(i) \\ &= m(i) - m(i \wedge j). \end{aligned} \tag{10.2}$$

The above definition is conform with Definition 10.12 of relative amount of information. In addition to the last six properties of Theorem 10.16, which hold by definition, it still has a bunch of interesting characteristics, resumed in the next Theorem. The proofs can be found in (Bar-Hillel & Carnap, 1952).

Theorem 10.22 *Consider two sentences i and j. The relative content-measure of j with respect to i, $cont(j|i)$, has the following properties:*

1. $cont(j|i) = cont(j) - cont(i \vee j) = m(i \vee j) - m(j)$.
2. $cont(j|i) = cont(j)$ *iff i and j are content-exclusive.*
3. $cont(j|i) \leq cont(j)$.
4. $cont(j|i) = cont(i \rightarrow j)$.
5. $cont(j|i) = m(i \wedge \neg j)$.

Property 1 states that, when first learning the sentence i and thereafter j, the relative content-measure can also be obtained by subtracting the content-measure of their disjunction from the content measure of j. A reformulation using the measure-function m is also given. If the two sentences have no content-element in common, the relative content-measure equals the one of j, as nothing that is asserted by j has already been asserted by i. Otherwise, the relative content-measure is strictly smaller than the content-measure of j itself (properties 2 and 3). From property 4 we learn the relation between absolute and relative content-measure. If i implies j, its absolute content-measure equals the relative content-measure of j with respect to i. This means that in a situation where i is given, the sentence j does not convey more information than $i \to j$, which is by itself a much weaker statement. Finally, the relative content-measure can also be seen as the inductive probability of $i \wedge \neg j$, as asserted by the last property.

The explicatum *cont* for amount of information proposed above just involves probabilities, but no use is made of the logarithm, as in the widely accepted approach from (Hartley, 1928; Shannon, 1948), where entropy is suggested as a measure of uncertainty leading to a measure of information. Since *cont* is not additive under independence, but under content-exclusiveness, Bar-Hillel and Carnap propose yet another explicatum for the concept of amount of information which will have this property, as explained below.

10.4.4 The Second Explicatum: Information-Measure (inf)

In Section 10.4.1 above, some requirements of a concept for amount of information are given. But it is also pointed out that Bar-Hillel and Carnap avoid to make a statement about the additivity of amount of information, as they consider two different formalisms, suitable for instantiating the amount of information concept. These two formalisms do not agree about additivity. In contrast to the first explicatum *cont*, presented in the previous section, the second one is conform with the usual sense of additivity, known from Equation 7.9. This formalism, denoted by inf, will be called information-measure or measure of information and is additive under independence, see Definitions 7.27 and 10.20. It may either defined by means of *cont*, as it is most of the time done by Bar-Hillel and Carnap, or directly using the m-function (Definition 10.17), as *cont* can also be traced back to this measure-function.

Definition 10.23 (Information-Measure inf)

Let $m(j)$ be the belief-type probability of the sentence j. Then,

$$\mathrm{inf}(j) \quad := \quad -\log m(j)$$

is a measure for $Cont(j)$, the content of sentence j, or any other adequate explicatum for $In(j)$, the semantic information carried by sentence j. This measure is called the information-measure *of j.*

Definition 10.23 may obviously be rewritten as

$$\mathrm{inf}(j) = -\log m(j) = \log \frac{1}{m(j)} = \log \frac{1}{1 - cont(j)}; \tag{10.3}$$

the last expression is Bar-Hillel and Carnap's favored one. However, we prefer the notation of Definition 10.23, as it is analogous to our quantitative measure of information proposed in Chapter 7, as well as to the consensus in information and communication theory. However, note that m is the belief-type probability. When Shannon's approach is followed, the frequency-type probability is used. See Section 10.1.1 for the difference between the two concepts.

There are various theorems for inf, some of them are mentioned in the following theorem:

Theorem 10.24 *Consider two sentences i and j. Their information-measures* $\inf(i)$ *and* $\inf(j)$ *have the following properties:*

1. $0 \leq \inf(i) \leq \infty$.
2. $\inf(i) = 0$ *iff i is true.*
3. $\inf(i) = \infty$ *iff i is false.*
4. $\inf(i)$ *is positive finite iff i is factual.*
5. *If $i \rightarrow j$, then* $\inf(i) \geq \inf(j)$.
6. *If $i \leftrightarrow j$, then* $\inf(i) = \inf(j)$.
7. $\inf(i \wedge j) \geq \inf(i) \geq \inf(i \vee j)$.
8. $\inf(i \wedge j) = \inf(i) + \inf(j)$ *if i and j are independent.*

The first four points of the above theorem state that the information-measure of a sentence i is never negative. In the case where i is a tautology, its information-measure is 0; in any other case, it is positive. If i is neither a tautology nor a contradiction, its information is measured by a finite positive value. Only if i is a contradiction, its measure is not finite any more, but $+\infty$. The fifth property states that a sentence i implying another sentence j provides a greater amount of information than the implied sentence j. If however the two sentences are equivalent, their information-measures equal, as stated by property 6. Property 7 points out that the information-measure of the conjunction of two sentences always results in a greater value than the information-measure of one of both sentences, which is again greater than the information-measure of their disjunction. Finally, the important additivity property states that the conjunction of two sentences which are independent in the sense of Definition 10.20 provides just as much information as the sum of the information-measures of the single sentences. In general, the information-measure of the conjunction of two sentences is however less than the sum of the inf-values of the single sentences. Only if their conjunction results in a contradiction, this sum is less than the information-measure of the conjunction. Clearly, Requirements 10.13 to 10.15 are fulfilled, as the following correspondences hold: Requirement 10.13 $\widehat{=}$ Property 5, Requirement 10.14 $\widehat{=}$ Property 2, Requirement 10.15 $\widehat{=}$ Properties 3 and 4.

Relative Information-Measure

As for *cont*, there is also a relative information-measure $\inf(j|i)$, determining how much information j adds to the information already expressed by i. It is defined in an analogous way:

Definition 10.25 (Relative Information-Measure)

Consider two sentences i and j. The relative information-measure *of j with respect to i is the information-measure that j adds to the value already provided by the information-measure of i:*

$$\inf(j|i) \quad := \quad \inf(i \wedge j) - \inf(i).$$

This definition agrees with Definition 10.12 of relative amount of information. As in the case of *cont*, it can be reformulated by applying Definition 10.23:

$$\begin{aligned} \inf(j|i) &= \inf(i \wedge j) - \inf(i) \\ &= \log m(i) - \log m(i \wedge j). \end{aligned} \tag{10.4}$$

Linking the Relative and the Absolute Information-Measure

In (Bar-Hillel & Carnap, 1952; Bar-Hillel & Carnap, 1953) not only the above definitions of relative information are given, but a further notion is introduced which links the relative information-measure to the conditional belief-type probability, as defined in (Carnap, 1950). Recall Carnap's definition of belief-type probability (which he calls "probability$_1$"):

> **Probability$_1$:** the logical concept of probability, degree of confirmation
>
> **Degree of confirmation:** a quantitative concept representing the degree to which the assumption of the hypothesis h is supported by the evidence e

It turns out that this degree of confirmation mentioned in the definition in (Carnap, 1950) is the conditional belief-type probability. He also calls it relative inductive probability or relative logical probability. As explained in Section 10.4.2, $m(j)$ is the belief-type probability of j on no evidence. The belief-type probability of j given i is $m(j|i) = \frac{m(i \wedge j)}{m(i)}$ and may also be described as the belief-type probability of j on the evidence i. $m(j|i)$ is thus the degree of confirmation of the hypothesis j by the evidence i, denoted as $c(j,i)$. As it holds that $c(j,i) = m(j|i) = \frac{m(i \wedge j)}{m(i)}$ and by applying Definition 10.23 to the definition of relative information-measure, it can be rewritten as follows:

$$\begin{aligned} \inf(j|i) &= \inf(i \wedge j) - \inf(i) \\ &= -\log m(i \wedge j) + \log m(i) \\ &= \log \frac{1}{m(i \wedge j)} + \log m(i) \\ &= \log \frac{m(i)}{m(i \wedge j)} \\ &= -\log \frac{m(i \wedge j)}{m(i)} \\ &= -\log m(j|i). \end{aligned}$$

The result

$$\inf(j|i) = -\log m(j|i) \tag{10.5}$$

is analogous to the Definition 10.23 of information-measure, $\inf(j) = -\log m(j)$. Note that there is no corresponding result for the content-measure *cont*, as

$$1 - m(j|i) \neq cont(j|i) = m(i) - m(i \wedge j).$$

This is due to the fact that $cont(j)$ just determines the complement of the belief-type probability of a sentence j, and is not involving the logarithm.

Independence

As a consequence of the above result, we obtain a further notable property of the relative information-measure, which sets it apart from its *cont* counterpart:

$$\inf(j|i) = \inf(j) \text{ iff } i \text{ and } j \text{ are inductively independent.} \tag{10.6}$$

There are some clear differences between *cont* and inf, even if they have not been completely revealed. Bar-Hillel and Carnap do however, or perhaps a fortiori, insist that the presystematic concept of amount of information has (at least) two explicata. This will be the topic of Section 10.4.5.

10.4.5 Two Explicata for Amount of Information

In Sections 10.4.3 and 10.4.4, Bar-Hillel and Carnap's two instances (explicata), which fulfill the requirements of their framework of Section 10.4.1, are described. The table in Figure 10.7 shows a juxtaposition of the two measures *cont* and inf. Only the crucial points are considered.

cont	inf
$cont(j) = 1 - m(j)$	$\inf(j) = -\log m(j)$
$0 \leq cont(j) \leq 1$ $cont(\perp) = 1$	$0 \leq \inf(j) \leq +\infty$ $\inf(\perp) = +\infty$
$cont(i \wedge j) = cont(i) + cont(j)$ if $Cont(i) \cap Cont(j) = \emptyset$	$\inf(i \wedge j) = \inf(i) + \inf(j)$ if $m(i \wedge j) = m(i) \cdot m(j)$
measure of substantial aspect	measure of unexpectedness (surprise value)
$cont(j|i) = m(i) - m(i \wedge j)$	$\inf(j|i) = \log m(i) - \log m(i \wedge j)$
$cont(j|i) \leq cont(i) + cont(j)$	no such property
$cont(j|i) \neq 1 - m(j|i)$	$\inf(j|i) = -\log m(j|i)$

Figure 10.7: Differences between *cont* and inf

Looking at the definition (first row), it turns out that *cont* is a purely probabilistic measure. By taking the complement to 1 of the probability of a sentence as its measure, our intuition is reflected: The more improbable an event (or sentence) j is, the more informative is its taking place. So a bigger content-measure is attributed to it than to a quite probable sentence. In the third row, the additivity characteristic proper to *cont* is stated. This property is very unusual, but it is – after some reflection – also justifiable. The content-measure of a conjunction of two sentences equals the sum of the single measures if the contents of the two sentences are disjoint, i. e. content-exclusive, as shown in Figure 10.4. Bar-Hillel and Carnap entitle *cont* as a measure of substantial aspect.

They are aware that the behavior of *cont* is at odds with the more customary additivity condition of independence (Definitions 7.27 and 10.20), but they justify their double-track approach by stressing that both, the additivity property related to content-exlusiveness, as well as the additivity property related to independence, are reasonable, but that no formalism can fulfill both properties at the same time:

> "Then, *cont* has certain plausible properties though it lacks certain other properties which are equally plausible. Since, however, no concept can have both theses plausible properties simultaneously, we are led to the idea – and there are many other arguments pointing in the same direction – that we do not have in our mind *one* clearcut, unique, presystematic concept of amount of information but at least two of them (and both still in the semantic dimension). This is not so strange. On the contrary, it is a rather common phenomenon that two related but different concepts are regarded as being identical although contradictory properties are required for their explicata." (Bar-Hillel, 1964)

The other measure proposed by the authors is inf. They point out that the definition of this second information-measure is a monotonic, but not linear transform of *cont*. It actually corresponds to the measure proposed by (Hartley, 1928) and has a game-theoretical background. The logarithm bounds the number of questions needed to determine an element from the set $Cont(j)$ and is thus a reasonable measure of the information provided by the sentence j. Its additivity property is well-known and used in every measure of amount of information which is in accordance with the Hartley-Shannon tradition. In contrast to *cont*, Bar-Hillel and Carnap indicate that inf is a measure of the unexpectedness or surprise value of some sentence.

A big advantage of *cont* is that it is mathematically very simple. The greater the probability of a statement, the smaller its content-measure. The information-measure inf is a little bit more complicated, as it uses the logarithm, but this turns out to be also an advantage when the relative information-measure is considered (see second point below). Furthermore, inf comes along with the usual additivity property based on independence. As inf is based on a question-game, it offers the possibility to illustrate the process of measurement by a decision tree where the base of the logarithm is the number of possible answers. At the same time, a code is generated. Bar-Hillel and Carnap remark that the information-measure inf has found a large field of application in communication engineering.[7] In (Bar-Hillel, 1964), two reasons are given why *cont* did not become widely accepted:

1. In communication engineering, the additivity property relative to independence (Definition 10.20) is much more important and practical than the additivity property relative to content-exclusiveness (last point of Theorem 10.19). The former is a feature of inf, the latter lies in the nature of *cont*, and Bar-Hillel adds that "it is doubtful whether this condition [i. e. the additivity property based on content-exclusiveness] makes sense there at all."

2. When looking at the table given in Figure 10.7, one notices that inf has a very appealing property with regard to relative information that *cont* does not have: $\mathrm{inf}(j) = -\log m(j)$ and $\mathrm{inf}(j|i) = -\log m(j|i)$, but the relative content-measure $cont(j|i)$ cannot be traced back to the absolute content-measure $cont(j)$. Bar-Hillel explains that most authors want the relative amount of information to be of the same function as the absolute amount of information. Such a requirement leads indeed to a log-type of function.

The above considerations lead to the conclusion that *cont* has interesting properties, but never became popular, due to the advantages that inf offers.

Carnap and Bar-Hillel finally point out that it is possible that there are further measures of information which correspond with the Requirements 10.13 to 10.15, but they are not considered by the authors.

10.5 Comparison: Information

After having introduced Bar-Hillel and Carnap's theory of semantic information in the foregoing sections, we will now compare it to our algebraic theory of semantic information. One can observe a gap between the weighting inside both theories: Bar-Hillel and Carnap do not look in great detail at information; they rather concentrate on the amount of information, whereas in our approach, it is of particular importance to understand the nature of information in order to draw conclusions about its (computational) nature. The measure of information, a problem discussed in many fields, is from our point of view a part of the algebraic theory of semantic information, but not of top priority. This is why this section about information is more extensive than the following section about measure of information. We will first compare both frameworks (Bar-Hillel and Carnap's presystematic concept of information and the information algebra framework as presented in the Chapters 4 and 5) from several points of view. Thereafter, the explicatum *Cont* is confronted with the information algebra instance predicate logic (see Chapter 9), involving the sample language proposed by Bar-Hillel and Carnap.

[7] They nevertheless point out that it is not inf in the version presented above (dealing with inductive probabilities), but its "statistical correlate", using a frequency-type probability, which is used.

10.5.1 Information on a Conceptual Level

Framework in General

The algebraic theory of information, as founded in (Kohlas, 2003), defines two basic operations for information processing on a set Ψ of pieces of information and a lattice D of questions. The operations are combination $\otimes$ and focusing $\Rightarrow$. A set of axioms is imposed on this two-sorted algebra, leading to a mathematical structure capturing the nature of information from an algebraic point of view. Thus the fundamental properties of information are described by the abstract framework of information algebra. This theory was outlined in Chapters 2 to 7; two formalisms which fit in the framework were thereafter presented (Chapters 8 and 9).

It is important to see that there is a clear difference between the framework as an abstract entity and the various instances that satisfy the information algebra axioms. Unfortunately, it is often the case that both things (the conceptual entity on the one hand and the concrete example on the other hand) are confounded.

In (Bar-Hillel & Carnap, 1952), however, the concept and its instance are also very well distinguished and we really want to stress this quality. The authors first present what they call "presystematic concept of semantic information". It has the same purpose as the abstract framework of information algebra and will be examined regarding its properties right now. The "explicatum" $Cont$, which is an instance of Bar-Hillel and Carnap's abstract theory of semantic information, is the topic of Section 10.5.2 below. Compared with the information algebra framework, which is characterized by Ψ, D, $\otimes$, $\Rightarrow$ and a set of five axioms, Bar-Hillel and Carnap's framework is less developed: In is used for capturing semantic information, but there is no set of all pieces of information. Questions are not considered at all, and operations on semantic information are not given explicitly. They are more or less borrowed from the logical background on which Bar-Hillel and Carnap's theory is founded. Only one axiom is provided by Carnap and Bar-Hillel, namely Requirement 10.3. We will see in this section that this requirement determines the the authors' view on semantic information, which is dual to the one we propose.

Bar-Hillel and Carnap also draw a clear distinction between the concept and the instance. But the algebraic theory of semantic information has the advantage of providing many different instances, whereas Carnap and Bar-Hillel only consider a reduced predicate logic as the only instance. There are no further instances which are related to other fields of application.

Semantic Information as a Set

Bar-Hillel and Carnap's presystematic concept of semantic information consists of a single requirement from which various theorems are derived (see Section 10.3). In other words, their framework has only one axiom, or information has one basic property, stated by Requirement 10.3:

$$In(i) \supseteq In(j) \quad \text{iff} \quad i \rightarrow j.$$

Bar-Hillel and Carnap use $In(i)$ to abbreviate "the information carried by statement i". The above requirement shows the first common point of the theories: They both treat information as a set of something. Earlier in this work we have explained that from our algebraic view on semantic information it is desirable to represent information in a set-theoretic way instead of the representation by a formula for example, and now it turns out that Bar-Hillel and Carnap are of the same opinion.

As already pointed out above, Bar-Hillel and Carnap do never consider all possible sentences in general. They do so when they look at an example, but they do not provide a set of all possible In-sets which would correspond to our set Ψ of pieces of information.

In the algebraic theory of semantic information, if some piece of information $\phi \in \Psi$ implies another piece of information $\psi \in \Psi$, the former is a subset of the latter: $\phi \subseteq \psi$. Bar-Hillel and Carnap's Requirement 10.3 cited above maintains the contrary, which corresponds to our dual view of information, taking into account the complement of each piece of information. If ϕ implies ψ, then $\phi^c \supseteq \psi^c$. It is the choice of this requirement which makes Carnap and Bar-Hillel's and our perception of information complementary.

Ordering Information

Another similarity is a further consequence of the above requirement: Information is ordered. In (Bar-Hillel, 1964), the author describes the situation regarding the instance *Cont* as follows, but it also holds for the conceptual entity *In*:

> "So long as we are talking only about [information] content itself, the most we can say is that a certain statement has a larger content than another one [...]" (Bar-Hillel, 1964)

Having a larger information content means providing more information. This leads, under certain conditions, to a partial order, which acts as a qualitative measure of information, stating whether some statement is more (or less) informative than another. However, as Bar-Hillel and Carnap do not consider a set containing all pieces of information, their study of the order of information is not as formal as the one we propose in Section 6.2, neither is it entitled "partial order" or "qualitative measure". But with the knowledge of Chapter 7, we can say that Bar-Hillel and Carnap's order leads to such a measure, even if the authors did not think of measuring information qualitatively.

Bar-Hillel and Carnap's order of information is expressed by set inclusion $\supseteq$. The elements making up a less informative statement j are by definition included in the elements constituting a statement i which provides more information:

$$i \geq j \quad \text{iff} \quad In(i) \supseteq In(j). \tag{10.7}$$

The authors never mention that this actually defines a partial order, which is not surprising. A partial order is defined in a set, but they do not consider a set of all possible In-sets. Therefore, no formal proof regarding the partial order $\geq$ is given. But it is well-known that set inclusion actually defines a partial order, see (Davey & Priestley, 2002), and thus the set of all such In-sets is partially ordered. This partial order goes (in terms of set inclusion) in the opposite direction of the partial order we have proposed in Section 6.2. We say that a piece of information ψ is more informative than another piece of information ϕ iff their combination yields the more informative one: $\psi \geq \phi$ iff $\psi \otimes \phi = \psi$. When looking at information from a set-theoretic point of view, as it should be done when dealing with semantics, the right hand side may be rewritten as $\psi \cap \phi = \psi$. This leads to the following representation of the partial order we have proposed earlier:

$$\psi \geq \phi \quad \text{iff} \quad \psi \subseteq \phi. \tag{10.8}$$

Comparing Equation 10.7 to Equation 10.8, the former has the clear advantage that it goes hand in hand with the cardinality of the set. Every statement j whose $In(j)$ carries only a subset of some $In(i)$ is clearly less informative than the statement i. This is not the case in our approach, where a more exclusive information is more informative. The reason for that difference lies in the two different representations of information. Even if, in both cases, sets are considered, Bar-Hillel and Carnap take into account the complement of what we call piece of information.

The minimum set min and maximum set max of all In-sets are of special interest. The minimum set is the In-set generated by a statement which is always true, a tautology. It is included in every other In-set and corresponds to the neutral element e in the information algebra framework. The maximum set is the In-set generated by a statement which is always false, a contradiction. It includes every other In-set and matches with the information algebra null-element z. Due to our inverse direction of the partial order, our neutral element e equals in the settings described in Chapters 8 and 9 the set of all possible valuations whereas the minimum set in the sense of Bar-Hillel and Carnap's theory would correspond to the empty set. However, in both cases the information described is a tautology. In the same manner, we express a contradiction in such settings by the empty set (as a contradiction does not have any models), which is our null-element z. Bar-Hillel and Carnap would represent this information by the In-set containing all possible elements. So the top element of their partial order is the bottom element of the ours and vice versa. The perception of the contradiction coincides in both approaches: It is the most informative statement in the partial order, with a wonderful explanation cited in Section 10.3.1. The following table gives an overview of the top and bottom elements in both partial orders:

	$\geq$	contradiction	tautology
Bar-Hillel and Carnap	$\supseteq$	min ($\bot$)	max ($\top$)
information algebra	$\subseteq$	z ($\top$)	e ($\bot$)

Operations

Focusing There is no such operation in Bar-Hillel and Carnap's presystematic concept of semantic information.

Combination Interestingly, Bar-Hillel and Carnap are not very much concerned with the combination of information. They do not introduce this operation separately; it seems that they take its existence for granted, which is certainly due to their logical background where such an operation is part of the basic equipment. Bar-Hillel and Carnap do not even point out the fact that combination, i. e. the conjunction of two statements i and j, leads to a new statement $i \wedge j$. This is in some way surprising, as they propagate a theory of semantic information. Nevertheless, the conjunction plays an important role as we will see right now when dealing with prior information.

Prior Information

Bar-Hillel and Carnap consecrate themselves to a more detailed investigation of the relativity of information. They point out that information and informativeness always depend on what is already known. So the authors make a difference between information in an absolute sense (without previous knowledge) and the information uttered in addition to some fact which was known before – the relative information. According to Requirement 10.3, which states that semantic information is to be treated as a set, they define the information which came up be learning j in addition to the statement i as

$$In(j|i) \quad = \quad In(i \wedge j) \setminus In(i),$$

see Definition 10.5. Prior information is perceived as a set, too, so consequently the set difference $\setminus$ is used. This above definition is due to their working in the complement (refer to Section 10.5.2 for more details). From Section 6.1.2 it is known that prior information can only be found in the algebraic theory of semantic information when information is measured. The concept of prior information constitutes the second principle of relativity of information in the algebraic theory of semantic information. As seen in Section 7.4, the measure of information relative to prior information involves the combination operation.

The concept of prior information does not exist per se in the algebraic theory of semantic information, only regarding the measure of information, whereas Bar-Hillel and Carnap consider prior information as an entity of its own. However, the important remark is that in both cases, the combination ($\wedge$ or $\otimes$, respectively) of the statements is involved.

Differences and Commonalities

Summarizing what we have seen so far, Bar-Hillel and Carnap's and our frameworks agree in the following points:

- A clear distinction between the conceptual framework and the instance(s) fitting in the framework is made.
- The nature of information is described by a requirement or a set of axioms.
- Semantic information should be and is treated as a set.
- Information is partially ordered.
- The contradiction provides the most information.
- Information is always relative to prior information.
- The concepts of prior information are both based on combination of information.

However, the information algebra framework is richer and larger as the "presystematic concept of semantic information" of Carnap and Bar-Hillel. The most important points that the information algebra framework provides in addition to Bar-Hillel and Carnap's theory of semantic information are the following:

- A set Ψ of pieces of information is defined on a general, abstract level.
- A lattice D of questions and an associated theory (see Chapter 3) is provided.
- There exists an operation of focusing for information extraction.
- The idempotent nature of information is pointed out.
- The algebraic properties of information are formalized.

It is a very important property of information that it relates to a question. In Chapter 3, questions and their associated lattice structures are discussed. Bar-Hillel and Carnap did not pay attention to this aspect at all. Furthermore, we point out that information cannot only be combined, but also focused on a certain field of interest. Such an operation does not occur at all in Carnap and Bar-Hillel's theory. But the focusing operation carries a lot of interesting qualities with it, namely a bunch of the information algebra axioms which state further properties of information. It furthermore enhances our theory with a customized inference automatism, by local computations, see Section 4.5. The sets Ψ of pieces of information and D of questions, together with the two operations, lead to an algebraic structure providing a formal, mathematical framework, which is not given in Bar-Hillel and Carnap's approach. Idempotency, which is a very fundamental property of information (perhaps even *the* property of information, as we get an information algebra from a valuation algebra by adding the idempotency axiom), is not really considered by Carnap and Bar-Hillel. Idempotency ensures that the combination of a piece of information with itself, or a coarser version of itself, gives nothing new. There are some implicit hints that the authors could have thought of something similar, when giving properties of the information carried by a conjunction of two statements or when introducing and describing relative information. But they never encapsulate the idea of idempotency.

After a first comparison of the frameworks, we will now turn to the instance *Cont* and its counterpart predicate logic described before.

10.5.2 Instance *Cont*

Bar-Hillel and Carnap do not only introduce a framework for semantic information, but also an instance of this framework (an "explicatum" in their terminology). This instance is the content of a statement i, abbreviated by $Cont(i)$, with a capital C. It was introduced in Section 10.3.2. In the previous section, we have seen that the framework presented by Carnap and Bar-Hillel covers certain aspects which are also provided by the information algebra framework, but the latter is more elaborated in terms of algebraic structure, operations and questions. We will see in this section that the information algebra counterpart of the instance *Cont* is predicate logic, as described in Chapter 9, and how *Cont* and predicate logic are related. It will furthermore turn out that our description of predicate logic extends the limited language system that is known from Section 10.2 and associated with *Cont*.

Valuations, State-Descriptions and Content-Elements

In order to refresh the notions "state-description" and "content-element" we will describe them using the vocabulary of predicate logic introduced in Sections 9.2 and 9.3. At the same time, this will show the existing relations and differences between Bar-Hillel and Carnap's and our approach.

State-Despcriptions To put it simply, a state-description has the same task as a the variable assignment function h_Σ of predicate logic, see Definition 9.7. However, things are more complicated. We have associated with the variable assignment function a so-called valuation (see Definition 9.8) which is a vector $\langle h_\Sigma(v_1)h_\Sigma(v_2)\ldots\rangle$ of the values that have been allocated to the variable v_j. Whereas a valuation is purely semantic (only values of the corresponding frame are provided), a state-description has still a very syntactic form, even if it has a semantic purpose. Bar-Hillel and Carnap define a state-description to be a *conjunction* of n Q-sentences (see below), where n is the cardinality of the finite set of variables Vbl of the considered language. So for every $v_j \in Vbl$, a Q-sentence is given, describing the state of the world for this very variable. A Q-sentence is a conjunction of all predicates of the language, occurring either negated or not negated. It is important to point out that the argument of the predicates in the Q-sentence relating to v_j is not the variable v_j itself (as it would be required by Definition 9.3), but $h_\Sigma(v_j)$, the value that has been ascribed to v_j, which is an individual constant and thus already a semantic component in a syntactic environment (conjunctions, negations, predicates). As a state-description is a conjunction of the Q-sentences for all variables considered, it completely supplies one possible state of the world. So one could say that a state-description is the syntactic form of a valuation. A state-description provides not only the values of the variables of the language, but also explicitly lists all possible predicates. Such a listing is, however, not necessary. In our approach, this is covered by the language $\mathcal{L}$. Therefore, a state-description has the same functionality and the same effect as a valuation; it has neither more nor less expressive power.

Content-Elements Now that the relation between our valuations and Carnap and Bar-Hillel's state-descriptions is clear, let us turn to an element introduced in (Bar-Hillel & Carnap, 1952) which does not appear in our theory, namely content-elements. From Definition 10.7 it is known that a content-element is a negated state-description. So it is once again a hybrid mix of syntax (connectives, predicates) and semantics (individual constants instead of variables), with a semantic purpose. Applying de Morgan's laws leads to an ad hoc definition of a Q-sentence in the content-element perspective, which we simply call Q^c-sentence.[8] A Q^c-sentence is the *disjunction* of all predicates of the language, occurring either negated or not negated. As in the case of Q-sentences, a Q^c-sentence always relates to one variable v_j, thus the arguments of its predicates are the individual constants obtained by $h_\Sigma(v_j)$. So we can reformulate the definition of content-elements by saying that a content-element is a *disjunction* of n Q^c-sentences, where n is the cardinality of the finite set of variables Vbl of the considered language.

So, what is the difference? From a semantic point of view, a valuation (or a state-description, which is conceptually the same) describes one possible state of the world by assuming an assignment of values (individual constants) to the variables and by stating which properties then hold simultaneously. Therefore, a valuation says the most that can be said about the world. A content-element on the other hand says the least that can be said about the world, besides tautology. As it is the negation of a state-description, a content-element excludes this state of the world to be possible. All the other states remain possible. So a single valuation provides information about the world by a detailed description of one state. It therefore excludes all other states of the world to be possible. A content-element provides a detailed description about how the world is *not*. By excluding one state of the world, it includes the remaining other ones.

Until now, we have only looked at single valuations (or state-descriptions) and single content-elements. But they usually do not appear alone; things get interesting when several of them are involved. Therefore, we will now consider sets of valuations, state-descriptions and content-elements.

[8] This definition is not given by Bar-Hillel and Carnap. We just introduce it to simplify the description of content-elements.

Set of Models, Range and Content

Valuations, state-descriptions and content-elements have all been introduced with the goal of expressing the semantic information carried by a predicate logic formula. This leads to the definition of a valuation ω being a model of a formula f iff $\omega \models_\Sigma f$, see Section 9.3. Such a valuation satisfying formula f gives one possible state of that world which is described by formula f. Thus the set

$$\hat{h}_\Sigma(f) = \{\omega : \omega \models_\Sigma f\}$$

of all valuations satisfying formula f, called set of models of f, is the information relative to the unknown values of the variables of f, or the information expressed by formula f. In Section 7.1, we presented two possible views on this set of models: On the one hand, the set of models is considered as a scheme of choice where one previously fixed, but not yet identified, valuation provides the actual values for the variables in the world described by formula f. In that case, the information is interpreted disjunctively, as the unknown valuation might either be ω_1 or ω_2 or ... for all $\omega_i \in \hat{h}_\Sigma(f)$. On the other hand, the set of models can be taken as an enumeration of all the possibilities described by formula f. This is the conjunctive interpretation of the information and the set of models states that one disposes of ω_1 and ω_2 and ... for all $\omega_i \in \hat{h}_\Sigma(f)$. The disjunctive view is much more usual than the conjunctive view, but it is important not to forget the latter.

Bar-Hillel and Carnap consider only the disjunctive view when dealing with state-descriptions. Actually, they do not say a lot about state-descriptions as they are only used for introducing content-elements which will serve as formalism for describing semantic information. Their conception of state-descriptions is the following:

> "Thus a state-description completely describes a possible state of the universe of discourse in question. For any sentence j of the system, the class of those state-descriptions in which j holds, that is, each of which implies j, is called the range of j. The range of j is null if, and only if, j is false; in any other case, j is equivalent to the disjunction of the state-descriptions in its range. [...] The sentence j says that the state of the universe (treated in $\mathcal{L}$) is one of the possible states which are described by the [state-descriptions] Z in $R(j)$." (Bar-Hillel & Carnap, 1952)

This is exactly the disjunctive view of information described above. Their notion of range $R(j)$ of some formula j (or sentence j, using their vocabulary) can be identified with our notion of set of models. Our disjunctive interpretation of the valuations in the set of models of a formula corresponds to Carnap and Bar-Hillel's equivalence between a sentence and the disjunction of the state-descriptions in its range. The same point of view is thus shared, but in our theory, it is made more explicit and the dual interpretation (conjunctive view) is proposed.

As already stated before, Bar-Hillel and Carnap are not so much concerned with state-descriptions and the range of some formula, but they attach great importance to content-elements and the content of a formula. Based on the idea that the state of the world, as described by a formula j, is one of the state-descriptions of $R(j)$ (i. e. one of the valuations in the set of models of j), the authors reason in the same way as we do that therefore the world is in none of the states $V \setminus R(j)$, where V is the set of all possible state-descriptions. In analogy to the above quotation, a formula j is not only implied by every state-description of $R(j)$, j also implies the negation of each state-description in $V \setminus R(j)$. This set corresponds to the set of content-elements which make up $Cont(j)$, the content of j. As a consequence, j is not only equivalent to the disjunction of the state-descriptions in $R(j)$[9], but also to the conjunction of the content-elements in $Cont(j)$[10]. Clearly, the information carried by a formula j is defined as the set of content-elements which are excluded by j. This is a dual point of view of information: On the one hand, Carnap and Bar-Hillel point out that the information described by the formula j is its set of models. (This view has nevertheless only been introduced with regard to the second point of view.) On the other hand, the authors stress that the information described by the formula j is given by all valuations which are excluded by j; this set is called $Cont(j)$, formally defined as

$$Cont(j) \quad = \quad V \setminus R(j). \tag{10.9}$$

The different ways of interpreting the range and the content of a sentence is linked to the syntactic form of state-descriptions and content-elements. A sentence j can be seen as the disjunction of the state-descriptions in its range $R(j)$. As a state-description has a conjunctive form, j is interpreted as a disjunction of conjunctions, when it is considered as a scheme of choice. By the application of the de Morgan laws, the sentence j can equally be seen as the conjunction of the content-elements in $Cont(j)$. As a content-element has a disjunctive form, j is interpreted as a conjunction of disjunctions, when it is considered as an enumeration. From this point of view, Bar-Hillel and Carnap deal with particular formulae j which can be connected to the notion of prime implicates and prime implicants known from (Haenni *et al.*, 2000).

Comparing the instance *Cont* of Bar-Hillel and Carnap's theory of semantic information to the predicate logic instance of our algebraic information theory leads to the consideration of Boolean information algebras and the associated duality theory (see Section 4.6). The domain-free information algebra (Ψ, D) associated with predicate logic is Boolean, so the set Ψ of pieces of information is a Boolean lattice. Consequently, there exists a further Boolean lattice which is the dual of the original one. Therefore, we dispose of a mapping $i : \Psi \to \Psi$, $\psi \mapsto \psi^c$, which is a Boolean isomorphism between the Boolean information algebra (Ψ, D) and its dual counterpart. This isomorphism is given by the negation on the level of syntax, or by the set complement on the level of semantics, in the sense of Lemma 9.25 and its proof. So to every piece of information ψ of the information algebra (Ψ, D) associated with

[9] As state-descriptions are conjunctions of predicates, this is a disjunction of conjunctions.

[10] As content-elements are disjunctions of predicates, this is a conjunction of disjunctions.

predicate logic, a dual piece of information ψ^c exists. If ψ is a formula, ψ^c is $\neg\psi$. If, however, ψ is a set of models,

$$\psi^c \quad = \quad \mathfrak{D} \setminus \psi, \tag{10.10}$$

where $\mathfrak{D}$ is the set of all valuations. Obviously, Equations 10.9 and 10.10 state the same. So Bar-Hillel and Carnap work from our point of view in the dual theory. They define semantic information in a way which is the dual of our definition. Thus the definitions are complementary and it is possible to capture their theory of semantic information with our algebraic information theory.

As to the partial order, the information order is the same, but the duality of sets of models and sets of content-elements introduces the opposite inclusion relation. The content-elements of a less informative statement i are included in the set of content-elements of a statement with larger content j:

$$i \leq j \quad \text{iff} \quad Cont(i) \subseteq Cont(j). \tag{10.11}$$

In Sections 6.2 and 9.5 we have defined the partial order between two pieces of information ϕ and ψ as $\phi \leq \psi$ iff $\phi \otimes \psi = \psi$. When considering pieces of information which are sets of models of predicate logic formulae, this partial order can be reformulated as

$$\phi \leq \psi \quad \text{iff} \quad \phi \supseteq \psi. \tag{10.12}$$

But it is also known that $\phi \supseteq \psi$ iff $\phi^c \subseteq \psi^c$, which corresponds to the above Equation 10.11, capturing Bar-Hillel and Carnap's partial order. Indeed, the same order is considered in both approaches.

Cont is exemplified using a simple predicate logic language. To conclude this section about the differences and commonalities regarding information in Bar-Hillel and Carnap's semantic theory of information and our algebraic theory of semantic information, we will now have a look at their sample language.

10.5.3 Sample Language

The language systems $\mathcal{L}_n^\pi$, made up of π predicates and n individual constants, proposed in (Bar-Hillel & Carnap, 1952) and presented in Section 10.2, are very restricted, as already pointed out. The authors there literally say:

> "The language-systems relative to which our theory of information will be developed are very simple ones, so simple indeed that the results to be obtained will be of only restricted value with regard to language-systems complex enough to serve as possible languages of science. The restriction, however, was partly imposed by the fact that inductive logic [...] has so far been developed to a sufficiently elaborate degree only for languages that are not much richer than those treated here, [...]"

As the measures of information that Carnap and Bar-Hillel propose are based on inductive logic, they also make their sample language rely on the state of the art of inductive logic at that time. Therefore, only a finite set of one-place predicates is considered. Besides the predicates of arity 1, the usual connectives and auxiliary symbols are part of the language system, as well as a finite set of individual constants. We are missing a clear differentiation between syntax and semantics, as predicates and connectives are part of the syntactic toolbox, whereas individual constants make up the semantics of a language. State-descriptons are also mixing syntax and semantics, as already illustrated above. Interestingly, there is no possibility to have variables in a $\mathcal{L}_n^\pi$-language. In other words, no distinction is made between variables and the values (individual constants) they can take. As a consequence, Bar-Hillel and Carnap's language systems lack in a structure $\Sigma = (F, R)$, see Definition 9.6. Here, the set R of relations specifies which individual fulfills which property. The other part of the structure, the set F of frames, is also absent, but this is due to the fact that the $\mathcal{L}_n^\pi$ language systems are not typed, i. e. the individual constants (and the missing variables) have no sort. Predicates with an arity greater than 1 would be desirable in order to have more expressive power. Finally, there are no quantifiers, which are notabene one of the most important elements of predicate logic!

A $\mathcal{L}_n^\pi$-language can be described by means of the elements introduced in Chapter 9. Our presentation of the language of predicate logic is however more comprehensive. It is therefore a generalization of their approach. Carnap and Bar-Hillel did not consider cases going beyond their sample language, a very restricted monadic predicate logic. We do not restrict ourselves to a finite set of variables, but present predicate logic dealing with a countable set of variables. In our case, predicates are not only unary, but n-ary and a sharp distinction is drawn between the syntax and the semantics of predicate logic. Furthermore, it is shown how predicate logic forms an information algebra. Thus this work may also be understood as an enhancement of Bar-Hillel and Carnap's work. The authors point out that their approach can be extended to more complex cases, but it was never done. However, we have seen that their concept of information is contained in the information algebraic point of view, so Chapter 9 on predicate logic continues their work.

10.6 Comparison: Measurement

Bar-Hillel and Carnap do not only describe the nature of (semantic) information, they also propose two different measurements for information. This second matter is actually the more important part in their theory, but it is less interesting for our purposes, as the algebraic theory of semantic information focuses more on the qualitative properties of information. For that reason this section, comparing both approaches of measurement, will be shorter than the foregoing one, which dealt with semantic information. This section is based on Sections 7.4 to 7.6 and Section 10.4. We will first address the concept of amount of information in general (Section 10.6.1) and look at some points which are important relative to our approach. Carnap and

Bar-Hillel propose two concrete measures of amount of information, *cont* and inf. We will start the comparison with the information-measure inf (Section 10.6.2), as it coincides with our approach of measuring information content. For reasons of completeness, some remarks are made about the content-measure *cont*, see Section 10.6.3.

10.6.1 The General Concept of Amount of Information

In (Bar-Hillel & Carnap, 1952), before proposing concrete measures of information, the authors make a couple of statements of how a measure of information should be. We will now compare their viewpoints to our idea of measure of information content.

Probability

Bar-Hillel and Carnap's theory of semantic information relies on inductive logic, which has been presented in detail in (Carnap, 1950). Inductive probability, as outlined in Section 10.1.1, is strongly involved in their measurements of amount of information. The authors stress several times that the theory is based on inductive probability and inductive logic. They use this argument to set themselves apart from other theories which use another probability, but the same measure-theoretical background, see also Section 10.6.2. Our approach is not as restrictive as theirs. It uses probability, whatever it is, as long at is conform with the Kolmogorov axioms.

Absolute and Relative Amount of Information

Right from the beginning, Bar-Hillel and Carnap point out that information can be measured absolutely, by just taking a piece of information (or a "sentence", using Bar-Hillel and Carnap's vocabulary) and determining its amount of information. The amount of information may be different, when some other piece of information is considered before. So the content of a piece of information has to be measured, *given* another piece of information. This corresponds to our second principle of information, see Section 6.1.2. In both approaches, the information measure relative to prior information is defined with respect to absolute information. Definition 10.12 (Carnap and Bar-Hillel) states that

$$in(j|i) \quad = \quad i(i \wedge j) - i(i),$$

and by our Theorem 7.21 (Chaining Theorem) we obtain

$$i(\psi|\phi) \quad = \quad i(\phi \otimes \psi) - i(\phi).$$

Thus we do not only agree with Bar-Hillel and Carnap about the necessity of measuring information relative to prior information. We also define this measure in the same way as they do, by referring to the amount of combined information.

The Requirements

Bar-Hillel and Carnap state three requirements for measuring the amount of information (Requirements 10.13 to 10.15). The second and the third requirement state together that the absolute amount of information is always ≥ 0. We completely agree with that, as seen in Section 7.4.1, where we point out that the measure of a piece of information, when initially nothing is known, results in a value ≥ 0. The first requirement is also met, even if it does not seem to be satisfied on first sight. Requirement 10.13 says that a piece of information must provide a numerically higher amount of information than another one if the latter is included in the former. In terms of $Cont$, $in(i) \geq in(j)$ means that the content-elements making up j are a subset of those of i:

$$in(i) \geq in(j) \quad \text{if} \quad Cont(i) \supseteq Cont(j). \tag{10.13}$$

We are postulating the opposite of Equation 10.13. From Section 7.4.1 one can conclude that

$$i(\phi) \geq i(\psi) \quad \text{if} \quad \phi \subseteq \psi, \tag{10.14}$$

where ϕ and ψ are given by their atoms. Clearly, Equations 10.13 and 10.14 differ in the direction of the set inclusion on the right hand side. However, when the two equations are reformulated using partial order, things become clearer. In Section 10.5.2 it was shown that the partial order proposed by Carnap and Bar-Hillel, given in Equation 10.11, is

$$i \geq j \quad \text{iff} \quad Cont(i) \supseteq Cont(j),$$

and our viewpoint of the partial order was cited in Equation 10.12 as

$$\phi \geq \psi \quad \text{iff} \quad \phi \subseteq \psi.$$

Thus, one can conclude that in both approaches the partial order (qualitative measure) is maintained by the quantitative, numerical measure:

$$\begin{aligned} in(i) \geq in(j) &\quad \text{if} \quad i \geq j, \\ i(\phi) \geq i(\psi) &\quad \text{if} \quad \phi \geq \psi. \end{aligned}$$

But Section 10.5.2 told us more than this, namely that Bar-Hillel and Carnap work from our point of view in the dual theory. So by looking again at Section 7.5, where our quantitative measure is presented from the dual perspective, one comes to realize that in as proposed by (Bar-Hillel & Carnap, 1952) corresponds to i^d, as given in Definition 7.10, because of $i^d(\phi) = i(\phi^c)$.

As already mentioned in Section 10.4, a requirement of additivity has been left out voluntarily by the authors, as the two concrete measures $cont$ and inf, which will be discussed below, do not share the same additivity property.

Differences and Commonalities

A general comparison between both approaches of measuring the amount of information may be summarized as follows: As from our point of view, Bar-Hillel and Carnap consider information in the dual algebra, their way of measuring information can be described by our dual measure i^d.

The common elements in both approaches are:

- The amount of information, when initial ignorance is assumed, is always a numerical value greater than or equal to 0.
- A tautology does not provide any information, its amount of information is 0.
- Information is measured relative to prior information, based on the measure of the amount of the combined information.
- The quantitative measure of information respects the qualitative measure (partial order).

However, there are also some important points in our approach, which are not covered by Bar-Hillel and Carnap's theory:

- Our theory provides two different interpretations (disjunctive and conjunctive) of the same piece of information and thus two different measures which share the same properties.
- Bar-Hillel and Carnap do not measure information relative to different questions, as there is no such thing as a question in their theory. This is nevertheless a very natural idea and an important extension of Carnap and Bar-Hillel's theory of semantic information.

10.6.2 Information-Measure

In (Bar-Hillel & Carnap, 1952), two concrete measures of information are proposed, one of them is inf, called information-measure, see Section 10.4.4. By Definition 10.23, the information of a sentence j is measured by

$$\inf(j) \quad = \quad -\log m(j),$$

where $m(j)$ is the belief-type probability of the sentence j, see Definition 10.17. This probability is computed by adding the respective probabilities of the state-descriptions in j's range. As content-elements are negated state-descriptions, it is also possible to obtain $m(j)$ by adding the probabilities of the content-elements in $Cont(j)$. Since our approach of measuring information works with any kind of probability (see above), $m(j)$ can be brought to the form of Equation 7.2:

$$m(j) \quad = \quad \frac{|Cont(j)|}{|C|},$$

where C is the set of all content-elements. Now it is easy to see that

$$\inf(j) = \log|C| - \log|Cont(j)| = -\log\frac{|Cont(j)|}{|C|} = -\log m(j). \tag{10.15}$$

So Bar-Hillel and Carnap's information-measure inf corresponds exactly to Hartley's approach of measuring information, which has been outlined in Section 7.4.1. Even if the authors always refer to Shannon and not to Hartley, they propose to measure information by reduction of uncertainty when initially, nothing is known. This approach is due to (Hartley, 1928). So we have seen that we share so far the same approach of measuring the information provided by a sentence or a piece of information. The measures are called inf and i, respectively, but in both cases, the idea of Hartley is implemented. Thus we completely agree with all the properties of inf listed in Theorem 10.24, they can be taken over to i one-to-one. Whereas Bar-Hillel and Carnap stop here, we go beyond. In Section 7.6, we show how to take probabilities into account in order to extend the measure towards entropy, as proposed in (Shannon, 1948). Such ideas cannot be found in Bar-Hillel and Carnap's theory.

10.6.3 Content-Measure

The other measure proposed by Bar-Hillel and Carnap is called content-measure, denoted by *cont*, see Section 10.4.3. It only relies on $m(j)$, as stated by its Definition 10.18:

$$cont(j) \quad = \quad 1 - m(j).$$

In Section 10.4.5, a comparison of *cont* and inf already took place. As we have seen in Section 10.6.2 above that inf and our measure i coincide, *cont* and our measure i have already been compared.

10.7 Conclusion

In their outline of a theory of semantic information, Bar-Hillel and Carnap present a framework where semantic information

- is looked at in a set theoretic way,
- can be partially ordered,
- can be combined and
- is to be considered relative to prior information.

As an instance of this framework, the authors provide a very restricted monadic predicate logic, where a piece of information is the set of models of the formula describing the information. Clearly our information algebra framework goes beyond Bar-Hillel and Carnap's framework of semantic information, but the two approaches are not contradictory; the latter might be seen as a subset in properties of the former. By comparing their monadic predicate logic to our information algebra instance, it turns out that Bar-Hillel and Carnap do not consider information in exactly the same way as we do, but they work in the dual algebra, which leads to dual results.

In a second step, Bar-Hillel and Carnap propose again a framework, but this time for the measure of semantic information, followed by two concrete measurement functions fitting in their framework. Concerning the framework, the same situation as before comes up, namely that their idea of measuring information is conform with ours in many points, but their theory lacks some elements which we consider essential for a good measure of information. As to the measurement functions, we completely agree in the one which goes back to Hartley; the other one (which has never been widely accepted, as the authors point out themselves) makes also sense, but does not seem to be useful for practical purposes.

11

Groenendijk and Stokhof's Semantics of Questions

Il est encore plus facile de juger de l'esprit d'un homme par ses questions que par ses réponses.

Pierre-Marc-Gaston, duc de Lévis (1764 – 1830)
Maximes et réflections sur différents sujets de morale et de politique

In 1984, Groenendijk and Stokhof published a joint Ph.D. thesis, entitled "Studies on the Semantics of Questions and the Pragmatics of Answers". It contains six studies on different subjects in the theory of questions and answers, written over a period of several years. The authors have originally come from the pragmatics of natural language and they are mainly interested in its epistemic aspects with regard to a general theory of meaning and understanding. The motivation for their studies is that in order to get a proper pragmatics, one needs a proper semantics. For this chapter, we have chosen the fourth paper of (Groenendijk & Stokhof, 1984), having the same title as the whole book and thereby hinting at its central position. When referring later on to Groenendijk and Stokhof's publication in 1984, we actually only mean this fourth chapter "On the Semantics of Questions and the Pragmatics of Answers" which is first and foremost about questions and question-answering. This implies, however of secondary importance, information theoretical aspects. The authors' view on information is probably best captured by the following quotation from (Groenendijk & Stokhof, 1984):

> "Information should be a crucial notion in any acceptable theory of question-answering. Whether a piece of information, a proposition, provides an answer to a question of a certain questioner, depends on the information it conveys and on the information the questioner already has. This makes the notion of answerhood essentially a pragmatic one. But no pragmatics without semantics. It is not information as such, but only information together with the semantics of a question, that determines whether a proposition counts as a suitable answer."

It is important to point out that Groenendijk and Stokhof talk of questions and answers not as syntactic objects (which they call "linguistic"), but they look at the semantic, model-theoretic objects that give an interpretation to the language-theoretic objects usually considered. This semantic flavor sets the authors apart from the established approaches and at the same time makes their theory highly interesting for our purposes.

More than 10 years after the publication of their joint Ph.D. thesis, Groenendijk and Stokhof contributed an article about questions to the Handbook of Logic and Language (Groenendijk & Stokhof, 1997), where the former ideas are taken up and refined, particularly with regard to two concrete fields of applications, namely propositional and predicate logic.

Thereafter, in the year 2000, van Rooij wrote a paper completing the above theory with a qualitative and a quantitative measure of informativity. It was published only 9 years later, in a collection about formal theories of information (van Rooij, 2009).

The three articles mentioned above are the basic elements of this chapter. Not their whole content is presented in this chapter, but only those parts which are important with respect to the algebraic theory of semantic information, introduced in the first part of this thesis. Further articles that have contributed to the overall picture of this chapter are (Groenendijk & Stokhof, 1991; Groenendijk, 1999; Stokhof, 2002; Groenendijk, 2003). After an explanation of the foundations of Groenendijk and Stokhof's theory (Section 11.1), we will have a look at their perception of questions in Section 11.2. This general idea is exemplified by two applications concerning propositional and predicate logic, in Section 11.3. As already pointed out above, answers are closely related to questions, so Section 11.4 is about answers, their order and measure is looked at in Section 11.5. As a consequence of the informativity of answers, a quantitative measure of the informative value of questions is introduced in Section 11.6, as proposed in (van Rooij, 2009). Finally, this chapter closes with a comparison between the presented theory of questions and answers and our algebraic theory of semantic information (Sections 11.7, 11.8 and 11.9).

11.1 Foundations

Groenendijk and Stokhof trace their theory of questions back to many other approaches, but two authors are of special importance: Hintikka and Hamblin.

11.1.1 Hintikka

In the 1970s, Hintikka came up with a theory of questions and answers (Hintikka, 1974; Hintikka, 1976; Hintikka, 1978), working with speech acts that are performed by uttering a question.[1] He was the first to incorporate questions - and therefore

[1] This is already the reason for not having included Hintikka's approach in this third part of the thesis. Speech acts are completely different of our algebraic way of describing semantic information

answers - in a theory which describes the nature of information and which captures the meaning of pieces of information.

11.1.2 Hamblin

Groenendijk and Stokhof's view of questions (and answers) is strongly influenced by Hamblin, and so is ours, as already pointed out several times before, even by the quotation from (Hamblin, 1973) preceding this thesis. From (Hamblin, 1958), Groenendijk and Stokhof derive three general principles of questions as semantic objects. The postulates, taken from (Groenendijk & Stokhof, 1997), provide a methodology of reducing questions to answers:

1. *An answer to a question is a sentence, or a statement.*
2. *The possible answers to a question form an exhaustive set of mutually exclusive possibilities.*
3. *To know the meaning of a question is to know what counts as an answer to that question.*

In the first postulate, Groenendijk and Stokhof point out the main purpose of an answer, namely providing information. It should be seen as a statement, in a semantic way and not be reduced to its syntactic form. The second postulate is about the main characteristics of those statements which may act as the answer to a question. If one answer is selected, the others become impossible answers; they are mutually exclusive. Each possible answer provides the whole information asked for; the set of answers that a question allows is exhaustive. Groenendijk and Stokhof deduce from the first two postulates that the possible answers to a question form a partition. Finally, the third postulate identifies the meaning of a question with the set of its possible answers as semantic objects. So the meaning of a question can be traced back to the partition it induces. These characteristics of (the meaning of) a question will be looked at in detail in the following sections.

11.2 Questions

The idea behind Groenendijk and Stokhof's semantic theory of questions is to trace it back to the meaning of its possible answers. According to the Hamblin quotation in the beginning of this thesis, the semantics of a question is given by a set of those propositions which are the possible answers to the question. Here, the term proposition has to be interpreted as in the context of possible worlds semantics. In such a setting the conditions or circumstances under which some statement ϕ is true are called possible worlds. The meaning of the statement ϕ is thus its set of

---

and questions that a comparison would not have been possible. Note, however, that he is one of the rare persons using many-sorted predicate logic, which reflects the characteristics of variables in a more natural way.

possible worlds. This set is known as the proposition expressed by a statement.[2] Groenendijk and Stokhof are not so much interested in the syntactic form of the statement ϕ, but in its meaning. In order to avoid additional notational complexity, we will also denote the meaning of a statement, i. e. the set of possible worlds, as well as the statement itself, by a Greek lower case letter like ϕ, ψ, ζ. Questions are denoted by Latin lower case letters like q, r, s and describe a set of propositions.[3]

11.2.1 Partition

In accordance with the Hamblin quotation, Groenendijk and Stokhof see in a question a partition (see Definition 3.7) of what they call the logical space. The logical space corresponds to what is usually called universe and therefore denoted by U. It describes all circumstances which could occur and have effect upon the truth value of all possible statements which can be uttered. Formally, the logical space U is the set of all possible worlds u. In (Groenendijk & Stokhof, 1984), this set of all possible worlds is called set of indices. Groenendijk and Stokhof actually identify a question q with a partition of U and refer to the partition as U/q. Hence, each block of U/q is a set of possible worlds. In other words, a block is a proposition, which is an answer to the question q, described by the equivalence class $[u]_q = \{v \in U : v \text{ gives the same answer to } q \text{ as } u\}$. When we want to highlight that an equivalence class $[u]_q$ of U/q is an answer to the question q, it will be denoted by a Greek lower case letter like ϕ. The family of equivalence classes

$$U/q \quad = \quad \{[u]_q : u \in U\} \tag{11.1}$$

can be regarded as the semantic object expressed by the question q.

Groenendijk and Stokhof distinguish between two types of questions: yes-/no-questions and constituent questions. A question of the first type, a yes-/no-question, has only two possible answers. So it makes a bipartition on the logical space, as illustrated on the left hand side of Figure 11.1. The second type of questions, constituent questions, can be regarded as an n-fold partition of U, where n depends on the relation underlying the predicate asked for, see the right hand side of Figure 11.1.

11.2.2 Order

Let $Part(U)$ denote the set of all partitions, i. e. of all questions. Groenendijk and Stokhof define an order relation between questions in $Part(U)$. Furthermore, some operations on questions are provided. The order relation and the operations are now introduced one at a time. Their properties are discussed. Finally, it is concluded that

[2] Several times Groenendijk and Stokhof point out that such a set of possible worlds corresponds to a set of models in the sense of Definitions 8.5 and 9.13. We will look at these two special cases in Section 11.3.

[3] This is not the notation which can be found in all of the publications of Groenendijk and Stokhof and van Rooij, as they use different notations over the years. We now fix one notation for the rest of the chapter, which will hold for questions as well as for answers.

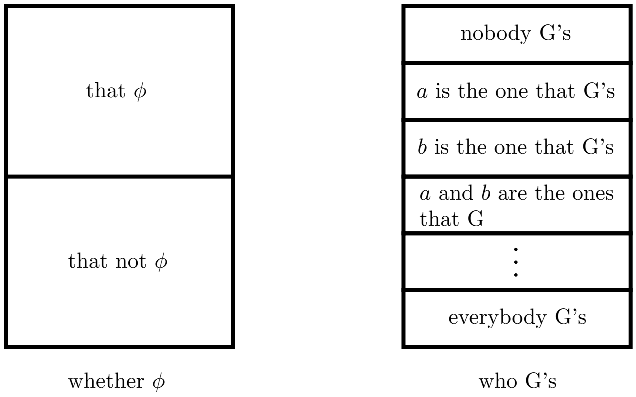


Figure 11.1: Partitions of the logical space

questions form a lattice. All illustrations are taken from (Groenendijk & Stokhof, 1984).

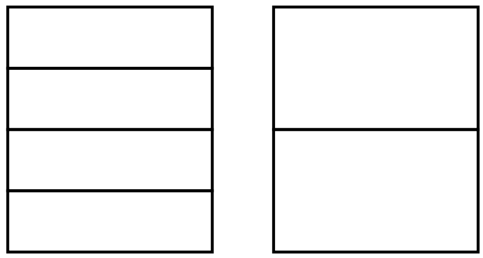
Inclusion Relation Between Two Partitions

The set $Part(U)$ of all questions can be ordered by means of the inclusion relation $\sqsubseteq$ which is closely related to the cardinality of a question. The cardinality of a question q equals the number of possible answers to it. As each answer to q corresponds to a block of the partition U/q, the cardinality of q is the number of blocks in U/q. A question q has the lowest possible cardinality if it consists of only one block, i. e. $U/q = \{U\}$. Such a question with cardinality 1 is called the tautological question in U. The only answer that can be given to such a question is a tautological statement. Groenendijk and Stokhof argue that a question which has only one possible answer is not a proper question at all. However, it is important for the partial order that will be defined by means of the inclusion relation between partitions:

$$U/q \sqsubseteq U/r \quad \text{iff} \quad \forall \phi \in U/q \; \exists \psi \in U/r \text{ such that } \phi \subseteq \psi. \tag{11.2}$$

One can paraphrase this inclusion relation between two questions q and r as follows: q implies or entails r iff every answer ϕ to q implies a unique answer ψ to r. Therefore, it is also named implication or entailment relation between two questions. Furthermore, Groenendijk and Stokhof point out that $U/q \sqsubseteq U/r$ means that U/q is a refinement of U/r, as illustrated in Figure 11.2.

Groenendijk and Stokhof point out that $\sqsubseteq$ is a reflexive, antisymmetric and transitive relation, a fact which is well-known in the theory of partition lattices. So $\sqsubseteq$ gives rise to a partial order on $Part(U)$. By means of the operations $\sqcup$ and $\sqcap$, introduced below, which satisfy idempotency, commutativity, associativity and absorption, $Part(U)$ even forms a complete lattice under $\sqsubseteq$ with $\sqcap$ and $\sqcup$ being meet and join.



$U/q \sqsubseteq U/r$

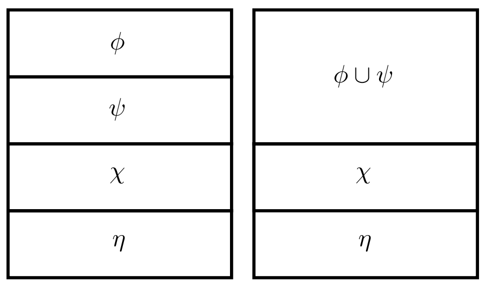
Figure 11.2: Inclusion relation between two partitions

Union Inside a Partition

The union-operation $\bigsqcup$ inside a partition is, according to Groenendijk and Stokhof a one-place operation. In fact, it is a three-place operation, taking the union of two specific blocks of an indicated partition. It is specified for a partition U/q and two of its blocks ϕ, ψ. The result $\bigsqcup_{\phi,\psi} U/q$ is a new partition, where the two answers ϕ, ψ form a single answer. (It will turn out in Section 11.7.2 that this is a coarsening.)

$$\bigsqcup_{\phi,\psi} U/q \quad = \quad \{\zeta : \zeta = \phi \cup \psi \text{ or } \zeta \in U/q, \zeta \neq \phi, \zeta \neq \psi\} \tag{11.3}$$

So the new partition $\bigsqcup_{\phi,\psi} U/q$ consists of the union of the possible worlds in the two specified blocks ϕ and ψ and of all remaining blocks of the original partition U/q, which are neither ϕ nor ψ.



U/q $\quad$ $\bigsqcup_{\phi,\psi} U/q$

Figure 11.3: Union of two blocks of one partition

Intersection Between Two Partitions

The intersection-operation $\sqcap$ on two partitions results in a new partition. It is a two-place operation taking the non-empty intersections of all the blocks of the two partitions on which it operates:

$$U/q \sqcap U/r \quad = \quad \{\phi \cap \psi : \phi \in U/q, \psi \in U/r, \phi \cap \psi \neq \emptyset\} \tag{11.4}$$

The intersection of two partitions U/q and U/r gives rise to a new partition $U/q \sqcap U/r$ whose non-empty blocks result from the intersection of every block of U/q with every block of U/r. Intersecting two questions means asking them simultaneously. The operation $\sqcap$ is the meet in the partition lattice that will be established below.

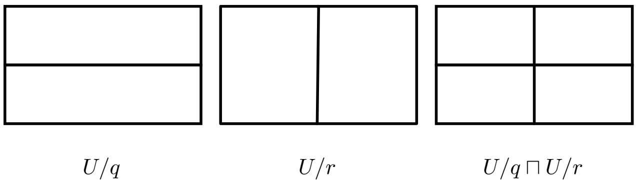


Figure 11.4: Intersection of two partitions

Union Between Two Partitions

There is not only a union-operation inside one partition, but also a union-operation $\sqcup$ on two partitions, which is in contrast to the former two-place. It results in a new partition $U/q \sqcup U/r$, expressing the common part of the two involved questions q and r. Its definition in (Groenendijk & Stokhof, 1984) is difficult to understand and will therefore not be reproduced here. However, the operation can sketched as resulting in a partition containing the largest blocks of the original partitions q and r. It is easy to realize that the operation $\sqcup$ is the join of the lattice of questions that will now be established.

Lattice of Questions

$Part(U)$ is not only partially ordered by means of the inclusion relation $\sqsubseteq$. Groenendijk and Stokhof stress that $Part(U)$ even forms a complete lattice, where $\sqcap$ is meet and $\sqcup$ is join. Some properties of $\sqsubseteq, \sqcap$ and $\sqcup$ are listed in the following theorem:

Theorem 11.1 *Let $U/q, U/r$ be any two partitions in $Part(U)$, the set of all possible partitions of U. It holds that*

1. $U/q \sqsubseteq \{U\}$, *for all* $U/q \in Part(U)$,
2. $\{\{u\} : u \in U\} \sqsubseteq U/q$, *for all* $U/q \in Part(U)$,

3. $U/q \sqcap U/r \sqsubseteq U/q$,

4. $U/q \sqsubseteq U/r$ *iff* $U/q \sqcap U/r = U/q$,

5. $U/q \sqsubseteq U/q \sqcup U/r$,

6. $U/q \sqsubseteq U/r$ *iff* $U/q \sqcup U/r = U/r$.

The tautological question $\{U\}$ is its maximal element (property 1 of Theorem 11.1), it is the least demanding question. The minimal element of the lattice is $\{\{u\} : u \in U\}$ (property 2 of Theorem 11.1). It is the most demanding question, asking everything that can be asked. $\sqcap$ and $\sqcup$ are meet and join. For the sake of simplicity, we will use $\sqsubseteq, \sqcup$ and $\sqcap$ as operations on partitions, but also as operations on the questions having induced the partitions. In particular, we will write $q \sqsubseteq r$ to express that q is a refinement of r which means that q is a finer question (providing finer answers) than r.

11.3 Questions in Propositional and Predicate Logic

In their early works in the 1980s, Groenendijk and Stokhof are concerned with the semantics of questions in general as shown in Section 11.2. Even if the authors have a background in logics, they do not refer to a concrete, underlying formalism, which would provide a syntax to the questions and answers they are dealing with. But in (Groenendijk & Stokhof, 1997), the point of view has changed: Two formalisms for expressing information have been chosen, propositional and predicate logic. It lies in the nature of propositional logic that it only allows yes-/no-questions. Predicate logic is an extension of the former in the sense that it also allows constituent questions. The starting point of their argumentation is the syntax of the respective formalism. The semantics of a question is in both cases traced back to the possible answers of the question. They are, however, in a first step, considered in a syntactic way. Only in a second step, the semantics of the possible answers is also taken into account, leading to a bipartition or an n-fold partition of the logical space. Depending on whether yes-/no-questions or constituent questions are considered, the procedure of slightly different, but the idea is the same.

We will first look at the case of propositional logic which is restricted to yes-/no-questions (Section 11.3.1). Thereafter, predicate logic is considered in Section 11.3.2, allowing also constituent questions. Groenendijk and Stokhof use a notation which is different from and goes beyond the one we have introduced in Chapters 8 and 9. As the aim of this thesis is not only to compare their theory with the ours, but also to present their theory at first, we decided to make use of Groenendijk and Stokhof's notation, but to state at each step to what this corresponds in the notation known from Chapters 8 and 9. The most important result to retain is surely that, in the case of propositional and predicate logic, Groenendijk and Stokhof speak of a question as being a set of sets of models.

11.3.1 Propositional Logic

When information is expressed by propositional logic, the possible answers to a question are formulae whose syntax is given in Section 8.1. Groenendijk and Stokhof use a notation different from that of Section 8.2 to describe their semantics. We will now introduce their notation for the semantics of statements and give an example. Thereafter, the semantics of questions is looked at, also illustrated by an example.

Semantics of Statements

As before, a formula ϕ is made of propositions which can either be true or false, determining whether the whole formula evaluates to true or to false. Ascribing a truth value of $\{0,1\}$ to each proposition, where 0 means false and 1 means true, results in a setting called possible world. A possible world thus corresponds to the valuation mapping v of Definition 8.3 and can be thought as a 0/1-vector, providing a truth value for every proposition. Groenendijk and Stokhof denote the set of all such vectors, i. e. the set of all possible worlds, by M and call it a model, whereas in Section 8.2, model has another meaning. A possible world $w \in M$ assigns a truth value to a statement which is given by a formula ϕ. Groenendijk and Stokhof call this truth value of $\{0,1\}$ the extension of the statement ϕ and denote it by $[\phi]_{M,w}$. It corresponds to the truth assignment $\hat{v}$ from the set of formulae to $\{0,1\}$ of Equation 8.1 and Definition 8.4. This allows to define the intension of the statement given by the formula ϕ. It is the set of those possible worlds $w \in M$ in which ϕ evaluates to 1 (true). This set is denoted by $[\phi]_M$ and defined to be $\{w \in M : [\phi]_{M,w} = 1\}$. In terms of Section 11.2, $[\phi]_M$ is the proposition expressed by ϕ. Using our notation introduced in Section 8.2, $[\phi]_M$ is ϕ's set of models $\mathcal{M}(\phi)$, see Definition 8.5. The table given in Figure 11.5 provides a summary of the corresponding terms:

notation	name	explanation	correspondence in Section 8.2
w	possible world	valuation, 0/1-vector	v
M	model	set of all possible worlds	$\{0,1\}^\omega$
$[\phi]_{M,w}$	extension of ϕ	value of $\{0,1\}$ assigned by $w \in M$	$\hat{v}$
$[\phi]_M$	intension of ϕ	$\{w \in M : [\phi]_{M,w} = 1\}$	$\mathcal{M}(\phi)$

Figure 11.5: Linking Groenendijk and Stokhof's propositional logic notation to the one introduced in Section 8.2

The notions of extension and intension, which were introduced for statements so far, are now illustrated by an example.

Example 11.3.1 (Extension and Intension of a Statement) Consider a simple propositional language $\mathcal{L}_P$, with $P = \{p_1, p_2\}$. Since there are two propositional symbols, the set M of all possible worlds is constituted of four 0/1-vectors: $M = \{(0,0),(0,1),(1,0),(1,1)\}$. We will now specify the extension and the intension of the formula $\phi = p_1 \vee p_2$.

- The extension of $\phi = p_1 \vee p_2$, relative to the world $w = (1,0) \in M$, is the truth value 1: $[p_1 \vee p_2]_{M,(1,0)} = 1$.
- The extension of $\phi = p_1 \vee p_2$, relative to the world $w = (0,0) \in M$, is the truth value 0: $[p_1 \vee p_2]_{M,(0,0)} = 0$.
- The intension of $\phi = p_1 \vee p_2$, relative to the model M, is a set of worlds: $[p_1 \vee p_2]_M = \{(0,1),(1,0),(1,1)\}$.

$\odot$

Semantics of Questions

The semantics of questions can now be defined by making use of the above knowledge about the semantics of statements, which act as possible answers to a question. From Section 11.2, it is known that a question q is identified with the set of its possible answers. In (Groenendijk & Stokhof, 1997), the authors take one of these answers in its *syntactic* form. This syntactic answer to q is a formula, denoted by ϕ. In order to get in touch with the semantics of the question q, the set of possible worlds in which the selected ϕ evaluates to 1 is taken. This is $[\phi]_M$. As questions in propositional logic give only rise to the answers "yes" and "no", the bipartition of the logical space M is obvious: One block is formed by the possible worlds of $[\phi]_M$, the second block is given by $M \setminus [\phi]_M$. Considering a question q as being given by its semantic answers, leads to the set $q = \{[\phi]_M, M \setminus [\phi]_M\}$, where ϕ is one possible syntactic answer to q.

The above paragraph is a short résumé of Groenendijk and Stokhof's way to arrive at the bipartition of M. For the sake of completeness, we will now give their detailed approach, involving the extension and the intension of the question q. A possible world w of the model M is fixed, and an answer ϕ to q is chosen in its syntactic form. An extension $[?\phi]_{M,w}$ of the question q in the world w regroups all worlds $w' \in M$, under which the syntactic answer ϕ evaluates to the same truth value as under w:

$$[?\phi]_{M,w} \quad = \quad \{w' \in M : [\phi]_{M,w'} = [\phi]_{M,w}\} \tag{11.5}$$

$[?\phi]_{M,w}$ is not unique, it depends on the answer ϕ and the chosen w. It is an intensional object, i. e. a proposition which is a subset of the set M of all possible worlds. An extension $[?\phi]_{M,w}$ is the proposition expressed by ϕ if the fixed w is part of the intension of ϕ, and the proposition expressed by $\neg\phi$, otherwise. Formally,

$$[?\phi]_{M,w} = \begin{cases} [\phi]_M & \text{if } w \in [\phi]_M, \\ [\neg\phi]_M & \text{else.} \end{cases} \tag{11.6}$$

So an extension of a yes-/no-question is the proposition expressed by one of the answers to that question. It depends on the answer chosen and therefore, there are always two possible extensions in the propositional logic case. The introduction of the extension of a question q paved the way for the definition of the intension of a question q in the model M. Groenendijk and Stokhof identify the intension of q in a model M with the set of its possible extensions in M:

$$q \quad = \quad \{[?\phi]_{M,w} : w \in M\}. \tag{11.7}$$

The propositions in this set are the two listed above in Equation 11.6, they are mutually exclusive, exhaust the logical space, which consists of all possible worlds in M, and are therefore all possible answers to q. As Groenendijk and Stokhof identify questions with partitions, q makes a partition on M. The blocks of the partition are the propositions $[?\phi]_{M,w}$, which in turn correspond to the equivalence classes $[u]_q$ of Equation 11.1.

We have seen that the notions of extension and intension are also defined for questions. They are now illustrated by an example.

Example 11.3.2 (Extension and Intension of a Question) Reconsider the simple propositional language $\mathcal{L}_P$, with $P = \{p_1, p_2\}$ and $M = \{(0,0), (0,1), (1,0), (1,1)\}$ of Example 11.3.1. For determining the extension and the intension of a question q, a syntactic answer ϕ to q is taken. Here, $\phi = p_1 \vee p_2$. The extension of the question q now depends on $\phi = p_1 \vee p_2$, as well as on a selected world $w \in M$:

- The extension of q, relative to the syntactic answer $\phi = p_1 \vee p_2$ and the world $w = (1,0) \in M$, is a set of worlds: $[?p_1 \vee p_2]_{M,(1,0)} = \{(0,1), (1,0), (1,1)\}$. This corresponds to the intension of the statement $p_1 \vee p_2$, denoted by $[p_1 \vee p_2]_M$, of Example 11.3.1.
- The extension of q, relative to the syntactic answer $\phi = p_1 \vee p_2$ and the world $w = (1,0) \in M$, is a set of worlds: $[?p_1 \vee p_2]_{M,(0,0)} = \{(0,0)\}$.

The intension of the question q in the model M is the set of its possible extensions in M. In propositional logic, a question has only two possible extensions. They are given above.

- The intension of q, relative to the model M, is a set of sets of worlds: $[?p_1 \vee p_2]_M = \{\{(0,1), (1,0), (1,1)\}, \{(0,0)\}\}$.

The intension of a question q shows the partition that q makes on M. A block of the partition is a set of possible worlds. $\odot$

11.3.2 Predicate Logic

Predicate logic is a formalism allowing to ask more sophisticated questions as in the case of propositional logic, where the answers are restricted to “yes” and “no”. Groenendijk and Stokhof distinguish two different kinds of constituent questions, namely one-constituent ones and multiple-constituent ones. One-constituent questions ask for the actual denotation of a particular property. As an example, the authors give "Which students passed the test?", which is typically answered by listing the students that actually fulfill this property. Multiple-constituent question ask for the actual denotation of a relation of interest, such as in "Who plays with whom?". So predicate logic extends the questions we have seen so far by inquiring also after objects which have certain properties and stand in certain relations.

This presentation of questions in predicate logic is a summary of the approach found in (Groenendijk & Stokhof, 1997), influenced by (van Rooij, 2009). In predicate logic, the possible answers to a question are formulae, whose syntax is given in Section 9.2. Again, we start with the presentation of a notation (different from the one proposed in Section 9.3) for the elements of predicate logic semantics, illustrated by an example. Thereafter, a special syntax for questions, used by Groenendijk and Stokhof, is introduced and exemplified. Finally, the semantics of questions is given. Here, the case of a fixed structure, as proposed in (van Rooij, 2009), has been chosen, since only this situation is useful for the comparison of the two theories in Section 11.8.

Semantics of Statements

Syntactically, a predicate logic formula ϕ consists of predicates, which involve a certain number of variables, and in addition quantifiers, negation symbols and the usual connectives. For the semantics of predicate logic, we refer to the approach of (van Rooij, 2009). It is very similar to the one of Groenendijk and Stokhof for propositional logic shown above. Their approach for the semantics of predicate logic will not be taken into account here, as it involves different structures. Such a setting does not correspond to our situation of a fixed structure, which is also provided by van Rooij. This difference will be taken up and looked at in more detail in Section 11.8.[4]

A predicate logic formula formula ϕ evaluates to true (1) or to false (0), depending on the values attributed to the variables. The values are ascribed by a so-called possible world, which corresponds to the valuation ω of Definition 9.8. It can be thought as a sequence of values, one for each variable. The set of all possible worlds is denoted by M, known from Section 9.3 as $\mathfrak{D}$, the Cartesian product of the frames of all variables. A possible world $w \in M$ assigns a truth value to a formula ϕ. This truth value of $\{0, 1\}$ is called the extension of the statement ϕ and is denoted by $[\phi]_w$. It corresponds to the truth assignment $\tilde{h}(\omega, \phi)$ from Definition 9.11. This

[4] For the sake of completeness, Groenendijk and Stokhof's semantics of statements can be linked to Section 9.3 as follows: Groenendijk and Stokhof use the same terms as in the case of propositional logic, but with a different meaning. A world w is a structure in the sense of Definition 9.6, a model M is the set of all different structures considered, and the variable assignment function (see Definition 9.7, where it is called h_Σ) is denoted by g. The truth value of a formula ϕ under g is either 0 (false) or 1 (true), according to Definition 9.11. This truth value is called the extension of a formula ϕ, with respect to a model M, a world w and an assignment g, written $[\phi]_{M,w,g}$. The intension of ϕ does not play any role and is thus not considered by Groenendijk and Stokhof. As Groenendijk and Stokhof usually take a set M of structures (called models) into account, they define the extension and the intension of statements and questions by means of a set of structures, which makes no sense in our approach.

allows to define the intension of the formula ϕ. It is the set of those possible worlds $w \in M$ in which ϕ evaluates to 1 (true). The intension of ϕ is denoted by $[\phi]$ and defined to be $\{w \in M : [\phi]_w = 1\}$. In terms of Section 11.2, $[\phi]$ is the proposition expressed by ϕ. Using our notation introduced in Section 9.3, $[\phi]$ is the set of models $\tilde{h}(\phi)$, see Definition 9.13. The table given in Figure 11.6 provides a summary of the corresponding terms:

notation	name	explanation	correspondence in Section 9.3
w	possible world	valuation	ω
M		set of all possible worlds	$\mathfrak{D}$
$[\phi]_w$	extension of ϕ	value of $\{0,1\}$ assigned by w	$\tilde{h}(\omega, \phi)$
$[\phi]$	intension of ϕ	$\{w \in M : [\phi]_w = 1\}$	$\tilde{h}(\phi)$

Figure 11.6: Linking van Rooij's predicate logic notation to the one introduced in Section 9.3

The notions of extension and intension, which were introduced for formulae so far, are now illustrated by an example.

Example 11.3.3 (Extension and Intension of a Statement) Consider a simple predicate language $\mathcal{L}$ with only two variables v_1, v_2 and the set of constants (values which can be ascribed to the variables) being $\mathbb{N}$. There is only one predicate $\leq$ specified, with two arguments, which have to be natural numbers. The relation $R_{\leq}$ underlying the predicate $\leq$ is a set of pairs (n_1, n_2) of natural numbers, such that $n_2 - n_1 \in \mathbb{N}$. The set M of all possible worlds is the set $\mathbb{N} \times \mathbb{N}$. We will now specify the extension and the intension of the formula $\phi = \leq (v_1, v_2)$.

- The extension of $\phi = \leq (v_1, v_2)$, relative to the world $w = (1, 3) \in M$, is the truth value 1: $[\leq (v_1, v_2)]_{(1,3)} = 1$.
- The extension of $\phi = \leq (v_1, v_2)$, relative to the world $w = (3, 1) \in M$, is the truth value 0: $[\leq (v_1, v_2)]_{(3,1)} = 0$.
- The intension of $\phi = \leq (v_1, v_2)$ is in this case $[\leq (v_1, v_2)] = R_{\leq}$.

$\odot$

Syntax of Questions

We have stated above that, from a semantic point of view, constituent questions in predicate logic ask for objects having certain properties, or standing in certain relations. This should be reflected in the syntactic form of the question, therefore, the following syntax for questions is proposed in (Groenendijk & Stokhof, 1997). It involves the properties and relations the question is asking for:

Definition 11.2 (Syntax of a Question)

Let ϕ be a formula in which all and only the variables $x_1, \ldots, x_n$ $(n \geq 0)$ have one or more free occurrences, then $?x_1 \cdots x_n \phi$ is a question.

Some illustrating examples are listed subsequently, at first for constituent questions, thereafter for a yes-/no-question.

One-Constituent Question A one-constituent question asks for a particular property, such as "Which students passed the test?". Being a student is formalized by the predicate P, having passed the test is described by the predicate Q. So the question, in its syntactic form, reads as $?x(P(x) \wedge Q(x))$.

Multiple-Constituent Question A multiple-constituent question asks for a particular relation, such as "Who loves whom?". Let the two-place predicate R express the relation *loves*, then this very question is $?xy(R(xy))$. Obviously, whenever $?xy(R(xy))$ is answered, an answer to the question $?yx(R(xy))$ is provided, too.

Yes-/No-Question A yes-/no-question is just a special case of the above constituent questions, since Definition 11.2 allows for zero variables to be queried over. This results in a question like $?\exists x(P(x) \wedge Q(x))$, asking whether at least one student passed the test. As in the case of propositional logic, there are only two possible answers to such a question. Note that the question mark in this syntactic form is immediately followed by the formula $\phi = \exists x(P(x) \wedge Q(x))$, and that there are no variables between them.

This syntax, introduced by Groenendijk and Stokhof, is interesting, as it extends the known predicate logic syntax with a further symbol, namely the question mark ?. One may even perceive it as a further quantifier, asking for a list of those objects which are ascribed to the variables following the question mark. However, van Rooij does not use this notation which involves variables. He proposes to precede a predicate P with a question mark, when one wants to ask for the objects attributed to the variables involved in the predicate. Syntactically, van Rooij thus presents a question by $?P$. Even if this amounts to the same, Groenendijk and Stokhof's approach is more explicit, since a list of variables is given, and more general, as it applies to formulae, not only to predicates. However, we will use van Rooij's notation below, as we also refer to his perception of the semantics of questions.

Semantics of Questions

In (van Rooij, 2009), a question is syntactically represented by $?P$, where P is an n-ary predicate. In order to provide a semantics of questions, van Rooij introduces the extension of a predicate P, which is not to be confused with the extension of a formula ϕ, introduced above. The extension of a predicate P is related to a possible world w, but concerns only the values that w provides for the variables in P. It is denoted by $[P]_w$ and corresponds to what we have called a projected valuation in Definition 9.10, denoted by $\omega^{\Rightarrow x}$. The extension $[P]_w$ of the n-ary predicate P is therefore a sequence of length n, telling which value is provided by w for every variable involved in P. One could say that $[P]_w$ is the possible world w "projected" to the variables of the predicate P.

This gives rise to the extension of the question $?P$, which depends on a possible world w, or rather on $[P]_w$, the projected possible world w. The extension of the question $?P$, denoted by $[?P]_w$ regroups all those possible worlds which provide the same values as w for the variables in P. In other words, every possible world which corresponds after projection to P to the projected world $[P]_w$, is part of the extension $[?P]_w$ of the question $?P$:

$$[?P]_w \quad = \quad \{w' \in M : [P]_{w'} = [P]_w\}. \tag{11.8}$$

Clearly, this defines a set of possible worlds. One can think of more than one extension $[?P]_w$. Each extension of the question $?P$ is a possible semantic answer to that question.

Finally, the intension of the question $?P$ is the set of all possible answers to it:

$$[?P] \quad = \quad \{[?P]_w : w \in M\}. \tag{11.9}$$

From (Groenendijk & Stokhof, 1984) it is known that the meaning of a question is provided by its possible answers. So the intension of a question, given by Equation 11.9, expresses its semantics. At the same time, the intension provides a partition of the logical space M, consisting of all possible worlds w. The blocks of the partition are the extensions $[?P]_w$, which are sets of possible worlds. All the worlds in such a set $[?P]_w$ provide the same values for the variables involved in the predicate P.

In order to clarify the notions introduced above, they are now illustrated by an example.

Example 11.3.4 (Extension and Intension of a Question) Reconsider the simple predicate language $\mathcal{L}$ of Example 11.3.3, with the binary predicate $\leq$ and its underlying relation $R_\leq \subset \mathbb{N} \times \mathbb{N}$. For a better illustration of the above notions, we will consider this time three variables v_1, v_2, v_3, taking values out of $\mathbb{N}$. Therefore, the set M of all possible worlds is the set $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$.

- The extension of the predicate $\leq$, relative to the world $w = (1,3,5) \in M$, is the projected world $[\leq (v_1, v_2)]_{(1,3,5)} = (1,3)$.
- The extension of the question $? \leq (v_1, v_2)$, relative to the world $w = (1,3,5) \in M$, is the set of worlds $[? \leq (v_1, v_2)]_{(1,3,5)} = \{w \in M : [\leq (v_1, v_2)]_w = (1,3)\}$.
- The intension of the question $? \leq (v_1, v_2)$ is a set of sets of possible worlds: $[? \leq (v_1, v_2)] = \{[? \leq (v_1, v_2)]_w : [\leq (v_1, v_2)]_w \in R_\leq\}$.

$\odot$

.

11.4 Answers

Groenendijk and Stokhof propose a theory of question-answering. This is why this section is entitled "answers" and not "(piece of) information", the term we would use in our algebraic theory of semantic information. According to (Groenendijk & Stokhof, 1984), the purpose of answers is "to fill in a gap in the information of the questioner", so answers provide information. Note however that this information is always related to one specific question, there is no general theory about information, as in the case of information algebras. Nevertheless, it is possible to filter from their papers the components that will be necessary for a later comparison of their theory with the ours (Section 11.8). Even if Groenendijk and Stokhof are not so much concerned with answers, and rarely look at them in a global, not question-specific way, it is possible to describe their opinion about answers in propositional and predicate logic (Sections 11.4.1 and 11.4.2). For the latter case, we could even find operations that can be performed on answers, see Section 11.4.3. It might be the case that an answer is provided when already some information relative to the question exists. This aspect is looked at in Section 11.4.4. We start this section about answers with a quotation from (Groenendijk & Stokhof, 1984), stating the most basic property of an answer:

> "Within the limits of possible world semantics, the information [...] can be represented as a non-empty subset of the set of indices. Each index in such an information set represents a state of affairs that is compatible with the information in question."

So Groenendijk and Stokhof consider answers in a semantic way, an answer is a set of possible worlds.[5] Information is always an answer to a question, and it is seen as a semantic object, made up of different states, describing how the world might be.

11.4.1 Answerhood in Propositional Logic

The semantics of answers or statements in propositional logic is known from Section 11.3.1 above. In the table given in Figure 11.5, the intension of a formula ϕ, which is the answer to some question, is defined by $[\phi]_M = \{w \in M : [\phi]_{M,w} = 1\}$. The intension of the formula ϕ is actually its set of models, containing all worlds in which ϕ evaluates to 1 (true). This set is also called the proposition expressed by ϕ.

But when is a formula an answer to a question q? We know that the question q is given by its possible answers, which equals, according to Equation 11.7, the set $\{[?\phi]_{M,w} : w \in M\}$. An extension $[?\phi]_{M,w}$ of the question q is given by Equation 11.6. Answerhood is now defined as follows:

[5] In (Groenendijk & Stokhof, 1984), the set of possible worlds is called set of indices, as pointed out in Section 11.2.

Definition 11.3 (Answerhood, Propositional Logic)

Let $q = \{[?\phi]_{M,w} : w \in M\}$ be a yes-/no-question in its semantic form. The formula ψ is an answer to q iff $\exists w \in M$ such that $[\psi]_M \subseteq [?\phi]_{M,w}$.

In terms of the partition q makes on M, ψ is an answer to q, iff $[\psi]_M$, the proposition expressed by ψ in M, is always a (possibly empty) part of one of the blocks of the partition.

11.4.2 Answerhood in Predicate Logic

Our presentation of questions in predicate logic in Section 11.3.2 was based on (van Rooij, 2009). However, no formal definition of answerhood in predicate logic, in the sense of Groenendijk and Stokhof's above definition of answerhood in propositional logic, is provided by van Rooij. But it is possible to reconstruct such a definition, based on (Groenendijk & Stokhof, 1997) and (van Rooij, 2009). The latter states that an answer to some question q entails one of the possible answers to q.[6] This gives rise to the following definition:

Definition 11.4 (Answerhood, Predicate Logic)

Let $q = \{[?P]_w : w \in M\}$ be a question in its semantic form, based on the n-ary predicate P. The formula ψ is an answer to q iff $\exists w \in M$ such that $[\psi] \subseteq [?P]_w$.

In terms of the partition q makes on M, ψ is an answer to q, iff $[\psi]$, the proposition expressed by ψ, or the set of models of ψ, is always a (possibly empty) part of one of the blocks of the partition $q = \{[?P]_w : w \in M\}$.

11.4.3 Operations on Answers

As already mentioned above, Groenendijk and Stokhof's main interest lies in questions. Information is not considered as a stand-alone entity, but only as an answer to a fixed question. This is also the reason why operations for information processing are rarely looked at. Nevertheless, it was possible to discover in the foundational paper of dynamic predicate logic (Groenendijk & Stokhof, 1991) two operations on predicate logic answers, which will be of interest in the comparison of Section 11.8. The purpose of dynamic predicate logic (henceforth DPL) is to rewrite the standard predicate logic semantics, using a dynamic semantics. The approach is inspired by systems of dynamic logic as they are used in the denotational semantics of programming languages. DPL is an alternative approach, having certain benefits that Groenendijk and Stokhof missed in the standard predicate logic semantics. We do not intend, however, to go into the details of DPL; we will just introduce those elements which are necessary to understand the two operations which are interesting

[6] For entailment, see Definition 9.14.

for us. They are called "dynamic conjunction" and "dynamic existential quantification", which is not really surprising after the results of Sections 9.4 and 9.5.

The idea underlying DPL is the following: The meaning of a statement lies in the way it changes the information state before its utterance to the information state afterwards. The relation to the semantics of programming languages is obvious, as an interpreting machine can be in different states. Such a machine state may be identified with an assignment of objects to variables, the variable assignment function known from the semantics of predicate logic. The interpretation of a program (a statement, a formula) is consequently regarded as a set of ordered pairs $\langle g, h \rangle$ of assignments, the first element of the pair being the input assignment g, the second one the output assignment h. The interpretation is thus the set of all the possible input-output-pairs $\langle g, h \rangle$. It can be translated to the semantics of predicate logic as follows: One considers a formula ϕ as well as assignments g, h in a fixed structure and asks for the meaning of ϕ, which is its DPL-semantic object. It is denoted by $[[\phi]]$ and consists of input-output-pairs $\langle g, h \rangle$, such that ϕ evaluates to 1 (true) under both g and h. In other words, the assignments g and h are part of the intension[7] of ϕ. Formally, the DPL-semantic object $[[\phi]]$ of a predicate logic formula ϕ is given as follows:

$$\begin{aligned} [[\phi]] &= \{\langle g, h \rangle : [\phi]_g = [\phi]_h = 1\} \\ &= \{\langle g, h \rangle : g, h \in [\phi]\}. \end{aligned}$$

For atomic formulae, it is often assumed that $g = h$, so the input-assignment g is "tested" whether really $[\phi]_g = 1$ and then passed on as output. Every other assignment is "rejected".

We have seen that in DPL, the semantic object, associated with a formula, is a set of ordered pairs $\langle g, h \rangle$ of assignments. The input g and the output h assign values to variables. We say that g is different from h if a different value is assigned to at least one variable. By $h[x]g$, one states that the output assignment h differs from the input assignment g at most with respect to the value it assigns to x. Consequently, $h[x]g$ allows two possible cases: $g = h$ or g and h differ in the value assigned to one variable. Now we dispose of the necessary toolbox to introduce the two interesting operations, which are called dynamic conjunction and dynamic existential quantification.

Dynamic Conjunction

The DPL interpretation of the conjunction of two formulae ϕ and ψ can be described as follows: The input assignment g may result in the output assignment h iff one can find some assignment k which has the following two propoerties:

1. k is the output of the dynamic interpretation of ϕ with input g, and
2. k is the input of the dynamic interpretation of ψ with output h.

[7] Recall that the intension of a formula ϕ is termed in Section 9.3 set of models of ϕ.

In terms of the better known intension of formulae, it is required for the three assignments g, h, k to satisfy $g, k \in [\phi]$ and $k, h \in [\psi]$.

Definition 11.5 (Dynamic Conjunction)

Consider assignments g, h, k and two predicate logic formulae ϕ and ψ. The semantics of their conjunction $\phi \wedge \psi$ is given by

$$[[\phi \wedge \psi]] \quad = \quad \{\langle g, h\rangle : \exists k \textit{ such that } \langle g, k\rangle \in [[\phi]] \textit{ and } \langle k, h\rangle \in [[\psi]]\}.$$

The dynamic interpretation of $\phi \wedge \psi$ consists of all pairs $\langle g, h\rangle$ which are linked by an assignment k. This relation may be depicted as in Figure 11.7. The assignments k are common to the intensions of ϕ and ψ. In other words, the assignments k can be obtained by intersecting the intensions $[\phi]$ and $[\psi]$. Obviously, the result is different in case of other formulae ϕ and ψ.

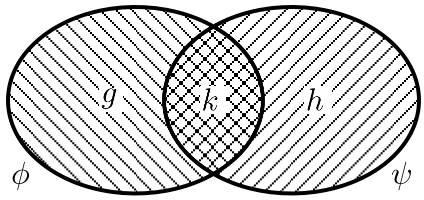


Figure 11.7: Graphical representation of dynamic conjunction

Dynamic Existential Quantification

The second interesting operation is the DPL interpretation of existential quantification. Let us first recall, for a formula $\psi = \exists x\phi$, how existential quantification is realized in standard predicate logic semantics, as described in Definition 9.11. An assignment g is in the intension (set of models) of ψ, denoted $g \in [\psi]$, iff one can find an assignment h such that ϕ evaluates to 1 under h (written $h \in [\phi]$). Additionally, h must differ from g at most with respect to the value it assigns to x (i. e. $h[x]g$). Let us now look at the semantics of the formula $\psi = \exists x\phi$ in a dynamic way. Groenendijk and Stokhof propose to take all assignments h with $h[x]g$ and $h \in [\phi]$ as possible outputs with respect to the input assignment g. Formally:

Definition 11.6 (Dynamic Existential Quantification)

Consider assignments g, h and a predicate logic formula $\psi = \exists x\phi$. Its DPL-semantic object is given by

$$[[\psi]] = [[\exists x\phi]] \quad = \quad \{\langle g, h\rangle : h[x]g \textit{ and } h \in [\phi]\}.$$

In (Groenendijk & Stokhof, 1991) a further, more refined version of the above definition is proposed, in addition. However, Definition 11.6 will be sufficient for the comparison in Section 11.8.

11.4.4 Answers Relative to an Information Set

When a question q is asked and some information, which (partially) answers q, is already available, this knowledge has to be taken into account. Each answer that is given to q has to be considered regarding this previous information. In (Groenendijk & Stokhof, 1984), the notion of "*a proposition giving an answer with respect to an information set*" is introduced. An information set might be the result of previous, consecutive answers, or simply something that always holds in the actual situation. Such an information set is given in the same way as an answer to a question, namely by a set of possible worlds. Let the information set be denoted by ψ and the answer that is given to q by ϕ. Groenendijk and Stokhof propose an updating process where the information of ψ is updated with the information provided by the answer ϕ. On the level of formulae, this is done using the conjunction $\wedge$, which is presented from a semantic point of view in Section 11.4.3 above. Depending on ϕ and ψ, the result of the updating process might cover completely one or more blocks of the partition U/q, induced by the question q (see Equation 11.1). But it is also possible that the resulting set of possible worlds is a subset of one, perhaps several blocks of U/q. See Section 11.5.1 for a classification of the different results which do not yet count as an answer in the sense of the answerhood defined above.

11.5 Order and Measure of Answers

Using Groenendijk and Stokhof's partition based analysis of questions, introduced in Section 11.2, a qualitative approach of comparing answers, the order $>_q$, is introduced in Section 11.5.1. It is taken from (Groenendijk & Stokhof, 1984). This order has been extended in (van Rooij, 2009), where a quantitative measure for the informativity of an answer is proposed, which is called inf; it is described in Section 11.5.2. The idea underlying the qualitative, as well as the quantitative approach is how good some statement ϕ answers a specific question q.

11.5.1 Order

The goodness of an answer ϕ (given by a set of possible worlds) is qualitatively measured in terms of the partition of the logical space U, induced by the question q. We are now interested in how good a statement ϕ answers a question q, represented by the partition $U/q = \{[u]_q : u \in U\}$, see Equation 11.1. For that purpose, the following notion of compatibility is introduced in (Groenendijk & Stokhof, 1984). The compatibility of an answer ϕ with a question q is determined by means of the set

$$\phi_q \quad = \quad \{[u]_q \in U/q : \phi \cap [u]_q \neq \emptyset\}. \tag{11.10}$$

The blocks of U/q which are in the set ϕ_q[8] are called compatible with ϕ, the other ones are called incompatible with ϕ. An example is given in Figure 11.8. A partition U/q with four blocks is considered. ψ is compatible with three of the four blocks, and ψ_q consists of the three lowest blocks in the figure. ψ is incompatible with only one block of the partition, the uppermost one. As for the compatibility of ϕ, it is compatible with half of the blocks of the partition, and thus incompatible with the other half of the blocks. The two lowermost blocks constitute ϕ_q.

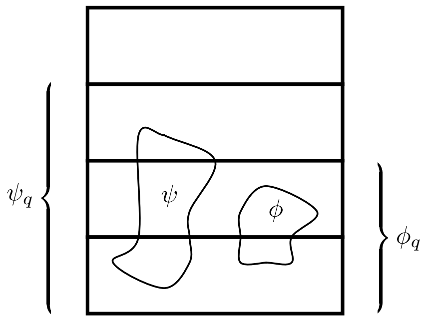


Figure 11.8: Compatibility of answers

The goodness of an answer ϕ is given by the number of (in)compatible blocks. An answer is the better the more blocks it is incompatible with. Put another way, ϕ is the better the less blocks constitute the set ϕ_q. Compatibility is used to compare two statements ϕ, ψ in how good they answer a question q. They can only be compared as described above if they do not concern different blocks of the partition, so if either $\phi_q \subset \psi_q$, where ϕ is a better answer to q than ψ, or if $\psi_q \subset \phi_q$, which allows to say that ψ is a better answer to q than ϕ.

But Groenendijk and Stokhof also propose to order answers ϕ, ψ which are compatible with the same blocks of the partition, i. e. where $\phi_q = \psi_q$. In that case, the more inclusive answer is preferred, as it said to be less overinformative. ϕ is therefore a better answer to q as ψ if $\phi_q = \psi_q$ and $\phi \supset \psi$.

Summing up what we have seen so far leads to the following order of answers:

Definition 11.7 (Order of Answers)

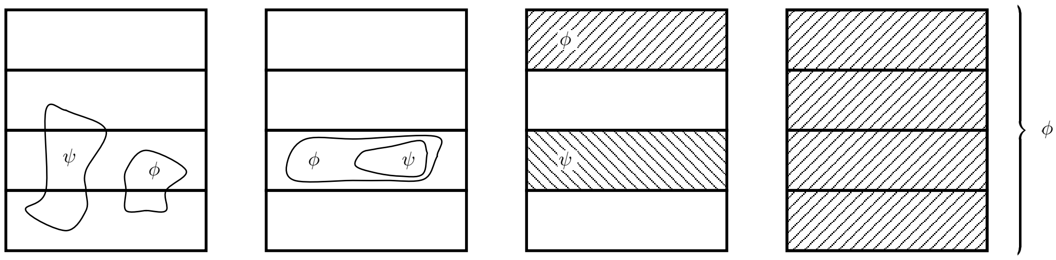
ϕ is (quantitatively) a better answer to question q than ψ, written $\phi >_q \psi$, iff either

1. *$\phi_q \subset \psi_q$, or*
2. *$\phi_q = \psi_q$, and $\phi \supset \psi$.*

If $\phi >_q \psi$, then ϕ is more informative *than ψ* regarding the question *q.*

[8]Note that ϕ_q is denoted as U/q^ϕ in (Groenendijk & Stokhof, 1984), which is rather unreadable when answers are compared to each other. This is why we have adopted the notation found in (van Rooij, 2009).

Groenendijk and Stokhof's order of answers is based on the interest to approximate one block of the partition, either from outside (first point) or from inside (second point) of the block. The two extreme cases of this order are given by the whole set U and the single blocks $[u]_q \in U/q$. If $\phi = U$, this statement ϕ is the worst (i. e. least informative) answer that can be given to the question q, as $\phi_q = U/q$. But every ϕ, that covers exactly one block $[u]_q$ of the partition U/q, counts as the best answer to the question q, in the sense that it provides the most information that can be provided regarding q. Obviously, there is only one least informative answer, but there are as many most informative answers as there are blocks in the partition U/q.



$\phi >_q \psi$
($\phi_q \subset \psi_q$)

$\phi >_q \psi$
($\phi_q = \psi_q$ and $\phi \supset \psi$)

ϕ, ψ are most informative answers

ϕ is the least informative answer

Figure 11.9: Graphical representation of the order of answers

Figure 11.9 depicts examples for the different cases that were introduced above. In all the four examples, a partition U/q with four blocks is considered. The leftmost situation illustrates the first case of Definition 11.7, where ϕ is compatible with less blocks of U/q than ψ and counts therefore as the better answer to q. In the second diagram from the left, the answers ϕ and ψ are compatible with the same block of the partition, but ψ is a subset of ϕ. According to the second case of Definition 11.7, ϕ is a better answer to q than ψ. Two answers ϕ and ψ, which are equally good, are given thereafter. They are both the most informative answers that can be given to q, as they completely cover one block $[u]_q$ of the partition U/q. Finally, the rightmost chart shows the weakest or worst answer ϕ that can be given to the question q. It is the least informative one as it states everything that can be stated; it equals the logical space U.

Note that in the order proposed above, any set of possible worlds is considered as an answer, as long as it is a subset of the universe U. However, in Definitions 11.3 and 11.4, where answers in the case of propositional and predicate logic are described, a set of possible worlds ϕ only counts as an answer to q if it is a subset of or equal to one of the blocks $[u]_q \in U/q$. Obviously, such an answer ϕ has the property $|\phi_q| = 1$. It is also referred to as complete answer in (Groenendijk & Stokhof, 1984; Groenendijk & Stokhof, 1997). The further answers, looked at above, range over more than one block, possibly as a subset. Such a ϕ with $|\phi_q| > 1$ does not give a precise answer to q, and is therefore called a partial answer.

11.5.2 Measure

Answers cannot only be ordered by means of $>_q$, but in (van Rooij, 2009), a measure for the amount of information of a proposition (i. e. a set of possible worlds) ϕ is introduced, giving rise to a total ordering. This quantitative measure inf takes probabilities into account. Given two answers ϕ, ψ, the fact that ϕ is more likely to happen than ψ is expressed by $\inf(\phi) > \inf(\psi)$. Probabilities are not primarily attributed to propositions (answers), but to the worlds making up the propositions. A world $u \in U$ is an element of the logical space.[9] There is a probability function which attributes a value out of $[0, 1]$ to each world u, such that the sum of all $u \in U$ amounts to 1:

$$p: \quad U \rightarrow [0,1] \quad \text{such that } \sum_{u \in U} p(u) = 1. \tag{11.11}$$

In order to assign a probability to an answer ϕ, the probabilities of its constituting possible worlds are summed up. Formally, this is expressed by

$$p(\phi) \quad = \quad \sum_{u \in \phi} p(u). \tag{11.12}$$

Van Rooij then bases himself on the proposal of (Bar-Hillel & Carnap, 1952), who in turn followed the approach of (Hartley, 1928), for measuring the informativity of an answer ϕ:

$$\inf(\phi) = \log \frac{1}{p(\phi)} = -\log p(\phi). \tag{11.13}$$

This measure of informativity, proposed in Equation 11.13, is Bar-Hillel and Carnap's information-measure inf, that they have taken as the second explicatum of amount of information. It is presented in Section 10.4.4, where also its properties can be found. We will not itemize these properties again. It is just interesting to mention that van Rooij puts emphasis on the fact that inf behaves monotone increasing with respect to the entailment relation, expressed by subsets of possible worlds. The total ordering induced by inf is therefore seen as an extension of the partial ordering induced by the entailment relation.

11.6 Measure of Questions

In Section 11.2.2, a partial order of questions has been introduced, by means of the inclusion relation $\sqsubseteq$ between question induced partitions, see Equation 11.2. An extension of this partial order is given in (van Rooij, 2009), where a total order between questions is proposed. A question q is again thought of as a semantic entity, a partition of the logical space, resulting in a set of sets of possible worlds. Each

[9] In the foregoing examples of propositional and predicate logic, a world was denoted by w and the set of all worlds was given by M. Here, we are considering the general case, this is why we prefer to speak of the logical space U, which was introduced in the general part about questions and answers of this chapter.

set of possible worlds is a proposition ϕ, that counts as an answer to q. As before, a question is given by the possible answers it allows.

Van Rooij applies Shannon's entropy-based measure, known from (Shannon, 1948), to the question structure. At first, the notion of entropy of a question is defined, followed by the concepts of common and conditional entropy. Their properties are described by the important chaining theorem. Thereafter, in Section 11.6.2, the topic of mutual information between a set of hypotheses and (the answers to) a question is covered. Finally, a total order of questions is proposed in Section 11.6.3.

11.6.1 Entropy of a Question

In order to describe the informative value of a question q, the expected or average amount of information, conveyed by its (complete, most informative) answers ϕ, is considered by van Rooij. It turns out that he measures the informative value of a question q by the entropy of the possible answers to q.

Informative Value of a Question

From Section 11.5.2, it is known that the informativity of an answer ϕ is measured by $\inf(\phi) = -\log p(\phi)$. The amount of information of every answer ϕ is now weighted by its probability to happen, $p(\phi) \cdot \inf(\phi)$, and summed up for all $\phi \in q$. This leads van Rooij to the definition of the informative value E of the question q:

$$\begin{aligned} E(q) &= \sum_{\phi \in q} p(\phi) \cdot \inf(\phi) \\ &= -\sum_{\phi \in q} p(\phi) \log p(\phi). \end{aligned} \tag{11.14}$$

Obviously, the informative value of a question q is the entropy of the set $U/q = \{[u]_q : u \in U\}$, which is known from Equation 11.1. From the properties of entropy, it is known that the informative value of a question is maximal if its possible answers are equiprobable. In that case, $E(q) = \log |q|$. On the other hand, the informative value of a question is minimal if the answer is already known. Such a situation is formalized by all besides one $\phi \in q$ having probability 0 to occur and, consequently, the one, remaining answer has probability 1, which leads to $E(q) = 0$. As entropy measures the uncertainty of the outcome, the informative value $E(q)$ of a question q expresses the uncertainty that exists about the answer to q. This informative value $E(q)$, proposed by van Rooij, respects and extends the partial ordering on questions, as for any two questions q and r, it holds that, whenever $q \sqsubseteq r$, it will also be the case that $E(r) \geq E(q)$.

Informative Value of a Combined Question

If two questions q and r are asked simultaneously, this may be described by a new question, characterized by other answers. Only those sets of possible worlds, that are shared by an answer ϕ to q and an answer ψ to r, count as answers to this new question. The uncertainty about the answer to the new question is given by the common entropy $E(q,r)$:

$$\begin{aligned} E(q,r) &= \sum_{\phi \in q} \sum_{\psi \in r} p_{q,r}(\phi \cap \psi) \cdot \inf(\phi \cap \psi) \\ &= -\sum_{\phi \in q} \sum_{\psi \in r} p_{q,r}(\phi, \psi) \log p_{q,r}(\phi, \psi). \end{aligned} \tag{11.15}$$

Actually, van Rooij points out that this is the informative value of the combined question $q \sqcap r$ (see Equation 11.4), so $E(q,r) = E(q \sqcap r)$. Note that

$$E(q,r) \leq E(q) + E(r), \tag{11.16}$$

which will be important below.

Conditional Informative Value of a Question

Thereafter, van Rooij proposes to see the universe U, which consists of worlds u, as a kind of artificial partition, induced by the question b, asking for what is believed. It is the most fine-grained partition

$$b = \{\zeta = \{u\} : p(u) > 0\}, \tag{11.17}$$

having the entropy $E(b)$. This entropy changes if an answer ϕ to some question q is learned. The conditional entropy of b, given a specific answer $\phi \in q$, is defined by

$$\begin{aligned} E_\phi(b) &= \sum_{\zeta \in b} p_{b|\phi}(\zeta|\phi) \cdot \inf(\zeta|\phi) \\ &= -\sum_{\zeta \in b} p_{b|\phi}(\zeta|\phi) \log p_{b|\phi}(\zeta|\phi). \end{aligned} \tag{11.18}$$

Sometimes, $E_\phi(b)$ is also written $E(b|\phi)$. It measures the uncertainty about b, when it is known that the answer to q is ϕ.

Expected Conditional Informative Value of a Question

However, it is not always the case that some specified answer ϕ to the question q is known. We may assume that an answer to q is given, but it is not known which one. So all $\phi \in q$ have to be considered, which leads to the expected conditional entropy of b, given the question q:

$$\begin{aligned} E_q(b) &= \sum_{\phi \in q} E_\phi(b) \cdot p(\phi) \\ &= \sum_{\substack{\zeta \in b \\ \phi \in q}} p_{b|\phi}(\zeta|\phi) \cdot p(\phi) \cdot \inf(\zeta|\phi) \\ &= -\sum_{\substack{\zeta \in b \\ \phi \in q}} p_{b,q}(\zeta, \phi) \log p_{b|\phi}(\zeta|\phi). \end{aligned} \tag{11.19}$$

As before, there is an alternative notation for $E_q(b)$, namely $E(b|q)$. The expected conditional entropy of the question b, given the question q, measures the uncertainty about b, when some answer to q has been learned, without knowing which one.

Chaining Theorem, Shannon's Inequality

The chaining theorem, already known from Section 7.4, is also introduced by van Rooij, but without naming it. A consequence of the chaining theorem, which he calls "Shannon's inequality", given in Equation 11.21, is of greater importance for him. First, the chaining theorem expresses that, for any two questions q and r, it holds that

$$E(q,r) = E(q) + E_q(r) = E(r) + E_r(q). \tag{11.20}$$

This holds in particular for the question b, introduced above, and any other question q. The common uncertainty of the question b (asking what is believed) and the question q (asking for some fact) is composed of the uncertainty about the answer to q, plus the remaining uncertainty about b, when some answer to q is has been learned: $E(b,q) = E(q) + E_q(b)$. Now Shannon's inequality follows follows from the chaining theorem and from Equation 11.16:

$$E_q(b) = E(b,q) - E(q) \quad \leq \quad E(b) + E(q) - E(q) = E(b). \tag{11.21}$$

The importance of this inequality is not immediately obvious, but this will change in a moment.

11.6.2 Information Gained

The above formalisms and their properties, which are all well-known from Shannon's classical information theory, enable van Rooij to define two types of informational values, with the objective to measure the average information gained from the answer to a question.

Informational Value of an Answer

The informational value of an answer $\phi \in q$, with respect to the question b, written $IV_b(\phi)$, is defined by the reduction of uncertainty about b, when the specific answer ϕ is known:

$$IV_b(\phi) \quad = \quad E(b) - E_\phi(b). \tag{11.22}$$

$IV_b(\phi)$ is sometimes also denoted by $i(B;\phi)$. It is the information gained about b by observing the answer ϕ to q. Van Rooij points out that $IV_b(\phi)$ might be negative.[10]

[10] Nevertheless, he speaks of *reduction* of uncertainty, and not of change of uncertainty, but when $IV_b(\phi) < 0$, the uncertainty is not reduced, but increased.

Informational Value of a Question

However, the informational value of the question q, with respect to the partition induced by b, denoted by $EIV_b(q)$, will always be positive. It is the average reduction of uncertainty about b, when some answer to q is learned:

$$EIV_b(q) \quad = \quad E(b) - E_q(b). \tag{11.23}$$

By Shannon's inequality, see Equation 11.21 above, $E(b)$ will always be greater or equal to $E_q(b)$. From a global point of view, the average uncertainty about b can therefore never be increased by asking a question:

$$EIV_b(q) \quad \geq \quad 0. \tag{11.24}$$

The expected informational value $EIV_b(q)$ determines the expected information to be gained from the answer to a question q. In the literature, it is known as the mutual information between b and q. An alternative writing is $I(b; q)$.

Usefulness and Relevance of Questions

A further idea that is introduced in (van Rooij, 2009), is the usefulness of asking a certain question in a specific situation. We have seen above that the partition induced by the question b, asking for what is believed, can be considered as a special case of asking a question, since in the end, one disposes of the total information about how the world is like. It might, however, be the case that the uncertainty cannot or needs not to be completely reduced. In such a case, one could for example dispose of a finite set of hypotheses $h = \{\psi_1, \ldots, \psi_n\}$. The ψ_i are sets of possible worlds, which are mutually exclusive and exhaust the logical space U. Every $\psi_i \in h$ is a subset of U and can therefore be seen as a proposition. Obviously, h is a partition of the logical space and can be treated by the same means as the questions before. The stated aim is now not any more to determine a block $\zeta \in b$, i. e. to decide which world $u \in U$ is the actual one, but one aims rather for finding out which hypothesis $\psi_i \subseteq U$ should be chosen. In order to find that out, it is natural to ask questions. Van Rooij proposes to avoid asking arbitrary questions, by looking how useful and relevant a question is, with respect to the set of hypotheses h.

The usefulness of a question q, with respect to the set h of hypotheses[11], is equated with the reduction of uncertainty about h, due to q:

$$EIV_h(q) \quad = \quad E(h) - E_q(h). \tag{11.25}$$

Van Rooij measures the usefulness of a question by the mutual information between h and q. The higher the expected information value of q with respect to h is, the more useful is q, helping to choose a $\psi_i \in h$. In (van Rooij, 2009), it is furthermore

[11] In (van Rooij, 2009), the notion of "decision problem which of the hypotheses of h should be chosen" is introduced. In order to avoid additional terms, which are not of great importance hereinafter, we do not introduce the original names, chosen by van Rooij.

pointed out that a question is relevant, with respect to h, if the mutual information between h and q is strictly higher than 0. So q is relevant to h if one can expect an answer to q to reduce the uncertainty about h:

$$EIV_h(q) \quad > \quad 0. \tag{11.26}$$

11.6.3 Total Order of Questions

The partial order defined by means of $\sqsubseteq$ in Section 11.2.2 can now be extended by making use of what has been introduced in this section. The total order given below is defined with respect to a set of hypotheses h, as described above. In particular, h can be the most fine-grained partition of Equation 11.17, induced by the question b, asking for what is believed. b describes the case where no hypotheses are considered.

Definition 11.8 (Order of Questions)

With respect to the set of hypotheses h, the question q is better than the question r, written $q >_h r$, iff either

1. $EIV_h(q) > EIV_h(r)$, *or*
2. $EIV_h(q) = EIV_h(r)$, *and* $q \sqsupset r$.

The question q is better than the question r if the expected informational value of q is higher than that of r. In other words, $q >_h r$, if the mutual information between h and q is higher than that between h and r, as in such a case, it is more useful to ask the question q. If q and r are equally useful, the question with the "weaker" answers is preferred to the question inducing the finer partition.

11.7 Comparison: Questions

Until now, we have presented Groenendijk and Stokhof's semantic theory of questions and answers. This theory has been extended by van Rooij, who proposes a quantitative measure for questions and answers. The following lines concerning the literature on question-answering can be found in (Groenendijk & Stokhof, 1984):

> *"Many proposals for the analysis of questions and answers at different levels and in different fields and frameworks exist. The aim of this paper is no other than to add another proposal to this long list. We will not discuss the work of others, or point at the relative merits of our own."*

But we shall do what Groenendijk and Stokhof did not want to do. We shall contrast their theory with our algebraic theory of semantic information, presented in the first part of this thesis. The comparison is done in two steps: In this section, we will care about the nature of questions, their order and their measure, in both theories. Section 11.8 takes answers into account and compares them to the pieces of information of an information algebra.

As already pointed out above, this section compares Groenendijk and Stokhof's view of questions with the one given by our algebraic theory of semantic information. The following points will be looked at: questions as partitions or compatible frames (Section 11.7.1), the order of questions (Section 11.7.2), the measure of questions (Section 11.7.3) and finally, questions in concrete applications (Sections 11.7.4 and 11.7.5).

The perception of a single question is the same in both approaches. Generally speaking, without considering the details of the formalism chosen for the representation of information, a question is given by the possible answers to it. These answers are considered on the semantic level. Therefore, a question is a set of semantic entities, each of which counts as an answer to this very question. The Hamblin quotation in the beginning of this thesis puts this approach in a nutshell. It turns out that both parties, independently of each other, implement this concept in the same way, namely by claiming that a question induces a partition. However, there is an important difference on the conceptual level between both approaches when more than one question is considered. The information algebra framework requires a lattice of questions, which is central to the axiomatics; the partition induced by a single question is only a consequence of the fact that all questions form a lattice. But Groenendijk and Stokhof put in their theory the emphasis on a single question, which induces a partition. It is not important that questions form a lattice, it is rather a side effect of the partition lattice. Summing up we can say that regarding questions, both approaches are similar in application, but different in concept.

11.7.1 Partitions or Compatible Frames

The most important remark on this point is that Groenendijk and Stokhof's approach to questions and the one proposed in Chapter 3, going back to (Shafer, 1976; Kohlas & Monney, 1995), are the same. They have been developed independently of each other, but they do not only share the same view on the semantic nature of questions, but also translate it both by means of partitions. Obviously, the different authors put emphasis on different aspects, but one can speak without any problems of a one-to-one correspondence.

In both approaches, the starting point is the question as the set of its possible answers. Groenendijk and Stokhof start the discussion of questions by considering only one question and by describing the different answers it allows as subsets of some global universe U, a fixed set of worlds, called the logical space. This leads the authors in a straightforward way to the introduction of partitions. They conclude that a question induces a partition of U and that the possible answers to this question are the blocks (equivalence classes) of the partition. Obviously, there are many ways of partitioning U, leading to $Part(U)$, the family of possible partitions of the universe U.

In the approach of (Shafer, 1976; Kohlas & Monney, 1995), introduced in Section 3.2, it is stated in a first step that a question, seen as the set of its possible answers, is a so-called frame. Immediately, different frames are considered and related to each other. For that purpose, the notion of partition is introduced. Thus, a frame Ω is analyzed from two points of view: as the partition of another frame Λ or as being partitioned by the other frame Λ. In the first case, Ω is a coarser frame than, or a coarsening of Λ; in the second case, Ω is a a refinement of Λ. So it is stressed that there is not only one question (frame), but a whole bunch of them, and their relation seems to be the most important point. It is important to emphasize that a global, fixed universe is not so much the focus of attention as in Groenendijk and Stokhof's approach. Only in the definition of the family $\mathcal{F}$ of compatible frames, the universe U is introduced. Point 2 of this Definition 3.8 tells that the frames in $\mathcal{F}$ correspond to partitions of the set U and that $\mathcal{F}$ can be embedded in $Part(U)$.

11.7.2 Order

In Groenendijk and Stokhof's theory on the semantics of question, as well as in our algebraic theory of semantic information, questions are related. As a question induces a partition, its relation to another question is described in both approaches by operations on partitions. The juxtaposition of the operations is done first. Thereafter, we have a look at the order relation in $Part(U)$ and its lattice properties.

Operations

In (Groenendijk & Stokhof, 1984), one one-place operation (which is actually a three-place operation) and three two-place operations on partitions are introduced. We will now draw analogies between their operations and the ours, as far as possible. For that purpose, we refer to Groenendijk and Stokhof's definitions of the operations, as presented in Section 11.2, and to the operations in $Part(U)$ of Section 3.2.1, as proposed for the algebraic theory of semantic information.

Union Inside a Partition Groenendijk and Stokhof introduce a union operation $\bigsqcup$. From its definition in Equation 11.3, it can be seen that this operation actually requires three arguments: one partition and two blocks of this partition. The result is a new partition, where the two specified blocks are unified, the others remain the same. If the original partition is U/q, and the two specified blocks are $\phi, \psi \in U/q$, then the new partition is $\bigsqcup_{\phi,\psi} U/q$.

No identical operation can be found in Section 3.2.1, but a more general one. Groenendijk and Stokhof's $\bigsqcup$ operation can be seen as the inverse of a refining (Definition 3.5) between the two frames, i. e. partitions, $\bigsqcup_{\phi,\psi} U/q$ and U/q. The refining is described by the mapping

$$\tau : \bigsqcup_{\phi,\psi} U/q \to \mathfrak{P}(U/q),$$

so U/q is a refinement of $\bigsqcup_{\phi,\psi} U/q$, which is in turn a coarsening of U/q, see Definition 3.6. As τ is a mapping from the coarser to the finer frame, but the result of the one-place union operation $\bigsqcup$ is a coarsening of the original partition, the inverse of τ has to be taken.

$$\tau^{-1} : \mathfrak{P}(U/q) \to \bigsqcup_{\phi,\psi} U/q$$

is therefore defined as

$$\tau^{-1}(\zeta) = \begin{cases} \phi \cup \psi & \text{if} \quad \zeta = \phi \text{ or } \zeta = \psi, \\ \zeta & \text{if} \quad \zeta \in U/q. \end{cases}$$

Intersection Between Two Partitions In Equation 11.4, Groenendijk and Stokhof define the intersection $\sqcap$ between two partitions. This operation consists in taking the non-empty intersections of all the blocks of the two partitions. The operation corresponding to $\sqcap$ is the join $\bigvee_{j \in J} \mathcal{P}_j$ of an arbitrary collection of partitions $\{\mathcal{P}_j\}_{j \in J}$, given by Equation 3.4. The definitions are exactly the same, if $\bigvee$ is applied to two partitions. Thus, Equation 3.4 is a generalization of Equation 11.4, and in particular

$$\mathcal{P}_1 \sqcap \mathcal{P}_2 = \bigvee_{j \in \{1,2\}} \mathcal{P}_j.$$

We will see below, when the inclusion relation $\sqsubseteq$ between partitions is discussed, that Groenendijk and Stokhof's order runs in the opposite direction of the one defined in the lattice of questions D of information algebras. This is why the meet $\sqcap$ corresponds to the join $\bigvee$ in D, and the join $\sqcup$ to the meet $\bigwedge$ in D, as we will see right now.

Union Between Two Partitions Similarly, Groenendijk and Stokhof's definition of the union $\sqcup$ is only given for two partitions, whereas Equation 3.5 defines the meet $\bigwedge_{j \in J} \mathcal{P}_j$ of an arbitrary collection of partitions $\{\mathcal{P}_j\}_{j \in J}$. Groenendijk and Stokhof's $\sqcup$ operation results in a partition containing the largest blocks of the original partitions and will act as join in the lattice of partitions. In the algebraic theory of semantic information, the meet operation in D is the join of all coarser partitions (with larger blocks). Again, $\bigwedge$ is a generalization of $\sqcup$, and in particular

$$\mathcal{P}_1 \sqcup \mathcal{P}_2 = \bigwedge_{j \in \{1,2\}} \mathcal{P}_j.$$

Inclusion Relation Between Two Partitions The last operation to be compared is given by Equation 11.2, as introduced by Groenendijk and Stokhof, and by Equation 3.3, from Section 3.2.1. In the first case, the inclusion relation is denoted by $\sqsubseteq$, and a partition U/q is included in a partition U/r, written $U/q \sqsubseteq U/r$, if every block of U/q is the subset of some block of U/r. The same relationship is identified in the second case with $\geq$. Although the relation sign has the other direction, the meaning is the same: A partition $\mathcal{P}_1$ is called a refinement of another partition $\mathcal{P}_2$ if $\mathcal{P}_1 \sqsubseteq \mathcal{P}_2$ (Groenendijk and Stokhof's approach) or if $\mathcal{P}_1 \geq \mathcal{P}_2$ (our approach).

Partial Order and Lattice

Groenendijk and Stokhof consider $Part(U)$, the set of all partitions of U, and state that $\sqsubseteq$ is a partial order on $Part(U)$. They furthermore claim that $Part(U)$ forms a complete lattice under $\sqsubseteq$, with $\sqcap$ being meet and $\sqcup$ being join. However, the authors define these operations only for two partitions, a generalization to arbitrary collections of partitions is missing. Due to the fact that Groenendijk and Stokhof's order relation $\sqsubseteq$ runs in the opposite direction of the ours, the top and the bottom element are also interchanged: $\{U\}$ is the top element of their lattice and the partition having only trivial, singleton blocks, $\{\{u\} : u \in U\}$, is the bottom element of their lattice.

The lattice of questions D of the information algebra framework can also be a sublattice of $Part(U)$, whereas Groenendijk and Stokhof only consider the whole partition lattice $Part(U)$. Our theory is not restricted to this general lattice of questions, it allows also for more specific lattices of questions. Note that due to the inverse order already mentioned above, the bottom element of D has only one block, namely U itself, and the top element of D consists only of trivial blocks.

Conclusion

In Groenendijk and Stokhof's theory on the semantics of questions and in our algebraic theory of semantic information, the operations that can be performed on questions are defined in the same way, but in our approach, their formulation is more general. Questions are not only partially ordered, they even form a lattice. The information algebra framework allows for a lattice D of questions, which is a sublattice of $Part(U)$, whereas this is not possible in Groenendijk and Stokhof's approach. As the partial orders considered in the two approaches are opposite, they have different signs, but the same meaning. As a consequence, the meet operation of one approach is the join operation of the other approach and vice versa. The same holds for the top and the bottom elements.

11.7.3 Measure

The measure of questions is not a part of Groenendijk and Stokhof's theory, it has been added later in (van Rooij, 2009). The concept of entropy is used to measure the change of uncertainty, due to getting an answer to a question. We use the same approach in Section 7.6 for measuring the information of a piece of information. Both measures are an application of (Shannon, 1948).

11.7.4 Partitions Induced by Questions

In both approaches, a question induces a partition, and the blocks of a partition provide the possible answers to this question. The set $Part(U)$ contains all possible partitions of a universe U, but not imperatively all partitions of $Part(U)$ correspond to a question. So how do the partitions, that are actually induced by a questions, look like?

Groenendijk and Stokhof

In Groenendijk and Stokhof's theory, the semantics of questions is given by their possible answers. In the end of Section 11.5.1, it turned out, that there are partial and complete answers. A complete answer to a question does not leave any uncertainty about the state of the world, everything that was asked for is answered. In the case of a partial answer, however, there is still some uncertainty left, the state of the world is not uniquely defined. Complete answers, as well as partial answers, are both possible answers to a question. However, Groenendijk and Stokhof consider only complete answers when they identify a question with those statements that count as a an answer to it. In (Groenendijk & Stokhof, 1997), this choice is motivated by the properties of a partition:

> *"An immediate consequence of the exhaustive and mutually exclusive nature of the set of possible answers is the following: in each situation [. . .] a question has a unique complete and precise true answer, viz., the unique proposition among the possible answers that is true in that situation. "*

Figure 11.1, taken from (Groenendijk & Stokhof, 1984), and the definitions of answerhood in the case of propositional and predicate logic (Definitions 11.3 and 11.4) expressly underline this approach.When Groenendijk and Stokhof consider the semantics of a question, coming from the level of syntax, they arrive at the semantic level by taking as blocks of the partition the propositions expressed by the possible (syntactic) answers of a question. So the blocks describing the semantic answers to a question contain as many possible worlds as possible, the blocks are of maximum size. Summing up, not all partitions of $Part(U)$ are induced by questions, only those partitions where every block of the partition corresponds to a complete answer to the question inducing the partition.

Algebraic Theory of Semantic Information

In our algebraic theory of semantic information, the facts are similar. The mapping σ of Definition 3.8 is actually an embedding of $\mathcal{F}$, the set of all frames representing questions, into $Part(U)$. So one may suggest the assumption that there are cases where not all partitions of $Part(U)$ are actually induced by a question. This is why the information algebra framework allows for a lattice D of questions, which is a sublattice of $Part(U)$. The most general approach is the one presented in Section 3.2.1. A question is given by a frame, the set of possible answers to that question. There may be some semantically impossible partitions of the universe, as in Example 3.2.1 ("car that will not start"), where the refining mapping allows only some specific blocks in the finer frame; other blocks can simply not occur, they would contradict the reality, the common knowledge or any other given constraints.

Besides semantic constraints, which influence the partitioning of U, there are also formal constraints, due to computation. Most of the current information modeling formalisms, used for computer-operated data processing, make use of variables. Such computational applications may take advantage of the information algebra properties underlying the algebraic theory of semantic information. Then, one can provide detailed information about which partitions are induced by questions and which are not. Partitions relative to finite sets of variables, which have already been introduced in Section 3.2.2, play an important role for computational purposes. An information algebra is a generic structure for inference, which meets many formalisms. Formalisms which fulfill the axioms of an information algebra allow to perform the reasoning process using local computation, see Section 4.5. Local computation is based on the labeled version of information algebras (Chapter 4), where each piece of information comes along with a domain. This domain indicates the set of variables the piece of information is about. A factorization of the knowledge to be reasoned on is assumed, each factor being a piece of information. Inference is then done by organizing the computations in a series of combinations and variable eliminations, in a way that each of these operations takes place in a domain of one of the factors. This is most easily done using a lattice of finite sets of variables, which is distributive and therefore modular. Using partition lattices with those additional properties makes local computation easy, and fits out the algebraic theory of semantic information with a custom-made and efficient data processing mechanism. The algorithm is based on computations on so-called join trees with certain properties which rely on the distributive lattice of partitions, induced by subsets of variables. Note nevertheless that (Mellouli, 1988) proposed a way of performing local computation with a general partition lattice, based on changes on the join tree. It is an implementation of the more general approach from (Shafer, 1976; Kohlas & Monney, 1995).

So in most of the practical applications using instances of information algebras, like propositional and predicate logic presented in Chapters 8 and 9, only a very specific subset of $Part(U)$ is considered to be induced by questions. Given a countable set of variables Vbl the information algebra instance makes use of, the elements of a lattice D of finite subsets of Vbl are considered to induce questions. D is a distributive and therefore modular lattice, it is sometimes also denoted by $\mathfrak{P}(Vbl)$ in order to stress that it is composed of subsets of the set of all variables Vbl. See Section 3.2.2 for more details on $D \subset Part(U)$.

Conclusion

In both approaches, not every partition in $Part(U)$ is induced by a question. Groenendijk and Stokhof only consider partitions whose blocks correspond to complete answers of the inducing question. In the algebraic theory of semantic information, most of the time a distributive sublattice $D \subset Part(U)$ of finite subsets of some set of variables Vbl is used for computational purposes, as, given such a lattice, the mechanisms using join trees are then more easily applied.

11.7.5 The Propositional and Predicate Logic Case

We now come to the examples discussed by Groenendijk, Stokhof and van Rooij. They use the same formalisms as we do in the second part of this thesis, namely propositional and predicate logic. In Section 11.3, questions of both logics are looked at from the perspective of Groenendijk, Stokhof and van Rooij. As already pointed out in these sections, Groenendijk and Stokhof consider in the case of propositional logic several sets M of possible worlds, whereas we restrict ourselves to one set M of possible worlds. The same holds for predicate logic, where Groenendijk and Stokhof deal with several structures, but we fix one structure $\Sigma = \{F, R\}$. Due to the computational background of our approach, which is used to solve one precise problem and not all problems that might occur, we only work with a fixed structure. Our approach corresponds to the one of (van Rooij, 2009).

Propositional Logic

As already seen before, a question induces a partition. This also holds for the propositional logic example. However, the universe U, which is given by 0/1-vectors, called possible worlds or valuations, is partitioned differently.

Groenendijk and Stokhof arbitrarily choose one possible answer to the question. This answer is in its syntactic form, so it is a formula. The sets of models of this formula are taken and one block of the partition of U is found. The other one is obtained by the set complement in U. So in Groenendijk and Stokhof's point of view, a question induces a bipartition of the universe U.

The situation is different in the algebraic theory of semantic information, where questions are induced by finite sets of variables. Consider a set x of variables, inducing a question. Then, one can look for the possible answers to this question. For that purpose, a valuation $\omega \in U$ is chosen and all valuations θ which are equivalent to ω make up the equivalence class $[\omega]_{\pi_x} = \{\theta \in U : (\theta, \omega) \in \pi_x\}$. Being equivalent to ω means having the same values as ω for the variables in x. As there are only two possible values a variable can take, the set x of variables induces $2^{|x|}$ different equivalence classes. The set x of variables can be seen as the syntactic form of a question, whereas the partition the question induces is its semantic form.

Predicate Logic

In the case of predicate logic, however, van Rooij proposes the same kind of partitions as we do. Bipartitions are still possible, but now n-fold partitions, determined by the underlying predicate, come into play. The universe is composed of sets of possible worlds or valuations, both are entities where one precise value is ascribed to each variable. U is the set to be partitioned, it is denoted in Section 9.3 by $\mathfrak{D}$.

In our approach, a question is given, as before, by a set of variables. Van Rooij works with a predicate, involving implicitly a set of variables.[12] The way of determining the partition is the same in van Rooij's and in our approach: equivalence classes of elements of U, sharing the same values for the specified set of variables, are taken. As variables can take more than only two values, this gives rise to n-fold partitions.

Conclusion

Summing up, one can say that questions in propositional and predicate logic are given in the same way in both theories: they induce partitions. However, Groenendijk and Stokhof restrict themselves to bipartitions in the case of propositional logic, whereas the algebraic theory of semantic information allows for more general partitions. In the case of predicate logic, general partitions are possible in van Rooij's, as well as in our approach. Furthermore, note that Groenendijk and Stokhof do not have such an underlying framework as information algebras, providing a lattice D of questions, as well as a set Ψ of pieces of information, and capturing the properties of information and questions. It is probably for that reason that a syntactic form of questions is lacking in the case of propositional logic. These circumstances make an abstract, general discussion very difficult. However, we have to admit that Groenendijk and Stokhof's approach to the semantics of question is more general in the case of propositional and predicate logic, as they do not restrict themselves to a fixed set of possible worlds or to a fixed structure, as we do.

11.8 Comparison: Answers

After the comparison of questions in the previous section, we will now compare Groenendijk and Stokhof's answers to our pieces of information. Even if two different terms are used, it turns out that there are lots of similarities. First, Groenendijk and Stokhof's answers will be described using the information algebra vocabulary (Sections 11.8.1 and 11.8.2). Then, the operations which are performed on pieces of information and answers are juxtaposed in Section 11.8.3. Thereafter, in Section 11.8.4, the orders proposed in both approaches are looked at and we close with some words about the measure of answers in Section 11.8.5.

11.8.1 Answers vs. Pieces of Information

Groenendijk and Stokhof's theory is about the semantics of questions. Therefore, there is no stand-alone entity "piece of information", only answers to questions are considered. In our algebraic theory of semantic information, there is a set of questions, a set of pieces of information, and there exist links between the two sets. However, each set has a right to exist of its own, without depending on the

[12] Groenendijk and Stokhof also provide a syntax for questions, involving the set of variables one is interested in.

other one. It is important to state that both theories have a different intent and, therefore, questions and answers or pieces of information occur in a different context. Nevertheless, answers and pieces of information are considered in both theories in a semantic way.

Answers

Groenendijk and Stokhof have developed a theory of question-answering. Thus, the purpose of statements is being answers to questions and in this function, providing information. Answers are only given relative to one specific question, but never all possible answers to all possible questions in a given situation are considered. There is neither a set of all answers in a global, i.e. not question-specific situation, nor a general theory of statements being answers, their properties and the operations applied to them. Groenendijk and Stokhof are interested in the semantics of questions. Questions are given by their possible answers, and the semantics of a question is captured by the semantic answers to it. In Groenendijk and Stokhof's approach, answers will always be considered on the semantic level in the end. Answers are therefore a semantic entity, their meaning is of interest, not their syntactic form, which is extremely rare for authors with such a strong background in logic.

Answers are subsets of the universe U, called logical space in (Groenendijk & Stokhof, 1997):

> *"The logical space defined by a question is the space of possibilities it leaves for the world to be like. [...] The possible answers to a question form a partition of the logical space."*

Each block of a partition (which has been induced by a question) represents a possible semantic answer to that question. In particular, an answer is a subset of U. The elements, constituting a block of a partition, are complete specifications of those objects the question asks for.

Pieces of Information

The algebraic theory of semantic information is based on the concept of information algebras. This is a general framework describing the common features of many different formalisms for representing information, not only capturing different logics. All these formalisms can be traced back to a two-sorted algebraic structure, consisting of questions and pieces of information, which can be processed by means of two operations fulfilling certain axioms. The starting situation is thus very different from the one met in Groenendijk and Stokhof's theory.

The set of questions of an information algebra has already been discussed in detail before (Chapter 3 and the comparison in Section 11.7), it should be now clear that an information algebra provides a lattice of questions, denoted by D. The set of pieces of information, denoted by Ψ, can be described from a semantic point of view in a very similar manner as for Groenendijk and Stokhof above, namely by referring to the universe U. Pieces of information are subsets in a subfamily Ψ of $\mathfrak{P}(U)$. So the semantics of a piece of information is the same as that of an answer in Groenendijk and Stokhof's approach: it is a subset of the universe U.

Answers as Atomic Pieces of Information

Groenendijk and Stokhof identify the possible answers to a question with the blocks of a partition of the universe U. Answers which equal one block of a partition are in information algebra terms atoms relative the question q, inducing the partition: In Section 6.4 about atomic information, it is stated that no piece of information relative to the question q can be more informative than an atom, except the contradiction. From Definition 11.7 it is known that answers, which cover one whole block of a partition induced by q, provide the most information to q. Therefore, a possible semantic answer to a question q, which is one block of the partition induced by q, can be equated with an atom in the algebraic theory of semantic information.

Conclusion

Even if the two theories have a different background and a different intent, they consider statements (which are called answers or pieces of information in the respective theories) in the same, semantic way, namely as being subsets of some universe U. The answers of Groenendijk and Stokhof's approach, which stem from a partition of the universe U, are the atomic pieces of information of an information algebra.

11.8.2 Complete and Partial Answers

An information algebra (Ψ, D) provides pieces of information $\psi \in \Psi$, which relate to questions in D. The pieces of information are classified by means of a partial order $\leq$ or using of a pre-order $\leq_x$, when one is only interested in their order relative to a question $x \in D$. Groenendijk and Stokhof distinguish between complete and partial answers to some question q. Complete answers are compatible with one block of the partition induced by q, see Equation 11.10. Partial answers are compatible with more than one block of this partition. We will now describe both types of Groenendijk and Stokhof's answers using the information algebra terminology:

Complete Answers

In the class of complete answers to a question q, Groenendijk and Stokhof make a distinction between

1. answers which fill one block of the partition induced by q, and
2. answers which are a subset of such a block.

Answers which equal one block of the partition are known from Section 11.8.1 above. In terms of the universe U and an information algebra (Ψ, D), with $\Psi \subseteq \mathfrak{P}(U)$ and D a lattice of questions, such an answer can be described as follows: q is a question of D and induces a partition $\mathcal{P}_q$ of U. The blocks B of $\mathcal{P}_q$ are elements of $\mathfrak{P}(U)$. As D is isomorphic to the partition lattice induced by its elements $q \in D$, and as every $\psi \in \Psi$ is related to some element q of D, all blocks B of $\mathcal{P}_q$ are in Ψ. Therefore, the first type of complete answers corresponds to a piece of information

$$\psi \in \Psi \quad \text{with} \quad \psi = B.$$

ψ is a block of the partition of the universe induced by the question q.

Complete answers of the second type provide more information than asked for by the question. Such answers have been called overinformative in Section 11.5.1 and in (Groenendijk & Stokhof, 1997) their nature is described as follows:

> *"This means that such [overinformative] answers contain additional information that the interrogative does not ask for. This lack of precision is not as harmless as it may seem."*

Such an overinformative answer to q is not constituted of all elements of a block $B \in \mathcal{P}_q$, but only of a subset of them. In terms of an information algebras (Ψ, D), where a question $q \in D$ has given rise to a partition $\mathcal{P}_q$ of U, an overinformative answer to q corresponds to an element $\psi \in \Psi$ with the following property:

$$\text{There exists a block } B \in \mathcal{P}_q \text{ such that } B = \psi^{\Rightarrow q} < \psi.$$

According to the partial order of the pieces of information in Ψ, proposed in Sections 6.2 and 7.2.1, the piece of information ψ is really more informative than the whole block. ψ is strictly more informative ($>$) than $\psi^{\Rightarrow q}$, and as we are dealing with subsets of the universe U, this can be translated to $\psi \subseteq \psi^{\Rightarrow q} = B$.

Partial Answers

Every subset of U, which is not a complete answer to a question q, is a partial answer to it. In particular, a partial answer to q is constituted of elements of U which belong to more than one block B of $\mathcal{P}_q$, the partition induced by question q.

11.8.3 Operations

In the algebraic theory of semantic information, there are two basic operations associated with an information algebra (Ψ, D). The operations of combination and focusing are very central in our approach, which has a strong background in computer science. Information processing can be seen from a computer scientist's point of view as a sequence of information aggregation (i. e. combination) and information extraction (i. e. focusing) and is for example used by computer systems for query answering. Information processing is not the focus of interest of Groenendijk and Stokhof's theory, as they come from a linguistic and philosophical background. In Section 11.4.3, two operations from dynamic predicate logic (Groenendijk & Stokhof, 1991) are introduced. Dynamic conjunction corresponds to the combination operation and dynamic existential quantification to variable elimination and thus to the focusing operation of the information algebra instance predicate logic, described in Chapter 9.

In order to compare the operations, predicate logic has to be looked at on a semantic level. With the aid of the table given in Figure 9.2, recall that predicate logic is considered relative to a set of variables $Vbl(\mathcal{L})$, which gives rise to a lattice of questions D. The elements of D are subsets of $Vbl(\mathcal{L})$. As to the set of pieces of information Ψ, it is a subfamily of $\mathfrak{P}(\mathfrak{D})$, where $\mathfrak{D}$ is the Cartesian product of the frames of all variables in $Vbl(\mathcal{L})$, also referred to as the set of all valuations.[13] The operation of combination is implemented by set intersection, the operation of variable elimination corresponds to cylindrification as given in Definition 9.22.[14]

A valuation provides a value for each variable in $Vbl(\mathcal{L})$, it expresses a variable assignment, as provided in dynamic predicate logic (DPL). A central role is assigned to assignments in DPL. In order to link DPL to the information algebra instance predicate logic, we will now equate assignments with valuations and use only the latter term. Instead of working with sets of valuations to represent information, DPL makes use of sets of pairs of valuations which are considered to be input-output-pairs. We will now show how to discover the combination and variable elimination operations in the definitions of dynamic conjunction and dynamic existential quantification.

Combination / Dynamic Conjunction

In Definition 11.5, the dynamic conjunction of two pieces of information ϕ and ψ is realized by means of a set of valuations k. This set results from the intersection of ϕ and ψ, see Figure 11.7. The correspondence to combination as set intersection is obvious.

Variable Elimination / Dynamic Existential Quantification

Linking the operation of dynamic existential quantification (Definition 11.6) with the cylindrification operation from Definition 9.21 is not so straightforward as in the case of combination, but nevertheless evident. Applying the cylindrification operation on a piece of information $\psi \in \mathfrak{P}(\mathfrak{D})$ results in a set $\exists v\psi$ of valuations, which differ from the valuations in ψ at most in the value assigned to the specified variable

[13] Reusing the above term of the universe, $U = \mathfrak{D}$.

[14] In this definition, cylindrification is applied to a set of variables. It is a generalization of Definition 9.21, where only one variable is considered. This one-variable version will be used in the following comparison, as Groenendijk and Stokhof's operation takes only one variable into account.

v. The input-output-pairs, which are created by performing dynamic existential quantification, can be obtained by taking the pairs of $\psi \times \exists v \psi$.

Conclusion

The combination and the variable elimination operation could be discovered in the DPL-versions of conjunction and existential quantification. Groenendijk and Stokhof work with paris of valuations and finally obtain a set of pairs of valuations, whereas we are dealing with single valuations and end up with a set of valuation. However, the idea underlying the operations is the same.

11.8.4 Order

As Groenendijk and Stokhof's interest is question-answering, they do not have developed a general, overall order of answers, comparing all possible answers which can be given to all possible questions which can be asked. They only take into account the answers to one specific question q and look at how good different answers are in providing information relative to q. The partial order between the pieces of information in Ψ, proposed in Sections 6.2 and 7.2.1, must therefore not be compared to the order relation $>_q$ of Definition 11.7. It is the qualitative measure regarding a question q, the pre-order $\leq_q$ from Definition 7.4, which has to be considered. We obviously encounter a problem since in both approaches, the same notation is used. From now on, we will denote Groenendijk and Stokhof's order $>_q$ of Definition 11.7 by $\gg_q$, or $\ll_q$, respectively, but the pre-order from Definition 7.4 will continue running as $\leq_q$.

Applying this pre-order $\leq_q$ to pieces of information, which are subsets of the universe U, allows to rewrite it as

$$\phi \leq_q \psi \text{ iff } \phi^{\Rightarrow q} \otimes \psi^{\Rightarrow q} = \psi^{\Rightarrow q} \text{ iff } \phi^{\Rightarrow q} \supseteq \psi^{\Rightarrow q}.$$

This shows that, using $\leq_q$, one can only compare two pieces of information if one is the subset of the other, where both pieces of information have to be considered with respect to the question q. $\phi \leq_q \psi$ means that ϕ is less informative than ψ relative to the question q. Looking at a piece of information $\psi \in \Psi$ relative to the question q amounts to focusing ψ on q. This induces a partition $\mathcal{P}_q$ of U. $\psi^{\Rightarrow q}$ contains those blocks $B \in \mathcal{P}_x$ whose intersection with ψ is not empty. The same fact is expressed by Equation 11.10, defining the compatibility of an answer with a question q.

Groenendijk and Stokhof's order $\gg_q$ is more general, as it compares any two subsets of U. We will also rewrite it in order to have the same direction for both orders. Recall that ψ_q denotes the set of blocks of $\mathcal{P}_x$ in which ψ is involved (compatibility, see again Equation 11.10). $\phi \ll_q \psi$ means that ϕ is less informative than ψ relative to the question q:

$$\phi \ll_q \psi \quad \text{iff either} \quad \begin{array}{l} \phi_q \supset \psi_q, \text{ or} \\ \phi_q = \psi_q, \text{ and } \phi \subset \psi. \end{array}$$

The first case considered in this definition of $\ll_q$ corresponds, besides equality, to the definition of $\leq_q$, as $\psi_q = \psi^{\Rightarrow q}$. Therefore, the two orders can be expressed as follows:

$$\begin{aligned} \phi \leq_q \psi \quad &\text{iff} \quad \phi^{\Rightarrow q} \supseteq \psi^{\Rightarrow q}. \\ \phi \ll_q \psi \quad &\text{iff} \quad \phi^{\Rightarrow q} \supset \psi^{\Rightarrow q}. \end{aligned}$$

The second case of the definition of $\ll_q$ is treated differently by the pre-order $\leq_q$, which allows equality. If it holds for two pieces of information ϕ, ψ, with $\phi \neq \psi$, that $\phi^{\Rightarrow q} = \psi^{\Rightarrow q}$, then they provide the same information relative to the question q. The original pieces of information ϕ and ψ are not any more taken into account. Groenendijk and Stokhof, however, still look at them and derive at that point an order different from the ours.

Summing up, we have seen that comparing two pieces of information ϕ, ψ relative to the question q depends on the focused sets $\phi^{\Rightarrow q}$ and $\psi^{\Rightarrow q}$. The pre-order $\leq_q$ corresponds to Groenendijk and Stokhof's order $\ll_q$ if $\phi^{\Rightarrow q} \supset \psi^{\Rightarrow q}$. In the case of equality of the two focused sets the two orders disagree.

11.8.5 Measure

The measure that van Rooij proposes for Groenendijk and Stokhof's semantic answers to questions is taken from (Bar-Hillel & Carnap, 1952), so see Section 10.6.2 for the comparison.

11.9 Conclusion

In the previous two sections we have seen that there is a lot of common ground in Groenendijk and Stokhof's theory of the semantics of questions and our algebraic theory of semantic information. The main differences are due to the background of the theories; Groenendijk and Stokhof come from logic, philosophy (of language) and linguistics, our approach has its origins in mathematics and computer science.

Differences

The main difference between Groenendijk and Stokhof's and our approach is the intent: Groenendijk and Stokhof developed a theory of question-answering, giving a semantic description of the act of asking a question and afterwards, obtaining an answer to that question. As it is a theory about questions, the relations between questions are indeed described, but after all, every question is treated separately. The answers to different questions are not related at all.

Our field of interest is wider, it covers not only questions, but also pieces of information, which have not be called answers, as they are an independent entity. By describing the nature of information, which may best be seen in a semantic way, questions have a very important role. But our goal is to look at information in general, to state its algebraic and semantic properties, as well as the relations between pieces of information, its order and measure.

There are some slight conceptual differences between the two theories, but in the end, the same result is obtained in most of the cases. Finally, the theory proposed by Groenendijk and Stokhof only applies to logic, whereas the framework of information algebras, which defines the algebraic theory of semantic information, applies to lots of other formalisms than logic such as systems of linear (in)equations and data bases, just to cite some of them.

Commonalities

However, there are a lot of similarities between the two theories. First of all, and that sets both theories apart from most of the other theories of information, a semantic approach is chosen to describe information – not by a formula, but by sets of valuations. In both approaches, the entailment relation reflects the characteristics of being more informative, either for answers to a specific question or for pieces of information in general. The operations of focusing and combination, which are the basic operations of an information algebra, are not important within Groenendijk and Stokhof's approach, but their definition can also be found and is similar to ours. The second big common is the strong link between information and questions. Both of us point out that a information is always related to a question. A question induces a partition of the available knowledge (the universe) by regrouping those elements of the universe which are equivalent relative to that question. Thus, in both approaches, a question is seen as the set of possible answers it allows.

12

Barwise and Seligman's Information Flow

[...] computer science is rife with phenomena whose understanding requires close attention to the interaction between language and structure.

Scott Weinstein
Finite Model Theory and Its Applications

In 1997, Jon Barwise and Jerry Seligman published a book about yet another semantic theory of information. It is a theory on information flow and so is called their book: "Information Flow. The Logic of Distributed Systems." This theory is based on three fundamental notions: classification and infomorphism which constitute an information channel. Classifications are also referred to as Chu spaces (Pratt, 1999) or as contexts. The latter are known from formal concept analysis, see (Davey & Priestley, 2002) for a good introduction to this topic. Based on (Barwise & Seligman, 1997), the theory of information flow is presented in this chapter, and it is shown how this theory can be linked to information algebras. The first two sections of this chapter introduce the fundamental notions classification (Section 12.1) and infomorphism (Section 12.2), followed by the concept of a channel, which is a core concept of the theory, in Section 12.3. This concludes the first part of this chapter, where an overview of the theory of information flow is given, as presented in (Barwise & Seligman, 1997). The next two sections establish the link between the theory of information flow and information algebras: Section 12.4 provides a slightly modified formulation of classifications, namely contexts. This paves the way for Section 12.5, where the theory of information flow is presented from an information algebraic point of view, and it is shown that contexts form an information algebra. As we will see, contexts can act as a tool for generating new information algebra instances. The content of Sections 12.4 and 12.5 is inspired by (Mengin & Wilson, 2001), (Davey & Priestley, 2002), (Kohlas, 2002), (Kohlas & Stärk, 2007) and (Kohlas & Schneuwly, 2009).

Before going into details, we will give an informal overview of the theory of information flow and point out its main characteristics. This is done by citing the four principles of information flow, which are given in (Barwise & Seligman, 1997):

1. *Information flow results from regularities in a distributed system.*
2. *Information flow crucially involves both types and their particulars.*
3. *It is by virtue of regularities among connections that information about some components of a distributed system carries information about other components.*
4. *The regularities of a given distributed system are relative to its analysis in terms of information channels.*

The **first principle** tells a lot about Barwise and Seligman's perception of information and its flow. It points out that the theory presupposes a distributed system and that this system has some regularities triggering a division of the system into parts. Information flows between the different parts. As an example of such a system, Barwise and Seligman mention a flashlight which is conceptually divided into the following parts: bulb, switch, batteries, case. (The light bulb carries the information that the switch is on and the battery is charged.) However, such a decomposition is clearly not unique. There might be other parts which have not be represented, or even other ways (e. g. more detailed ones) of dividing the system. A different decomposition of the system will give rise to another kind of information flow from one part to another.

Types and particulars come up in the **second principle**. Types describe by a set of properties a certain set of particulars. Particulars have always a set of associated types and by it, carry information about their characteristics. The following example, given by Barwise and Seligman, illustrates the relation between types and particulars. A flashlight bulb (particular) can be described by some of these properties (types): LIT, UNLIT, LIVE. In their book, Barwise and Seligman refer to particulars or instances, that carry information, also as tokens. This terminology is also known from the field of philosophy of language, but should not be associated with the ideas arising from that context. The authors literally say:

> *"By a token, we mean only something that is classified; by type, we mean only something that is used to classify."*

This leads directly to classifications, the subject of Section 12.1. A classification is a formalism which ascribes properties (types) to objects (tokens), or from another point of view, which describes the characteristics (types) of some set of objects (tokens) and thereby carries or represents information. Recall the first principle, where a distributed system is mentioned. The system is decomposed into different parts, every part expressing information by a classification. So information flow takes place between classifications.

Now we can move on to the **third principle**, which tells more about a translation of information from one classification to another, which Barwise and Seligman call infomorphism. An infomorphism allows to identify information in one classification by the tokens or types of another classification. Loosely speaking, an infomorphism is a process which acts between two classifications and expresses the information of one classification by means of the other one and vice versa; it is a possibility to transport information from one component of the system to another one. However, information can only be transported when there is a connection between the tokens of the two classifications. These connections come from the regularities which reign the system. In (Barwise & Seligman, 1997), a set of classifications connected by infomorphisms is referred to as an (information) channel. The regularities, mentioned in the above enumeration in the third position, describe the set of all constraints which are in the system, modeled by a classification, as expressed by the authors:

> *"If the tokens of the classification represents all the possible instances of the device deemed relevant to the problem at hand, then the classification gives rise to a "theory" of the device, namely, those relationships between types that hold for all the tokens."*

A theory is a set of relationships, which are also called constraints. The example of the flashlight bulb (token) and its properties (types) LIT, UNLIT, LIVE gives rise to constraints like $\{\textsc{lit}\} \vdash \{\textsc{live}\}$ ("Every lit bulb is live.", called entailment constraint) or $\{\textsc{lit}, \textsc{unlit}\} \vdash \{\}$ ("No bulb is lit and unlit.", called incompatible types constraint).

Finally, the **fourth principle** might also be called the design principle. When modeling some situation by a distributed system, one has to decide what tokens and what types are present in the classification. By doing this, one takes also the decision on the constraints of the classification, i. e. on which regularities there are in the system and of what kind they are. In the same way, one decides on the flow of information in the system, as the regularities determine the connections between the classifications. Changing the regularities makes often changing the classifications and the infomorphisms between them, which results in a different information flow.

12.1 Classifications

In this section, classifications will be formally introduced. For Barwise and Seligman, a classification is a formalism which ascribes attributes to objects, or, from another point of view, which describes the characteristics of some set of objects. We will start with an introductory example, so that one gets a feeling for classifications. After that, a formal definition and some further notations are given.

Classifications, as introduced by (Barwise & Seligman, 1997), consist of types and tokens. A type is something that classifies a token. It may therefore be seen as a characteristic trait of some object. A token is classified by a type, so we may say that it is an object which has one or several properties. Classifications are often represented by tables, so-called classification tables, where the types are naming the columns and the tokens are naming the rows. The following example is taken from (Davey & Priestley, 2002) and modified according to the latest news in astrology.

Example 12.1.1 (Classification) In the table given in Figure 12.1, very rough information about the planets of our solar system is given. The rows provide a set of planets, the columns a set of attributes. The seven indicated properties are relating to size, distance from the sun and existence of a moon. This is a classification: a set of tokens (planets) and a set of types (attributes).

	size			distance from sun		moon	
	small	medium	large	near	far	yes	no
Mercury	×			×			×
Venus	×			×			×
Earth	×			×		×	
Mars	×			×		×	
Jupiter			×		×	×	
Saturn			×		×	×	
Uranus		×			×	×	
Neptune		×			×	×	

Figure 12.1: Classification table for planets

One may now ask which planet has which attributes. To answer this, a token is chosen, say the planet *Earth*, and the corresponding attributes are looked up in the table. They are marked by a ×: *size-small, distance-near, moon-yes*. If one wants to know which planets are described by a specific set of attributes, we have to look for a set of planets possessing the given set of attributes, i. e. having a × at the very properties. In the case of {*size-large, distance-far, moon-yes*}, there are two corresponding planets: Jupiter and Saturn. ⊙

Let us now define classifications, as introduced in (Barwise & Seligman, 1997):

Definition 12.1 (Classification)

A classification *is a triple* $C = (tok(C), typ(C), \models_C)$, *where*

- $tok(C)$ *is a set of objects* t *to be classified, called tokens,*
- $typ(C)$ *is a set of objects* θ, *used to classify the tokens, called types, and*
- $\models_C$ *is a binary relation between the set of tokens and the set of types, called classification relation:*

$$\models_C \subseteq tok(C) \times typ(C).$$

For $t \in tok(C)$ *and* $\theta \in typ(C)$, *we shall write* $t \models_C \theta$ *if* $(t, \theta) \in \models_C$, *i. e. the token* t *is classified as being of type* θ.

Barwise and Seligman depict the dual structure of a classification, composed of types and tokens, with the binary relation connecting them, by means of a diagram as the one given in Figure 12.2.

$$\begin{tikzcd} typ(C) \arrow[d, dash, "\models_C"] \\ tok(C) \end{tikzcd}$$

Figure 12.2: Illustration of a classification

Now that we dispose of a formal definition of classifications, we give two further examples for a better understanding of the formalism.

Example 12.1.2

1. *Powersets:* Let A be any set and and $\mathfrak{P}(A)$ its power set. Here, an element of A is linked to a set of $\mathfrak{P}(A)$, if it is contained in the latter. A token is therefore an element of A, a type is a subset of A, i. e. an element of $\mathfrak{P}(A)$. The binary relation of such a classification is $\in$ and so we can denote the powerset classification by $(A, \mathfrak{P}(A), \in)$.

2. *Propositional Logic:* Let P be a finite set of propositional symbols $p_1, p_2, \ldots, p_n$. $typ(C_{PL})$ is $Fml(\mathcal{L}_P)$, the set of all well-formed formulae. For $tok(C_{PL})$, we take the set $\{0,1\}^{|P|}$ of valuations. (See Sections 8.1 and 8.2 for the notation of propositional logic.) If $v \in tok(C_{PL})$ is a valuation for the symbols of P and $\gamma \in typ(C_{PL})$ a wff, then $v \models_{PL} \gamma$ means that the valuation v satisfies the formula γ. In other words, v is a model of γ. So the binary relation of the propositional logic classification C_{PL} is $\models$ and the classification is denoted by $(\{0,1\}^{|P|}, Fml(\mathcal{L}_P), \models)$.

$\odot$

Other examples of classifications are predicate logic (the types are formulae, the tokens are structures, the binary relation is the well-known entailment relation) or linear equations. In a system of linear equations, one disposes of a set of equations (types) and a set of solution-vectors (tokens). A set of equations describes (or not) some solution, and one can find for every solution-vector a set of equations for which this is a solution. Barwise and Seligman provide examples from less formal situations, like classifying the rolls of a pair of dice or the flashlight, as seen above. Even if these examples are quite different, there is always the same dual structure with a binary relation behind them.

The decision which part of the classification takes the place of the types and which one of the tokens is not always obvious. Some people may have argued the other way round in the preceding examples. This is possible and is referred to as type-token-duality in (Barwise & Seligman, 1997) and allows tokens to classify their types. When we want the types of some classification C to be classified by the tokens, the resulting classification is denoted by $C^{\perp}$ and called the flip of C. The types of C are the tokens of $C^{\perp}$ and the tokens of C are the types of $C^{\perp}$. In terms of classification tables, flipping interchanges rows and columns.

We need some more tools and notions for handling with classifications. Let us define the following two mappings for $t \in tok(C)$ and $\theta \in typ(C)$:

$$typ(t) \quad := \quad \{\theta \in typ(C) : t \models_C \theta\} \tag{12.1}$$

$$tok(\theta) \quad := \quad \{t \in tok(C) : t \models_C \theta\} \tag{12.2}$$

For each token t, there is a corresponding type set $typ(t)$, containing all types which classify this token. On the other hand, one can find for each type θ a token set $tok(\theta)$ in which all tokens, which are classified by this type, are collected.

Example 12.1.3 (Type Set and Token Set) Reconsider the classification table for planets of Example 12.1.1. The type set for the token *Earth* is the set of attribute it possesses: $typ(Earth) = \{$*small, distance-near, moon-yes*$\}$. The token set of the property *small* is the set of all the planets which are classified as being small: $tok(small) = \{$*Mercury, Venus, Earth, Mars*$\}$. $\odot$

In order to pave the way for the infomorphism mechanism, which is the subject of the following section, let us introduce the notion of isomorphic classifications:

Definition 12.2 (Isomorphic Classifications)

Two classifications $C_1 = (tok(C_1), typ(C_1), \models_{C_1})$ and $C_2 = (tok(C_2), typ(C_2), \models_{C_2})$ are said to be isomorphic*, denoted by*

$$C_1 \quad \cong \quad C_2,$$

if there exists a one-to-one correspondence between their types and between their tokens, given by a pair of functions. The corresponding tokens t and t' and the corresponding types θ and θ' must then satisfy

$$t \models_{C_1} \theta \quad \textit{if and only if} \quad t' \models_{C_2} \theta'.$$

12.2 Infomorphisms

Now that classifications have been introduced, we want to connect these classifications. The goal of establishing a relation between two classifications is to move information back and forth. This is made possible by so-called infomorphisms. Barwise and Seligman point out that the classifications connected by infomorphisms may be constituted of the same objects, then information relative to the same objects is exchanged. However, these objects can be viewed from different perspectives, e. g. that of different people, different time zones, different branches of science etc. It is also possible to relate two classifications made up of different objects. This is typically the case when a distributed system is modeled by one classification and one of its parts by another classification.

An infomorphism i is a pair of contravariant functions $\langle f, g\rangle$ between two classifications C_1 and C_2. f is a function from the types of C_1 to the types of C_2. g is a function from the tokens of C_2 to the tokens of C_1. As f and g are defined to map in opposite directions, the infomorphism i is a contravariant pair of functions.[1]

Definition 12.3 (Infomorphism)

Consider two classifications C_1 and C_2 with $C_1 = (tok(C_1), typ(C_1), \models_{C_1})$ and $C_2 = (tok(C_2), typ(C_2), \models_{C_2})$. An infomorphism $i : C_1 \rightleftarrows C_2$ from C_1 to C_2 *is a contravariant pair of functions $i = \langle f, g\rangle$*

$$\begin{array}{rccl} f: & typ(C_1) & \rightarrow & typ(C_2) \\ & tok(C_1) & \leftarrow & tok(C_2) \quad : g \end{array}$$

satisfying $\forall t \in tok(C_2)$ and $\forall \theta \in typ(C_1)$

$$g(t) \models_{C_1} \theta \quad \Leftrightarrow \quad t \models_{C_2} f(\theta).$$

The image $g(t)$ of a token t of the classification C_2 is a token of the classification C_1. $g(t)$ is classified as being of some type θ of the classification C_1. Furthermore, given this type θ of classification C_1, it always has an image $f(\theta)$ in the classification C_2, where $f(\theta)$ classifies the token t of C_2. Barwise and Seligman use classification diagrams, as the one in Figure 12.2, to depict infomorphisms. The contravariant pair of functions becomes manifest.

Example 12.2.1 (Infomorphism: Propositional Logic) Let P be a finite set of propositional symbols $p_1, p_2, \ldots, p_n$ and $Q \subset P$. Then, we consider the classifications $C_Q = (\{0,1\}^{|Q|}, Fml(\mathcal{L}_Q), \models)$ and $C_P = (\{0,1\}^{|P|}, Fml(\mathcal{L}_P), \models)$ and we want to show that there is an infomorphism $i : C_Q \rightleftarrows C_P$ between them, with

$$\begin{array}{rccl} f: & Fml(\mathcal{L}_Q) & \rightarrow & Fml(\mathcal{L}_P), \\ g: & \{0,1\}^{|P|} & \rightarrow & \{0,1\}^{|Q|}. \end{array}$$

This means that for $t \in \{0,1\}^{|P|}$ and $\theta \in Fml(\mathcal{L}_Q)$, it must hold that

$$g(t) \models \theta \quad \Leftrightarrow \quad t \models f(\theta).$$

[1] Here, we adopt a slightly different naming as in (Barwise & Seligman, 1997). The authors denote the pair of functions of an infomorphism by $f^{\wedge}$ (function between types) and $f^{\vee}$ (function between tokens). The infomorphism itself is usually named f.

$$\begin{tikzcd} typ(C_1) \arrow[r, "f"] \arrow[d, dash, "\models_{C_1}"] & typ(C_2) \arrow[d, dash, "\models_{C_2}"] \\ tok(C_1) & tok(C_2) \arrow[l, "g"] \end{tikzcd}$$

Figure 12.3: Illustration of an infomorphism using classification diagrams

f is a function from $Fml(\mathcal{L}_Q)$ to $Fml(\mathcal{L}_P)$. As $Q \subset P$, the formulae of $Fml(\mathcal{L}_Q)$ are only constituted of some of the propositional symbols the formulae of $Fml(\mathcal{L}_P)$ may be constituted of. Therefore, every formula of $Fml(\mathcal{L}_Q)$ is also a formula of $Fml(\mathcal{L}_P)$. So f is simply defined as

$$f(\theta) \quad := \quad \theta.$$

g is the projection from $\{0,1\}^{|P|}$ onto $\{0,1\}^{|Q|}$. Only the valuations relative to the set Q of propositional symbols are taken into account. This is done by considering a valuation $t' \in \{0,1\}^{|Q|}$ which provides the same values for the propositions in Q as the original valuation $t \in \{0,1\}^{|P|}$. The two valuations t' and t are said to be equivalent relative to Q, written $t' \equiv_Q t$, if they share the same values for the propositional symbols in Q. g is defined as

$$g(t) \quad := \quad t' \in \{0,1\}^{|Q|}, \text{ such that } t' \equiv_Q t.$$

This corresponds to the y-tuple known from Section 2.4, denoted by $f[y]$ in Equation 2.5.

Now we are going to bring to proof that $i = \langle f, g \rangle$ is actually an infomorphism.

- $g(t) \models \theta \Rightarrow t \models f(\theta)$:
  Assume that a formula $\theta \in Fml(\mathcal{L}_Q)$, containing propositional symbols from Q, is satisfied by the projection $g(t)$ of some valuation $t \in \{0,1\}^{|P|}$. Then, $\theta = f(\theta)$ is also satisfied by the original valuation t, since t merely provides some additional values for the propositional symbols in $P \setminus Q$, which, however, do not play any role.

- $g(t) \models \theta \Leftarrow t \models f(\theta)$:
  Assume that the valuation $t \in \{0,1\}^{|P|}$, which contains values for the propositional symbols in P, satisfies some formula $f(\theta) = \theta$. $\theta \in Fml(\mathcal{L}_Q)$ contains

only propositional symbols from $Q \subset P$. Projecting the valuation t to Q results in a valuation $g(t) = t' \in \{0,1\}^{|Q|}$. The projected valuation $g(t)$ obviously satisfies θ, since it carries the same values for the propositions in Q (and only these propositional symbols matter) as the original valuation t.

As both directions hold, the pair $i = \langle f, g \rangle$ of contravariant functions is an infomorphism, according to Definition 12.3. $\odot$

An infomorphism $i = \langle f, g \rangle$ is an isomorphism if both functions f and g are bijections. Therefore, an isomorphism might also be called bijective infomorphism. Now the notion of isomorphic classifications (Definition 12.2) becomes clear: Two classifications are isomorphic iff there is an bijective infomorphism between them.[2]

Lemma 12.4 *Consider two infomorphism i and j. Let $i = \langle f_i, g_i \rangle$ be an infomorphism from $C_1 = (tok(C_1), typ(C_1), \models_{C_1})$ to $C_2 = (tok(C_2), typ(C_2), \models_{C_2})$, $i : C_1 \rightleftarrows C_2$. Let $j = \langle f_j, g_j \rangle$ be an infomorphism from C_2 to $C_3 = (tok(C_3), typ(C_3), \models_{C_3})$, $j : C_2 \rightleftarrows C_3$. Then, the* composition *$ij : C_1 \rightleftarrows C_3$ of the two infomorphism i and j is an infomorphism from C_1 to C_3 and it is defined by $ij = \langle f_j \circ f_i, g_i \circ g_j \rangle$.*

Proof. Let $t \in tok(C_3)$ and $\theta \in typ(C_1)$. Assume that $t \models_{C_3} f_j(f_i(\theta))$, then $g_j(t) \models_{C_2} f_i(\theta)$, hence $g_i(g_j(t)) \models_{C_1} \theta$. The converse direction follows in the same way: Assume that $g_i(g_j(t)) \models_{C_1} \theta$, then $g_j(t) \models_{C_2} f_i(\theta)$, hence $t \models_{C_3} f_j(f_i(\theta))$.[3] $\square$

The composition of infomorphisms gives rise to another notion, introduced by Barwise and Seligman, namely that of a commuting diagram. In the following, infomorphisms will have special properties if their corresponding diagram commutes. Consider three infomorphism $i : C_1 \rightleftarrows C_2$, $j : C_2 \rightleftarrows C_3$ and $k : C_1 \rightleftarrows C_3$, as depicted in Figure 12.4 below. To say that this diagram commutes is to assert that $k = ij$, i. e. the infomorphism k is the composition of the two infomorphism i and j. Graphically speaking, you can go either way around the triangle and get the same result. As seen in the above Lemma 12.4, $f_k = f_j \circ f_i$ and $g_k = g_i \circ g_j$.

12.3 Information Channels

In the two preceding sections, the concept of classifications has been introduced as a formalism for representing information in its dual structure. Furthermore, infomorphisms, which relate classifications, have been presented for transporting information. As already mentioned in the introduction of this chapter, these notions serve to define information channels. Barwise and Seligman use information channels, often simply called channels, to model information flow in distributed systems.

[2] The inverse functions f^{-1} and g^{-1} need to exist between isomorphic classifications, as they allow to directly link their types and tokens. Definition 12.2 gives rise to the following equalities: $f(t) = t', g^{-1}(\theta) = \theta'$ and $f^{-1}(t') = t, g(\theta') = \theta$.

[3] This trivial proof has been omitted in (Barwise & Seligman, 1997). This is why it can be found here.

$$\begin{tikzcd}
 & C_2 \arrow[dl, shift left, "g_i"] \arrow[dr, shift left, "f_j"] & \\
C_1 \arrow[ur, shift left, "f_i"] \arrow[rr, shift left, "f_k"] & & C_3 \arrow[ul, shift left, "g_j"] \arrow[ll, shift left, "g_k"]
\end{tikzcd}$$

Figure 12.4: Diagram commutes if $k = ij$

Definition 12.5 (Channel)

A channel $\mathcal{C}$ *is an indexed family* $\{i_j : C_j \rightleftarrows C\}_{j\in J}$ *of infomorphisms. J is some index set, C is a common codomain, called the* core *of* $\mathcal{C}$*.*
The tokens of the core C are called connections*; a connection c is said to* connect *the tokens* $g_j(c)$ *for* $j \in J$*. The classifications* C_j*,* $j \in J$*, are called* component classifications *of the channel* $\mathcal{C}$*. A channel with index set* $\{0, \ldots, n-1\}$ *is called an n-ary channel.*

Barwise and Seligman describe a channel $\mathcal{C}$ to be a mathematical model for a distributed system – a whole made up of parts such that information can flow. The core C of the channel $\mathcal{C}$ represents the whole, the component classifications C_j the parts. The infomorphisms i_j are relationships between the whole and its parts. Channels can also be depicted by diagrams. Figure 12.5 shows a channel of arity four: There are four infomorphisms i_1 to i_4 ($|J| = 4$) implied in this channel, each of them is constituted of a contravariant pair of functions $\langle f_j, g_j \rangle$. A token c of the core C, called connection, connects the various tokens t_j of the indexed classifications: $t_j = g_j(c)$.

In order to exemplify that a token c of the core acts as a connection between the components of the system (the t_j), the authors mention in (Barwise & Seligman, 1997) that

> *"a light bulb, at a time is connected by a switch at a time if they are connected by being parts of, say, a single flashlight. It is this connection that allows one to carry information about the other."*

Reconsider Example 12.2.1, where an infomorphism, constituted by two propositional logic classifications, has been shown. This example will now be extended to illustrate a channel in action:

Example 12.3.1 (Channel: Propositional Logic) Let P be a finite set of propositional symbols $p_1, p_2, \ldots, p_n$ and $Q, R \subset P$. Then, we consider the three classifications $C_Q = (\{0,1\}^{|Q|}, Fml(\mathcal{L}_Q), \models)$, $C_R = (\{0,1\}^{|R|}, \mathcal{L}_R, \models)$ and $C_P =$

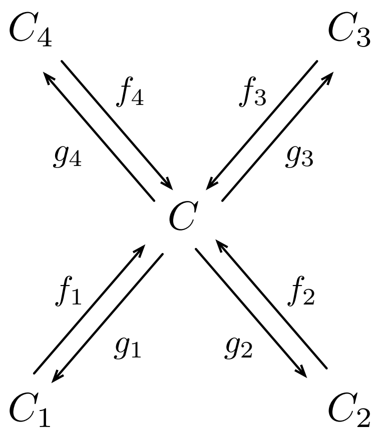


Figure 12.5: A channel of arity four

$(\{0,1\}^{|P|}, Fml(\mathcal{L}_P), \models)$ and the infomorphisms $i_1 : C_Q \rightleftarrows C_P$ and $i_2 : C_R \rightleftarrows C_P$ between them. The core of this channel is the classification C_P. We will see that its tokens are connections between C_Q and C_R. Take for instance $P = \{p_1, p_2, p_3\}$, $Q = \{p_1, p_2\}$ and $R = \{p_2, p_3\}$. We will now start with a formula of $Fml(\mathcal{L}_Q)$ and see how the information contained in this formula is moved from the classification C_Q to the classification C_R and what happens in the core. The processes is illustrated in Figure 12.6.

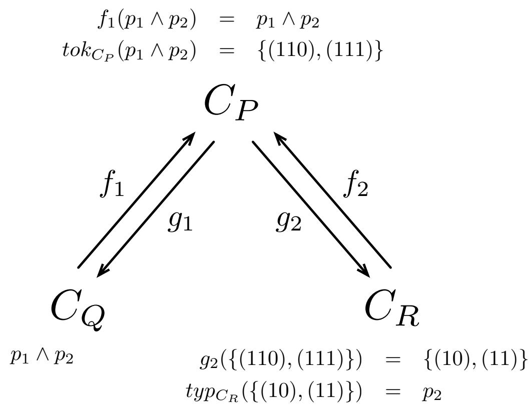


Figure 12.6: Moving $p_1 \wedge p_2$ from C_Q to C_R using a channel

The formula $\theta = p_1 \wedge p_2$ describes information by means of classification C_Q, namely that both p_1 and p_2 are set to 1. Using the infomorphism i_1, θ is expressed in the classification C_P, by $f_1(p_1 \wedge p_2) = p_1 \wedge p_2$. Now the information we want to transmit, and which is for the moment only classified by its type, is located in the core of the channel. It is ready to be transported to another classification. For that purpose, the corresponding token set has to be generated. (Remember that the incoming arrows of the core are the functions f_j from sets of types to sets of types and the outgoing arrows are the functions g_j from sets of tokens to sets of tokens). So we get $tok_{C_P}(p_1 \wedge p_2) = \{(110), (111)\}$. This set of tokens of the core contains the connections of our specific situation of transporting $p_1 \wedge p_2$. They describe the information present in the whole system, whereas the component classifications only describe parts of the whole system. Their tokens can only be connected by going via the core. The infomorphism i_2 requires an application of g_2, since we possess information in C_P and want that information to be expressed by C_R. So we obtain the information $g_2(\{(110), (111)\}) = \{(10), (11)\}$. This set of tokens is classified by $p_2 = typ_{C_R}(\{(110), (111)\})$, which is not a surprise. When moving information about the propositional variables p_1 and p_2 to some classification which only expresses information about p_2 and p_3, the information about p_1 will be "forgotten", the information about p_2 will be retained and nothing can be said about p_3. $\odot$

12.3.1 Refinement

As already mentioned in the introduction, the information flow in a distributed system can be modeled in different ways, i. e. by different channels. The channels differ from each other in their classifications and consequently also in their infomorphisms. If there are different channels for modeling a system, it might be interesting to compare these channels. There may be finer or coarser ones. Intuitively, a more "refined" channel provides a more reliable information flow, as more details of the system are modeled. Barwise and Seligman state that a channel is a refinement of another channel, if they are connected by a special kind of infomorphism, a so-called refinement infomorphism. The following definition formalizes the refinement relation between channels:

Definition 12.6 (Refinement)

Let $\mathcal{C} = \{i_j : C_j \rightleftarrows C\}_{j \in J}$ and $\mathcal{C}' = \{i'_j : C_j \rightleftarrows C'\}_{j \in J}$ be two channels with the same component classifications C_j, $j \in J$. A refinement infomorphism *r from channel $\mathcal{C}'$ to channel $\mathcal{C}$ is an infomorphism $r : C' \rightleftarrows C$ such that for each $j \in J$, the infomorphism i_j of $\mathcal{C}$ can be expressed as $i'_j r$, where i'_j is an infomorphism of $\mathcal{C}'$:*

$$i_j \quad = \quad i'_j r.$$

The channel $\mathcal{C}'$ is a refinement *of the channel $\mathcal{C}$ if there is a refinement infomorphism r from $\mathcal{C}'$ to $\mathcal{C}$.*

Example 12.3.2 (Refinement) In Figure 12.7, a situation with two channels $\mathcal{C} = \{i_j : C_j \rightleftarrows C\}_{j \in J}$ and $\mathcal{C}' = \{i'_j : C_j \rightleftarrows C'\}_{j \in J}$ is given, where $J = \{1, 2\}$. Channel $\mathcal{C}$ is depicted by a dashed line, channel $\mathcal{C}'$ by a dotted line. There are two component classifications, namely C_1 and C_2 with two infomorphisms, each. Note that for reasons of legibility, the arrows are labeled with the name of the infomorphism, not with the names of the contravariant pair of functions constituting the infomorphism. r is the refinement infomorphism from $\mathcal{C}'$ to $\mathcal{C}$: $i_1 = i'_1 r$ and $i_2 = i'_2 r$.

$\odot$

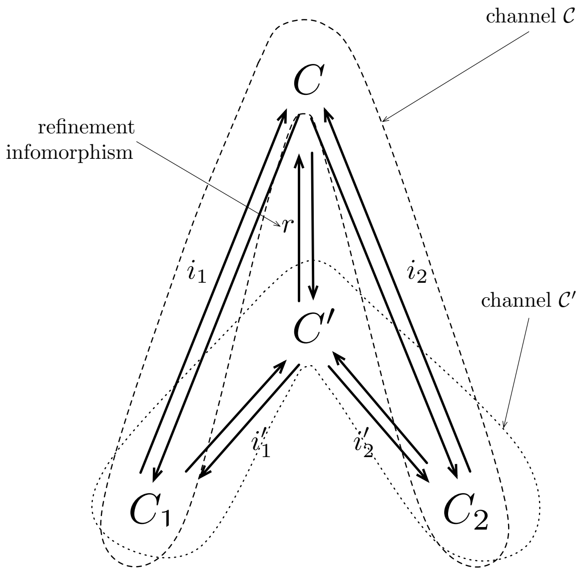


Figure 12.7: $\mathcal{C}'$ is a refinement of $\mathcal{C}$

12.4 Information in Context

After having introduced the basics of Barwise and Seligman's theory of information flow, we will now extend the theory in the following way: We give more expressive power to classifications by enriching them with closure operators and Galois connections, so that we finally dispose of a lattice of enriched classifications. This is mainly based on the results from formal concept analysis, as presented in (Davey & Priestley, 2002). The new features allow for a more specific meaning of classifications: An enriched classification formalizes the two aspects of information, namely the representation of information (syntactic point of view) and the meaning of information (semantic point of view). Thereby, a definition of what we understand by a piece of information is needed. This notion does not appear in (Barwise & Seligman, 1997), but it will be very useful for linking the theory of information flow to information algebras in Section 12.5. The content of this section mainly bears from (Kohlas, 2002) and (Davey & Priestley, 2002).

12.4.1 Contexts

First of all, let us rename classifications. We want to point out that there is a real difference between the classifications introduced by Barwise and Seligman and the entity that we have called "enriched classifications" above. Therefore, enriched classifications will be called contexts in the following. This is the term used in formal concept analysis. A context is given by a triple (G, M, I), where G and M are sets and $I \subseteq G \times M$. The elements of G are called objects (the G stands for the German "Gegenstände"), the elements of M are called attributes (the letter M comes from the German "Merkmale"). I is a binary relation, read as some object having a specific attribute. We thus get the same dual structure as for classifications, with G corresponding to the tokens of a classification and M to its types. However, there is something more.

Extent and Intent

An important notion of formal concept analysis is that of a concept of a context. A concept of a context consists of an ordered pair (A, B), where $A \subseteq G$ is called the extent and $B \subseteq M$ is called the intent. Furthermore, the sets A and B have to be closed (see Section 12.4.2 below). The extent of a concept consists of objects of G, the intent of a concept is the collection of all attributes of M shared by the objects of the extent. Here, we already see that contexts and their concepts provide important properties of information that do not occur in (Barwise & Seligman, 1997). Furthermore, contexts make it possible that questions come into the play. Questions are an important, even very basic concept of information theory, which has been completely neglected by Barwise and Seligman.

Towards Information Modeling by Contexts

We will use contexts for information modeling. Information is usually expressed by a formal language which gets a specific meaning by a valuation which attributes values to variables. Instead of (G, M, I), we will denote contexts in the following by a triple $(\mathcal{M}, \mathcal{L}, \models)$, where $\mathcal{M}$ is to be thought as a set of valuations or models, $\mathcal{L}$ as a language (a set of sentences) and $\models \subseteq \mathcal{L} \times \mathcal{M}$ as a binary relation between sentences and valuations. When a valuation $v \in \mathcal{M}$ satisfies a sentences $s \in \mathcal{L}$, meaning is given to the language $\mathcal{L}$. We write $v \models s$ instead of $(v, s) \in \models$.

Definition 12.7 (Context)

A context *is a triple $(\mathcal{M}, \mathcal{L}, \models)$, where*

- *$\mathcal{M}$ is a set of valuations v,*
- *$\mathcal{L}$ is a set of sentences s,*
- *$\models \subseteq \mathcal{M} \times \mathcal{L}$ is a binary relation.*

We say that v satisfies *s if $v \models s$.*

It is possible to determine from a given set $S \subseteq \mathcal{L}$ of sentences the corresponding set of valuations, which are described by S. This set $\hat{r}(S)$ contains all the valuations which satisfy every sentence in S. It is also called the model of S:

$$\hat{r}(S) = \{v \in \mathcal{M} : \forall s \in S, v \models s\}. \tag{12.3}$$

On the other hand, one can also start with a set $V \subseteq \mathcal{M}$ of valuations and have a look at all the sentences $s \in \mathcal{L}$ that describe V. In this set $\check{r}(V)$, there are sentences s whose satisfying valuations contain V, but the sentences $s \in \check{r}(V)$ may also be satisfied by other valuations, which are not in V. V is the set of valuations that all sentences in $\check{r}(V)$ agree on. $\check{r}(V)$ is also called the theory of V:

$$\check{r}(V) = \{s \in \mathcal{L} : \forall v \in V, v \models s\}. \tag{12.4}$$

Figure 12.8 illustrates the two mappings: $\check{r}$ maps a subset of $\mathcal{M}$ to a subset of $\mathcal{L}$. On the other hand, $\hat{r}$ maps a subset of $\mathcal{L}$ to a subset of $\mathcal{M}$. The sets which have not changed after a sequential application of both mappings (making a "round-trip") have interesting properties, as we will see in a moment.

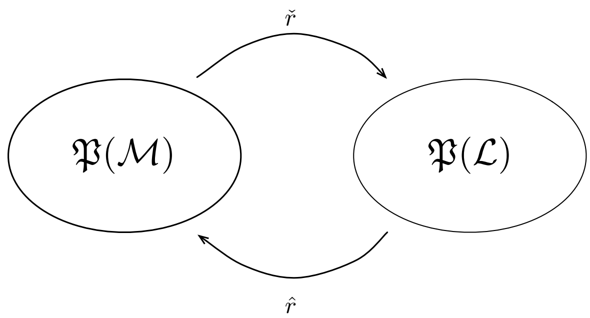


Figure 12.8: $\check{r}$ and $\hat{r}$ map sets to sets

There are some important properties of $\hat{r}$ and $\check{r}$, summarized in the following lemma, taken from (Davey & Priestley, 2002):

Lemma 12.8 *For a context $(\mathcal{M}, \mathcal{L}, \models)$, the sets $V, V_j \subseteq \mathcal{M}$ of valuations and the sets $S, S_j \subseteq \mathcal{L}$ of sentences, where $j \in J$, the following four dual pairs of properties hold:*

(P1)	$V \subseteq \hat{r}(\check{r}(V))$	$S \subseteq \check{r}(\hat{r}(S))$
(P2)	$V_1 \subseteq V_2$ *implies* $\check{r}(V_1) \supseteq \check{r}(V_2)$	$S_1 \subseteq S_2$ *implies* $\hat{r}(S_1) \supseteq \hat{r}(S_2)$
(P3)	$\check{r}(V) = \check{r}(\hat{r}(\check{r}(V)))$	$\hat{r}(S) = \hat{r}(\check{r}(\hat{r}(S)))$
(P4)	$\check{r}(\bigcup_{j \in J} V_j) = \bigcap_{j \in J} \check{r}(V_j)$	$\hat{r}(\bigcup_{j \in J} S_j) = \bigcap_{j \in J} \hat{r}(S_j)$

The first property tells that an application of both mappings always results in a set of valuations or sentences which contains the original set. Property 2 states that switching from valuations to sentences, or vice versa, inverts the inclusion relation which holds between two sets of valuations or two sets of sentences. From property 3 it is known that the set $\check{r}(V)$ of sentences, which corresponds to a given set V of valuations, is not changed any more when both mappings are applied at that stage. The same holds for the dual case with a given set S of sentences. Finally, by the fourth property, we can either take the union of a given set of valuations or sentences, and pass afterwards to the dual form of representation, or the translation to the other representation is done at first for each single set, followed by the intersection of the translated sets.

Context as Scheme of Choice

A context may be seen as a frame to formulate questions and find answers to these questions. One of these questions is typically "Which is the sentence s needed to describe some set of valuations?". This translates in the classification case to "How is a set of tokens classified?". In the planets example, we get "Which properties does some set of planets have?", or, for propositional logic, "What are the formulae satisfied by some Boolean vector?". Another question might be "Given some set of sentences, which are their satisfying valuations?". The classification formulation is "What does a given set of types classify?". Such a question translates for the planet example to "Which planets are described by some given properties?", or, for the propositional logic case, to "Which set of valuations satisfies some set of formulae?". These questions are formalized by the notion of a concept in a context, introduced above. A concept in a context is a pair (V, S), where the extent $V \in \mathcal{M}$ and the intent $S \in \mathcal{L}$ are of the same context $(\mathcal{M}, \mathcal{L}, \models)$. See also Section 12.4.3 for a our information-theoretic approach to a concept in a context. When we are looking for the value of the intent S or the extent V of some concept, the context tells that it must be a subset of $\mathcal{L}$ (first type of questions) or of $\mathcal{M}$ (second type of questions). From this point of view, the context acts as a scheme of choice where one or several unknown elements are selected and the goal is to get to know which one(s).

Contexts and Information

At the same time, a context is also a domain to express information. When looking for some valuation $v \in \mathcal{M}$, a source may provide further information about v, by specifying a set $S \subseteq \mathcal{L}$ of sentences which describes the unknown valuation. So, a context allows to express information about the unknown v, e.g. that for all $s \in S$, $v \models s$ holds. The information provided by S limits the possible values of the unknown valuation v to the set $\hat{r}(S)$, i.e. to all valuations which satisfy all the sentences in S. Once that we dispose of this set of valuations, we can look for the corresponding set of sentences which describe all the valuations of $\hat{r}(S)$. This set is $\check{r}(\hat{r}(S))$. Later in this section, we require information to satisfy $S = \check{r}(\hat{r}(S))$. Furthermore, when starting with a set of valuations M, in order to be called information, $M = \hat{r}(\check{r}(M))$ must hold.

12.4.2 Closure Operators

We will now define two operators $C^{\models}$ (mapping sets of sentences to sets of sentences) and $C_{\models}$ (mapping sets of valuations to sets of valuations) and verify that they are closure operators. Thereafter, information will be defined to be constituted of closed sets of valuations and closed sets of sentences (Subsection 12.4.3).

For $V \subseteq \mathcal{M}$ we define

$$C_{\models}(V) \quad := \quad \hat{r}(\check{r}(V)), \tag{12.5}$$

and for $S \subseteq \mathcal{L}$ we define

$$C^{\models}(S) \quad := \quad \check{r}(\hat{r}(S)). \tag{12.6}$$

These two operators are actually closure operators, as shown in the proof of the following lemma:

Lemma 12.9 *$C^{\models}$ and $C_{\models}$ are closure operators, as they fulfill the following three properties of a closure operator:*

(C1)	$V \subseteq C_{\models}(V)$	*and*	$S \subseteq C^{\models}(S)$,
(C2)	$C_{\models}(C_{\models}(V)) = C_{\models}(V)$	*and*	$C^{\models}(C^{\models}(S)) = C^{\models}(S)$,
(C3)	$V_1 \subseteq V_2$ *implies* $C_{\models}(V_1) \subseteq C_{\models}(V_2)$	*and*	$S_1 \subseteq S_2$ *implies* $C^{\models}(S_1) \subseteq C^{\models}(S_2)$.

Proof. The proof is very straightforward, based on the properties of $\hat{r}$ and $\check{r}$, given in Lemma 12.8. We will only show the proof for $C_{\models}$. The proof for $C^{\models}$ follows in the same way.

1. (C1) is property (P1) of Lemma 12.8: $V \subseteq \hat{r}(\check{r}(V)) = C_{\models}(V)$.
2. (C2) is proven by considering the equality (P3) of Lemma 12.8, $\check{r}(V) = \check{r}(\hat{r}(\check{r}(V)))$. Now $\hat{r}$ is applied to both sides of the equals sign:
$$\hat{r}(\check{r}(V)) = \hat{r}((\check{r}(\hat{r}(\check{r}(V)))) \quad \Leftrightarrow \quad C_{\models}(V) = C_{\models}(C_{\models}(V)).$$
3. (C3) is proven by applying two times property (P2) of Lemma 12.8 and the fact that $\check{r}(V_1), \check{r}(V_2) \subseteq \mathcal{L}$:
$$\begin{aligned} V_1 \subseteq V_2 \quad &\Rightarrow \quad \check{r}(V_1) \supseteq \check{r}(V_2) \\ &\Rightarrow \quad \hat{r}(\check{r}(V_2)) \supseteq \hat{r}(\check{r}(V_1)) \\ &= \quad C_{\models}(V_1) \subseteq C_{\models}(V_2). \end{aligned}$$

□

Now we can define what it means for a set V of valuations or for a set S of sentences to be closed:

Definition 12.10 ($\models$-closed Sets)

Sets $V \subseteq \mathcal{M}$ and $S \subseteq \mathcal{L}$ are called $\models$-closed, *if*

$$V = C_{\models}(V) \text{ or } S = C^{\models}(S),$$

respectively. $C_{\models}(V)$ is called the closure *of V, $C^{\models}(S)$ the closure of S.*

Note that by (P3) of Lemma 12.8, every set S of sentences determines a $\models$-closed set of valuations: $\hat{r}(S) = C_{\models}(\hat{r}(S))$. On could say that this is the information expressed by S. In the same way, any set V of valuations determines a $\models$-closed set of sentences – the theory of V: $\check{r}(V) = C^{\models}(\check{r}(V))$.

Let us illustrate the closure operator with the examples from the previous sections:

Example 12.4.1 (Closure Operators)

1. *Planets:* Looking at the table given in Figure 12.1, one can observe that $Earth \neq \hat{r}(\check{r}(Earth))$. $\{Earth\}$ is therefore not closed, it is a strict subset of its closure: $\{Earth\} \subset C^{\models}(\{Earth\}) = \{Earth,\ Mars\}$, see property 1 of Lemma 12.9. But $\{Earth,\ Mars\}$ is a closed set, which illustrates the second property of the above lemma. For exemplifying the third property, we will start with two sets of sentences: $\{moon\text{-}yes\}$ and $\{size\text{-}large,\ moon\text{-}yes\}$, where the former is a subset of the latter. The same has to hold for their closures. After a consultation of the table given in Figure 12.1, it is clear that $\{moon\text{-}yes\} = C_{\models}(\{moon\text{-}yes\}) \subset C_{\models}(\{size\text{-}large,\ moon\text{-}yes\}) = \{size\text{-}large,\ distance\text{-}far,\ moon\text{-}yes\}$.

2. *Powersets:* The powerset context introduced in Example 12.1.2 is $(A, \mathfrak{P}(A), \in)$, where the valuations are elements of some reference set A, the sentences are elements of the power set $\mathfrak{P}(A)$ and therefore subsets of A. The relation is $\in$. This context has the interesting property that every of its sets of valuations is closed.

   When starting with some set V of valuations, the corresponding theory $\check{r}(V)$ is constituted of all sentences $s \in \mathfrak{P}(A)$ which contain every $v \in V$. In order to obtain the corresponding set of valuations of this theory, the intersection of all elements of $\check{r}(V)$ is taken. This brings us back to the original set V and actually, $V = C_{\models}(V)$ for any set V of valuations.

   However, consider some set S of sentences whose elements s are subsets of A, i. e. elements of $\mathfrak{P}(A)$. The corresponding set of valuations is composed of those elements of A which occur in each of the $s \in S$: $\hat{r}(S) = \bigcap_{s\in S} s$. Looking at the closure of S means generating the theory of $\hat{r}(S)$. One obtains the set of all supersets of this intersection. $C^{\models}(S) = \{s' \subseteq A : \hat{r}(S) \subseteq s'\} \supseteq S$.

3. *Propositional Logic:* In Example 12.1.2, the propositional logic context $(\{0,1\}^{|P|}, Fml(\mathcal{L}_P), \models)$ has been introduced. The valuations are the propositional logic valuations $v \in \{0,1\}^{|P|}$, the sentences are wff $\gamma \in \mathcal{L}_P$ and the relation is the so-called consequence relation $\models$, indicating that some valuation v satisfies some formula γ, i. e. v is a model of γ.

   For some set of formulae S, the corresponding valuation set $\hat{r}(S)$ consists of those valuations v which satisfy every $\gamma \in S$. So the closure $C^{\models}(S)$ is constituted by those formulae which are at least satisfied by the valuations in $\hat{r}(S)$. The formulae in $C^{\models}(S)$ may have further models, but every formula must have $\hat{r}(S)$ among its models. The set $C^{\models}(S)$ is the set of all logical consequences of S: Let $S = \{p_1 \wedge p_2\}$, so $\hat{r}(S) = \{(11)\}$ and therefore $C^{\models}(S) = \{p_1, p_2, p_1 \vee p_2, p_1 \wedge p_2\}$.

   On the other hand, when starting with a set of models V, we will always have $V = C_{\models}(V)$, for a finite set P of propositions. In propositional logic, a set of valuations, modeling some set of formulae, is always $\models$-closed.

$\odot$

12.4.3 Pieces of Information in a Context

The above examples show that we can either start with some set S of sentences and conclude $C^{\models}(S)$, or we may start with some set V of valuations and conclude $C_{\models}(V)$. We now require that information in a context has to be closed. A piece of information may be determined by a closed set $C^{\models}(S)$ of sentences, to which always belongs a set $\hat{r}(C^{\models}(S)) = \hat{r}(S)$ of valuations, by (P3) of Lemma 12.8. Alternatively, a piece of information may be fixed by a closed set $C_{\models}(V)$ of valuations. The theory corresponding to this piece of information is given by $\check{r}(C_{\models}(V)) = \check{r}(V)$, again by (P3) of Lemma 12.8. This may be formalized by the following definition of a piece of information in a context. Note that what we call here "information" is called "concept" in formal concept analysis.

Definition 12.11 (Information in a Context)

A pair (V, S) is called a (piece of) information in a context $C = (\mathcal{M}, \mathcal{L}, \models)$*, if $V \subseteq \mathcal{M}$, $S \subseteq \mathcal{L}$ and*

$$V = \hat{r}(S) \text{ and } S = \check{r}(V).$$

$\mathcal{I}(\mathcal{M}, \mathcal{L}, \models)$ denotes the set of all pieces of information in the context $(\mathcal{M}, \mathcal{L}, \models)$.

Clearly, V and S must be $\models$-closed sets. So a piece of information (V, S) always has the property that $V = C_{\models}(V)$ and $S = C^{\models}(S)$. The above definition tells what information is, namely that it is always located in a context. Furthermore, the definition states that information may be either given by a $\models$-closed set V of valuations, which allows to determine the corresponding set S of sentences by $\check{r}(V) = S$. On the other hand, a piece of information is also fully determined by a $\models$-closed set S of sentences, from which one can deduce the set $\hat{r}(S) = V$ of valuations.

However, note that any set S of sentences gives rise to a piece of information, namely to $(\hat{r}(S), C^{\models}(S))$. By (P3) of Lemma 12.8, every set $\hat{r}(S)$ is closed. In the same way, any set V of valuations determines the piece of information $(C_{\models}(V), \check{r}(V))$.

There are two special pieces of information we want to point out. It is the null information, which is often considered as the contradiction in logics, and the vacuous information, often called tautology in logics. The null information is a piece of information determined by a set $S \subseteq \mathcal{L}$ of sentences, which are not satisfied by any valuations, i.e. $\hat{r}(S) = \emptyset$. By (P3) of Lemma 12.8, it holds that $\hat{r}(S) = \hat{r}(\check{r}(\hat{r}(S))) = C_{\models}(\emptyset)$. So the null information is given by the piece of information $(C_{\models}(\emptyset), S)$. The vacuous information is a piece of information which is given by a set of sentences which is satisfied by the set $\mathcal{M}$ of all possible valuations in some context $C = (\mathcal{M}, \mathcal{L}, \models)$. So the vacuous information is the piece of information $(\mathcal{M}, \check{r}(\mathcal{M}))$.[4] These two special pieces of information will be illustrated by three examples:

[4] $(\mathcal{M}, C^{\models}(\emptyset))$ is an alternative form of the vacuous information for the tautology in logics. See point three of Example 12.4.2 for further details.

Example 12.4.2 (Null Information, Vacuous Information)

1. *Planets:* The null information in the planets context is $(\emptyset, \{$*size-small, size-medium, size-large, distance-near, distance-far, moon-yes, moon-no*$\})$. There is no planet having all the seven properties. It is important to list all of them, otherwise, it would not be a closed set. The vacuous information is given by the set of eight planets, which have no properties in common: $(\{$*Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune*$\}, \emptyset\}$.

2. *Powersets:* In the powerset context $(A, \mathfrak{P}(A), \in)$, the only possibility to provide a set of sentences satisfied by no valuations is the empty set. It is the only set of sentences which has no corresponding valuations, as no element of A is contained in the empty set: $a \notin A, \forall a \in A$. Therefore, the null information is $(\emptyset, \emptyset)$. The vacuous information is the reference set A (the set of all possible valuations of the context) and the corresponding theory $\check{r}(A)$. The only set containing all elements of A is A itself. So (A, A) is the vacuous information of the powerset context.

3. *Propositional Logic:* A set S of propositional logic formulae can be thought as a conjunction over the formulae, as all formulae of S need to be satisfied by a valuation for the set S to be satisfied. Propositional logic formulae which are not satisfiable do not dispose of any satisfying valuations (models). A formula which does not have any model is called a contradiciton. The set $Fml(\mathcal{L}_P)$ of all formulae includes contradictions, such as $\perp$ or $p \wedge \neg p$, hence $Fml(\mathcal{L}_P)$ is not satisfiable. By (P1) of Lemma 12.8, $Fml(\mathcal{L}_P) = C^{\models}(Fml(\mathcal{L}_P))$, so the set of all formulae is closed. So the null information in propositional logic is $(\emptyset, Fml(\mathcal{L}_P))$.

   A formula which is satisfied by all valuations of $\{0,1\}^{|P|}$ is referred to as tautology in propositional logic. A tautology is entailed by the empty set of formulae. So $\hat{r}(\emptyset) = \{0,1\}^{|P|}$. Furthermore, the set $\check{r}(\hat{r}(\emptyset)) = C^{\models}(\emptyset)$ is the set of all tautologies. The vacuous piece of information is thus $(\{0,1\}^{|P|}, C^{\models}(\emptyset)))$.

$\odot$

Now that it is clear that there are two "extreme" pieces of information, we want to show that there is also an order in the context, between its pieces of information.

12.4.4 Natural Order of Information

In contrast to the partial order established in formal concept analysis, we define some piece of information (V', S') to be richer in information than a piece (V, S) if V' limits the set of possible valuations of the corresponding sentences more than V, i. e. $V' \subseteq V$. In formal concept analysis, the reverse order is considered, but we have chosen this order, having the scheme of choice in mind where the richer information (providing less valuations) reduces the uncertainty about the unknown element more than the poorer information (given by more valuations). By property (P2) of Lemma 12.8, the partial order can also be expressed on the level of sentences. A piece of information (V, S) provides less information than another piece (V', S') if $S \subseteq S'$. Every sentence asserted by the first piece of information is also asserted by the second one. Thus, there is an order established between all the pieces of information in the context $(\mathcal{M}, \mathcal{L}, \models)$, as given by the following definition:

Definition 12.12 (Partial Order Between Information in a Context)

Given two pieces of information (V, S) and (V', S') of the set $\mathcal{I}(\mathcal{M}, \mathcal{L}, \models)$ of all pieces of information in the context $(\mathcal{M}, \mathcal{L}, \models)$. The partial order

$$(V, S) \leq (V', S')$$

is defined

- *on the level of valuations by $V \supseteq V'$ and*
- *on the level of sentences by $S \subseteq S'$.*

In that case, (V, S) is said to be the coarser information*,* less informative *or* poorer in information *than (V', S'). Consequently, (V', S') is called the* finer information*,* more informative *or* richer in information *than (V, S).*

$\leq$ is a partial order on $\mathcal{I}(\mathcal{M}, \mathcal{L}, \models)$, as $\subseteq$ is known to define a partial order in a set of subsets, see (Davey & Priestley, 2002). We have

$$(\mathcal{M}, \check{r}(\mathcal{M})) \leq (V, S) \leq (C_{\models}(\emptyset), \mathcal{L}) \tag{12.7}$$

for all pieces of information (V, S) in the context $(\mathcal{M}, \mathcal{L}, \models)$. The vacuous information $(\mathcal{M}, \check{r}(\mathcal{M}))$ is the poorest possible information in a context, as the set of valuations of every piece of information is a subset of $\mathcal{M}$ and thus more informative than the vacuous information. The null information $(C_{\models}(\emptyset), \mathcal{L})$ must be the top information in a context, as for all pieces of information (V, S) in this context, it holds that $S \subseteq \mathcal{L}$. However, $(C_{\models}(\emptyset), \mathcal{L})$ is not really a piece of information in the sense that its sentences limit the possible values of the unknown valuation. As all sentences of the context are considered, one cannot really speak of a limitation. It is nevertheless mentioned for technical reasons. Note that the finest pieces of information among the non-null information are those which are atomic (see Section 6.4). Atomic pieces of information are generated by single valuations, i. e. $(C_{\models}(v), \check{r}(C_{\models}(v)))$ for $v \in \mathcal{M}$. These pieces of information are the richest in information among all (V, S) of $\mathcal{I}(\mathcal{M}, \mathcal{L}, \models)$ in the sense that $(C_{\models}(v), \check{r}(C_{\models}(v))) \leq (V, S)$ either implies $(V, S) = (C_{\models}(v), \check{r}(C_{\models}(v)))$ or $(V, S) = (C_{\models}(\emptyset), \mathcal{L})$.

We will conclude with an important lemma from formal concept analysis, saying that the pieces of information in a context are not only partially ordered, but also form a complete lattice. Its proof can be found in (Davey & Priestley, 2002).

Lemma 12.13 *$\langle \mathcal{I}(\mathcal{M}, \mathcal{L}, \models); \leq \rangle$ is a complete lattice in which join and meet are given by*

$$\begin{aligned} \bigvee_{j \in J} (V_j, S_j) &= \Big(C_{\models}(\bigcup_{j \in J} V_j), \bigcap_{j \in J} S_j \Big), \\ \bigwedge_{j \in J} (V_j, S_j) &= \Big(\bigcap_{j \in J} V_j, C^{\models}(\bigcup_{j \in J} S_j) \Big). \end{aligned}$$

12.5 Comparison: Contexts, Information Algebras and Information Flow

In Sections 12.1 to 12.3, Barwise and Seligman's theory of information flow has been presented. This theory is based on classifications. Infomorphisms describe the flow of information between two classifications. An information channel is established by a well-defined collection of classifications and the infomorphisms between them. As already indicated by its name, the theory of information flow concentrates on what happens *between* classifications in such a channel. In the foregoing Section 12.4, contexts (and thus classifications, as their is no conceptual difference between them) were said to be domains to express information. Looked at from this point of view, Barwise and Seligman's theory of information flow formalizes the transport of information between different domains. What we have really missed in their theory is a formal definition of the nature of information. Our unanswered questions are: What is information? If information flows between classifications, where is it located in such a classification? What are its properties? The answers to these questions are provided by formal concept analysis, as presented in Section 12.4. In formal concept analysis, a context is fixed and its concepts, which we have renamed "pieces of information", are examined. So we obtained the following results about the nature of information: Information is always to be considered within a context and there is usually more than one piece of information in a context. A piece of information is given by a pair of sets (V, S) being subsets of the sets $\mathcal{M}$ and $\mathcal{L}$, which constitute the context. Both sets of the pair, defining a piece of information, have to be $\models$-closed. The pieces of information in a context are partially ordered and form a complete lattice. Formal concept analysis describes the structure and the rules inside a classification, which we have missed in Barwise and Seligman's theory. However, formal concept analysis is limited to the inner life of a context. It does not consider anything outside a context, i. e. a context's relation to other contexts or how some information in this context may be expressed by means of another one. Loosely speaking, formal concept analysis tells us what happens *inside* a context and Barwise and Seligman's theory of information flow is about what happens *between* contexts. Putting both point of views together gives a detailed description of what information is and how to transport it from one context to another. It turns out that pieces of information in a context form an information algebra and that infomorphisms give rise to the transport operation introduced in Section 4.4.

This section is divided into two parts: First, in Section 12.5.1, we will go on with the subject of Section 12.4 and show that there is a domain-free information algebra associated with a context. In the second part, Section 12.5.2, Barwise and Seligman's theory comes into play. We explain why in our opinion, contexts, as proposed by formal concept analysts, are more convenient for information representation than Barwise and Seligman's classifications. Then, we consider more than one context and point out the infomorphisms between them. Finally, we establish the link between channels and the information algebra transport operation.

12.5.1 Contexts and Information Algebras

In order to show that the pieces of information in $\mathcal{I}(\mathcal{M}, \mathcal{L}, \models)$, i. e. all possible pieces of information in a context, form an information algebra, some requirements have to be met. We will now define two operations, the combination of information and the focusing of information. Furthermore, a domain structure is provided. This paves the way for the proof that there is a domain-free information algebra associated with a context. As contexts are dual structures, one can either operate on the level of valuations or on the level of sentences. For the proof, we will consider combination and focusing for valuations. Thereafter, the case of sentences is looked at and the corresponding entities are stated.

Combination

One of the most basic operations on pieces of information is their combination. We will provide the combination operation for two pieces of information by using the valuation sets of both. Consider two pieces of information (V, S) and (V', S'). As they are constituted of $\models$-closed sets, we have $V = \hat{r}(S), S = \check{r}(V)$ and $V' = \hat{r}(S'), S' = \check{r}(V')$. Their combination $(V, S) \otimes (V', S')$ is defined by the intersection of their sets V and V' of valuations:

Definition 12.14 (Combination)
The combination *of two pieces of information (V, S) and (V', S') is defined by*

$$(V, S) \otimes (V', S') \quad := \quad (V \cap V', \check{r}(V \cap V')).$$

Now that combination has been defined, we have to add a domain structure and a focusing operation in order to approach information algebras. This is done using so-called similarity model structures.

Similarity Model Structures

Similarity model structures have been introduced in (Mengin & Wilson, 1999). The authors use similarity model structures for showing how to embed different kinds of logics in what they call the local computation framework. This framework consists only of those three axioms of information algebras which are necessary for local computation (see Section 4.5). In (Mengin & Wilson, 2001), a more detailed insight is given, on which the following is based. Note, however, that the authors are dealing with the labeled version of information algebras, in the sense used in Section 4.2.

Definition 12.15 (Similarity Model Structure)

A similarity model structure *is a triple* $(M, \omega, (\equiv_x)_{x \subseteq \omega})$*, where*

- M *is a set whose elements are called models,*
- $\omega = \{1, 2, \ldots\}$ *indexes sets* x *of variables,*
- *each* $\equiv_x$ *is an equivalence relation on* M*.*

The models in M are typically tuples, ascribing values to some set of variables. ω provides indices for specifying subsets of variables. For every set of variables determined by $x \subseteq \omega$, an equivalence relation is given. It regroups those models of M which have the same values for the variables specified by x.

Example 12.5.1 (Similarity Model Structure) Consider a similarity model structure with $M = \{(0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1),\ (1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)\}$, so the models are sequences of values out of $\{0,1\}$ and there are three variables taking values. The first value of a model is assigned to the variable called p_1, the second and the third to the variables called p_2 and p_3, respectively. So ω indexes sets of $\mathfrak{P}(\{p_1, p_2, p_3\})$. In order to identify equivalent models, not only the variables have to be specified, but also their values have to be fixed. As we are in a Boolean situation, there are $|\mathfrak{P}(\{p_1, p_2, p_3\})| \times 2 = 16$ possible equivalence relations. Let some subset $x \subseteq \omega$ designate the set $\{p_2\}$, where p_2 is set to 0. The corresponding equivalence class regroups the following models of M: $\{(0,0,0), (0,0,1), (1,0,0), (1,0,1)\}$. As another example take $y \subseteq \omega$, indexing the set $\{p_2, p_3\}$, with p_2 set to 0 and p_3 set to 1. Then, the corresponding equivalence class is $\{(0,0,1), (1,0,1)\}$. $\odot$

As we are aiming to show that the pieces of information of a context $(\mathcal{M}, \mathcal{L}, \models)$ form an information algebra, we will use the following similarity model structure: The set M containing the models will be the set $\mathcal{M}$ of valuations of the context. A simple indexing set will not be enough, more structure has to be provided for describing domains, so we consider a lattice D. Finally, the equivalence relation regroups those valuations of $\mathcal{M}$ which are equivalent relative to an element of D, i. e. which have the same values for the variables in $x \in D$. So we will consider in the following the similarity model structure $(\mathcal{M}, D, (\equiv_x)_{x \in D})$.

Monotonicity Property

After having introduced similarity model structures, Mengin and Wilson bring in the very important monotonicity property which is to be assumed for all equivalence relations $(\equiv_x)_{x \in D}$:

Property 12.16 (Monotonicity Property)

For $x, y \in D$ it holds that

$$\text{if } x \leq y \text{ then } \equiv_x \supseteq \equiv_y .$$

The monotonicity property says that given two valuations $v, v' \in \mathcal{M}$ and $x \leq y$, we know that if $v \equiv_y v'$, then also $v \equiv_x v'$. This is a very natural property, as models which are equivalent relative to some set of variables y will remain equivalent if a subset x of the original set y is considered.

Example 12.5.2 (Monotonicity) We will illustrate the monotonicity property using the above Example 12.5.1. Clearly $x \leq y$, and we see that $\equiv_x \supseteq \equiv_y$, as $\{(0,0,0), (0,0,1), (1,0,0), (1,0,1)\} \supseteq \{(0,0,1), (1,0,1)\}$. $\odot$

Independence Property

A further property is required for the set of equivalence relations $(\equiv_x)_{x \in D}$ of a similarity model structure $(\mathcal{M}, D, (\equiv_x)_{x \in D})$:

Property 12.17 (Independence Property)

Let $v, v' \in \mathcal{M}$ and $x, y \in D$. If for two valuations v, v' it holds that $v \equiv_{x \wedge y} v'$, then there exists a valuation v'' such that $v'' \equiv_x v$ and $v'' \equiv_y v'$.

This property is best paraphrased in (Mengin & Wilson, 2001) as

> "knowing the $\equiv_x$-equivalence class A of an unknown model v'' doesn't tell us anything about its $\equiv_y$-equivalence class B (except that B and A are both subsets of the same $\equiv_{x \wedge y}$-equivalence class i. e., that containing v'')."

Now, we already dispose of a domain structure. But similarity model structures are also necessary for introducing the focusing operation.

Focusing

The other basic operation besides combination is focusing. Focusing allows to extract information by considering it only relative to some set of variables $x \in D$, which is of specific interest. As for combination, we define focusing on sets of valuations. This requires to state how to focus a single valuation. So we define for $v \in \mathcal{M}$ and $x \in D$

$$v^{\Rightarrow x} \quad := \quad \{v' \in \mathcal{M} : v' \equiv_x v\}. \tag{12.8}$$

Focusing a single valuation to some set of variables $x \in D$ results in the set of those valuations which are equivalent to this valuation regarding x, i. e. which have the same values as v has for all the variables in x. However, pieces of information always consist of a set of valuations, which is a subset of $\mathcal{M}$, and of the corresponding set of sentences. Therefore, the above focusing operation is now extended to sets in the following way:

Definition 12.18 (Focusing)

Focusing of a piece of information (V, S) on some $x \in D$ is defined on the level of valuations by

$$\begin{aligned} (V,S)^{\Rightarrow x} &:= (V^{\Rightarrow x}, \check{r}(V^{\Rightarrow x})), \text{ where} \\ V^{\Rightarrow x} &= \bigcup_{v \in V} v^{\Rightarrow x}. \end{aligned}$$

Focusing a piece of information in a context is defined by focusing its set of valuations. The sets of valuations which do not change when they are focused on some $x \in D$ are of special interest.

Definition 12.19 (Cylindric Set)

*Let $x \in D$. A set $V \subseteq \mathcal{M}$ of valuations is called x-*closed *or* cylindric over *x, if $V = V^{\Rightarrow x}$.*

Before going on, we will illustrate focusing by an example:

Example 12.5.3 (Focusing) Reconsider the scenario from Example 12.5.1, where three Boolean variables are given. This gives rise to the eight valuations $\mathcal{M}$ consists of. For focusing the set $V = \{(0,0,0), (0,0,1)\}$, $V \subseteq \mathcal{M}$, to $x \subseteq \omega$ (x is designating the set $\{p_2\}$, where p_2 is set to 0), we need to determine $(0,0,0)^{\Rightarrow x}$ and $(0,0,1)^{\Rightarrow x}$ and take the union of the results:

$$\begin{aligned} (0,0,0)^{\Rightarrow x} &= \{(0,0,0),(0,0,1),(1,0,0),(1,0,1)\} \\ (0,0,1)^{\Rightarrow x} &= \{(0,0,0),(0,0,1),(1,0,0),(1,0,1)\} \\ \{(0,0,0),(0,0,1)\}^{\Rightarrow x} &= \{(0,0,0),(0,0,1),(1,0,0),(1,0,1)\} \end{aligned}$$

$\odot$

Even if we are now equipped with the operations of combination and focusing, as well as with a domain structure, there is still something missing, before the information algebra axioms can be verified. We need two more assumptions on sets of valuations.

Closure Property

Not only monotonicity (Property 12.16) and independence (Property 12.17) are important prerequisites for the verification of the information algebra axioms. A closure property has to be imposed on sets of valuations. It relates to the focusing operation and presuppose the similarity model structure $(\mathcal{M}, D, (\equiv_x)_{x \in D})$.

Property 12.20 (Closure Property)

Let $V \subseteq \mathcal{M}$ and $x \in D$. If $V = C_{\models}(V)$, then $V^{\Rightarrow x} = C_{\models}(V^{\Rightarrow x})$.

So for all sets V of valuations, we assume that if V is $\models$-closed, then $V^{\Rightarrow x}$ is $\models$-closed, too. Thus the family of $\models$-closed sets is closed under the focusing operation $\Rightarrow$.

The three Properties 12.16, 12.17 and 12.20 (monotonicity, independence and closure) are needed for the following verification of the information algebra axioms, especially for the transitivity axiom, as we shall see below.

Verification of the Axioms

The verification of the axioms will only be done for sets of valuations. The corresponding set of sentences is omitted, but it can easily be retrieved by means of $\check{r}$. Sets of valuations will be denoted by Greek lower case letters like ϕ and ψ.

D is a lattice of finite subsets of variables. The set Ψ of all pieces of information is the set of all cylindric sets of valuations which are $\models$-closed:

$$\Psi \quad := \quad \{\psi \in \mathcal{M} : \psi = C_{\models}(\psi), \psi = \psi^{\Rightarrow x}, \text{ for some } x \in D\}. \tag{12.9}$$

By $\models$-closeness it holds for all $\psi \in \Psi$ that $\psi = C_{\models}(\psi)$ and so the set ψ of valuations determines a piece of information in a context, as $\check{r}(\psi)$ is the corresponding ($\models$-closed!) set of sentences. Each $\psi \in \Psi$ is cylindric over some set $x \in D$. According to Definition 12.14, the combination of $\phi, \psi \in \Psi$ is

$$\phi \otimes \psi \quad := \quad \phi \cap \psi. \tag{12.10}$$

From Definition 12.18 it is known that focusing a piece of information ψ on some domain $x \in D$ is done by

$$\psi^{\Rightarrow x} \quad := \quad \bigcup_{v \in \psi} v^{\Rightarrow x}. \tag{12.11}$$

Now we are ready to state that pieces of information in a context (represented by their sets of valuations) form an information algebra.

Theorem 12.21 *The two-sorted algebra (Ψ, D), where Ψ is defined by Equation 12.9 and D is a lattice of finite subsets of variables, together with the operations of combination $\otimes$ and focusing $\Rightarrow$, as given by Equations 12.10 and 12.11 above, is a domain-free information algebra if the following properties are satisfied: monotonicity (Property 12.16), closure (Property 12.20) and independence (Property 12.17).*

Proof.

1. *Semigroup-Axiom*:
   Combination is realized by the intersection of sets of valuations. As the intersection operation is commutative and associative, the semigroup axiom holds for combination. $\mathcal{M}$ is the neutral element and $C_{\models}(\emptyset)$ the null element. For the proof that the intersection of two cylindric sets still results in a cylindric set, is proved formally in the same way as the first point of the proof of Theorem 8.15.

2. *Transitivity-Axiom:* $(\psi^{\Rightarrow x})^{\Rightarrow y} = \psi^{\Rightarrow x \wedge y}$*, for* $\psi \in \Psi$ *and* $x, y \in D$*.*
   ψ is cylindric, so $\exists z \in D$ such that $\psi = \psi^{\Rightarrow z}$. Each valuation $v \in \psi$ is thus also contained in $\psi^{\Rightarrow z}$, according to the definition of focusing, $\psi^{\Rightarrow z} = \bigcup_{v \in \psi} \{v' \in \mathcal{M} : v' \equiv_z v\}$.
   Let $a \in D$ and as D is a lattice, it holds that $a \wedge z \leq z$. The monotonicity property now states: If $v \equiv_z v'$, then also $v \equiv_{a \wedge z} v'$. Above (definition of focusing) we have seen that $v \equiv_z v'$, so $v \equiv_{a \wedge z} v'$ actually holds. By the independence property we know that there exists a valuation $t \in \mathcal{M}$ such that $t \equiv_a v$.

   Instead of proving $(\psi^{\Rightarrow x})^{\Rightarrow y} = \psi^{\Rightarrow x \wedge y}$, we will prove that $\forall v \in \psi$, it holds that $(v^{\Rightarrow x})^{\Rightarrow y} = v^{\Rightarrow x \wedge y}$.

   - $\leftarrow$ We have seen above that $\forall v \in \psi$, $\exists t \in \mathcal{M}$, such that $t \equiv_a v$. Take $a = x \wedge y$, so the proof is done for $t \equiv_{x \wedge y} v$: By definition $t^{\Rightarrow x \wedge y} = \{u \in \mathcal{M} : u \equiv_{x \wedge y} t\}$. For all such $u \equiv_{x \wedge y} t$, there exists by the independence property a valuation $w \in \mathcal{M}$ such that

     $$\left.\begin{array}{lll} t \equiv_x w & \text{i.e.} & w \in t^{\Rightarrow x}, \\ w \equiv_y u & \text{i.e.} & u \in w^{\Rightarrow y}, \end{array}\right\} u \in (t^{\Rightarrow x})^{\Rightarrow y}.$$

     All $u \in t^{\Rightarrow x \wedge y}$ are thus also in $(t^{\Rightarrow x})^{\Rightarrow y}$.
   - $\rightarrow$ By the same reasoning as above, $(t^{\Rightarrow x})^{\Rightarrow y} = \bigcup_{w \in t^{\Rightarrow x}} w^{\Rightarrow y}$, where $w^{\Rightarrow y} = \{u \in \mathcal{M} : u \equiv_y w\}$. This states that $w \equiv_x t$ and $w \equiv_y u$. By the fact that $x \wedge y \leq x$, $x \wedge y \leq y$ and by the monotonicity property, $t \equiv_{x \wedge y} w \equiv_{x \wedge y} u$ holds. All $u \in (t^{\Rightarrow x})^{\Rightarrow y}$ are thus also in $t^{\Rightarrow x \wedge y}$.

3. *Combination-Axiom:* $(\phi^{\Rightarrow x} \otimes \psi)^{\Rightarrow x} = \phi^{\Rightarrow x} \otimes \psi^{\Rightarrow x}$*, for* $\phi, \psi \in \Psi$ *and* $x \in D$*.*
   This axiom is proven by applying the definitions of focusing and combination:

   $$\begin{aligned} (\phi^{\Rightarrow x} \otimes \psi)^{\Rightarrow x} &= \bigcup_{v \in (\phi^{\Rightarrow x} \otimes \psi)} v^{\Rightarrow x} \\ &= \bigcup_{v \in (\phi^{\Rightarrow x} \cap \psi)} v^{\Rightarrow x} \\ &= \left(\bigcup_{v \in \phi^{\Rightarrow x}} v^{\Rightarrow x} \right) \cap \left(\bigcup_{v \in \psi} v^{\Rightarrow x} \right) \\ &= \phi^{\Rightarrow x} \otimes \psi^{\Rightarrow x} \end{aligned}$$

4. *Idempotency-Axiom:* $\psi \otimes \psi^{\Rightarrow x} = \psi$, *for* $\psi \in \Psi$ *and* $x \in D$.
   By the definition of focusing it holds that $\psi^{\Rightarrow x} \supseteq \psi$.

$$
\begin{aligned}
\psi \otimes \psi^{\Rightarrow x} &= \psi \cap \underbrace{\psi^{\Rightarrow x}}_{\supseteq \psi} \\
&= \psi
\end{aligned}
$$

5. *Support-Axiom:* $\forall \phi \in \Psi\ \exists x \in D$ *such that* $\psi^{\Rightarrow x} = \psi$.
   This axioms holds since all the elements of Ψ are cylindric (see above).

Hence, (Ψ, D) is a domain-free information algebra. □

Finding New Instances

The foregoing proof has a great impact. By having proved that a domain-free information algebra can be associated with a context, contexts become a formalism inducing information algebra instances. A context $(\mathcal{M}, \mathcal{L}, \models)$ is a general framework for formalisms whose representation (the sentences in $\mathcal{L}$) and whose semantics (the valuations in $\mathcal{M}$) can be linked by a binary relation $\models$. We have already given examples for formalisms being contexts, such as propositional or predicate logic, linear equations, powersets, etc. Each of these examples provides pieces of information in a context, according to Definition 12.11, i. e. the mappings $\hat{r}$ and $\check{r}$ exist. As they fulfill the monotonicity property, the closure property and the independence property, the cylindric sets of $\mathcal{M}$, which are $\models$-closed, form a domain-free information algebra, by Theorem 12.21. This procedure allows to find new information algebra instances, induced by contexts.

From Valuations to Sentences

We have seen above that there is an information algebra associated with a context. The operations have been formulated using valuations. A context being a dual structure, there is is also a formulation with sentences. We will now show how to go from the valuations to the sentences. A context $(\mathcal{M}, \mathcal{L}, \models)$ allows to switch from $\mathcal{M}$ to $\mathcal{L}$ using the mapping $\check{r}$, see Equation 12.4. For a given piece of information, specified by a set $V \subseteq \mathcal{M}$ of valuations, $\check{r}(V)$ contains all the sentences that describe V, the theory of V. As V is part of a pair defining a piece of information, it is $\models$-closed and so is $S = \check{r}(V)$.

In Definition 12.14, the combination of two pieces of information (V, S) and (V', S') in a context is given in terms of valuations. Carrying forward Equation 12.10 results in

$$V \otimes V' = V \cap V' = \hat{r}(S) \cap \hat{r}(S') = \hat{r}(S \cup S'), \tag{12.12}$$

using $\models$-closeness and (P4) of Lemma 12.8. We are aiming for a formulation by sentences, so we will trace them back to their corresponding valuations, using $\hat{r}$. In the following, the above Equation 12.12, $V = \hat{r}(S)$ and the definition of the closure operator for sentences (Equation 12.6) are used:

$$\begin{aligned} S \otimes S' &= \check{r}(\hat{r}(S) \otimes \hat{r}(S')) \\ &= \check{r}(V \otimes V') \\ &= \check{r}(\hat{r}(S \cup S')) \\ &= C^{\models}(S \cup S') \end{aligned}$$

The focusing operation for sentences can also be reduced to Definition 12.18, using the mapping $\hat{r}$. So if we dispose of a piece of information (V, S) in a context, $\hat{r}(S) = V$ holds. Furthermore, we know how to focus V. By the mapping $\check{r}$, the result of the focusing is transformed to sentences again:

$$S^{\Rightarrow x} \quad := \quad \check{r}(\hat{r}(S)^{\Rightarrow x})$$

Now that it has been shown that a domain-free information algebra is associated with a context and that the operations of combination and focusing may either be expressed in terms of valuations or in terms of sentences, Barwise and Seligman's theory of information flow will be looked at in the light of the results just obtained.

12.5.2 Information Flow and Information Algebras

In Barwise and Seligman's theory of information flow, classifications are a core entity. With contexts, a very similar formalism, providing, however, more structure, has been proposed. We will explain why contexts are preferable to classifications for information representation. Thereafter, we take up Barwise and Seligman's idea of infomorphisms, but this time information is transported between contexts (not between classifications). Until now, we have only looked at a single context. For moving information between contexts, more than one context has to be considered. We show how to derive from a general context more specific contexts. This leads directly to Barwise and Seligman's notion of a channel, which turns out to give rise to the transport operation of information algebras.

Classifications, Contexts and Pieces of Information

Naming Clearly, there is no difference (besides naming) between the definition of a classification (Definition 12.1) and that of a context (Definition 12.7). Both are triples, composed of two sets and a binary relation between the two sets. The naming correspondences are shown in the table given in Figure 12.9.

Mappings The first difference appears when mappings are defined between the two constituting sets. In the classification case, the mappings *typ* and *tok* (Equations 12.1 and 12.2) apply to a single token or type, respectively. Applying *typ* to a token t results in the set of types which classify t. *tok* maps a type θ to the set of tokens

classification	**context**
$C = (tok(C), typ(C), \models_C)$	$C = (\mathcal{M}, \mathcal{L}, \models)$
token $t \in tok(C)$	valuation $v \in \mathcal{M}$
type $\theta \in typ(C)$	sentence $s \in \mathcal{L}$
$t \models_C \theta$: t is classified as being of type θ	$v \models s$: v satisfies s

Figure 12.9: Naming correspondences for classifications and contexts

θ classifies. In the context case, Equations 12.3 and 12.4 define the mappings $\hat{r}$ and $\check{r}$, which apply to sets of sentences and valuations, respectively. So $\hat{r}$ maps a set of sentences to a set of valuations, and applying $\check{r}$ to a set of valuations results in a set of sentences. By defining their mappings on sets, contexts offer the possibility to map the result of the mapping $\hat{r}$ back to sentences, using $\check{r}$ (or the other way round with valuations, first applying $\check{r}$ and then $\hat{r}$). Such a switching between types and tokens is not provided by classifications, so Barwise and Seligman had no need to extend their *typ* and *tok* mappings to sets, as they probably did not see the impact of this extension.

Closure In (Davey & Priestley, 2002), the mappings $\hat{r}$ and $\check{r}$ are also called polar maps. It is shown that they set up a Galois connection and the authors state that (P1)-(P3) of Lemma 12.8 are instances of properties which hold for any Galois connection. As one can associate with every context a Galois connection $(\hat{r}, \check{r})$, closure operators, as the ones defined in Equations 12.5 and 12.6, can be derived and the fact that $\mathcal{I}(\mathcal{M}, \mathcal{L}, \models)$ is a complete lattice (Lemma 12.13) can be shown. See (Davey & Priestley, 2002) for more details.

From our point of view, the closure operator is indispensable for a formal description of the nature of information. If one admits that information can either be given by a syntactic representation (sentences) or by its meaning (the semantics, expressed by valuations), intuitively, one wants to switch between both forms, always getting the same information, irrespective of the mode of representation. It is thus very natural to define pieces of information to be composed of two closed sets, one for the syntactic representation and one for its meaning. Closed means in this case that if one takes some set of either mode of representation, retrieves the corresponding set in the other mode and then transforms the result back to the original mode, one obtains again the set started with – a "round trip" in Figure 12.8, where the starting set equals the arrival set.

A further reason for information to be closed is that a statement always implies certain consequences. These consequences are captured by the closure of the statement. As the consequences are also part of the information uttered, information has to be closed.

There are lots of examples where the closure is needed. The most common is perhaps propositional logic, where the set $\mathcal{M}$ of valuations is given by $\{0,1\}^{|P|}$, the set $\mathcal{L}$

of sentences by $\mathcal{L}_P$ and the binary relation by the entailment relation $\models$. A set of models of $\{0,1\}^{|P|}$ is naturally closed. If we start with some set of models, determine the corresponding formulae and then retrieve their models, we will come back in any case to exactly the same set as we started with. This does not apply to formulae. When starting with some set of formulae, identifying their models and then determining the corresponding theory, the resulting set of formulae is not necessarily the original set, but a superset (in case the original set was not closed). The set obtained at the end contains all the consequences of the set started with. For propositional logic we thus have $S \subseteq C^{\models}(S)$, but $V = C_{\models}(V)$. Two examples will illustrate what happens if $\hat{r}$ and $\check{r}$ are applied to sets which are not closed. The first example deals with propositional logic, giving an illustration of the problem just depicted, the second one is about planets, continuing Example 12.1.3.

Example 12.5.4

1. Consider the propositional logic setting as described above and let us start with the formula $p_1 \wedge p_2$. The corresponding set of models is $\hat{r}(\{p_1 \wedge p_2\}) = \{(1,1)\}$. Applying $\check{r}$ results in the theory $\check{r}(\{(1,1)\}) = \{p_1, p_2, p_1 \vee p_2, p_1 \wedge p_2\}$, which is the logical consequence of $p_1 \wedge p_2$.
2. The set of properties $\{$*small, distance-near, moon-yes*$\}$ was obtained by $\check{r}(\{$*Earth*$\})$. Asking for the planets which have these properties results in a superset of $\{$*Earth*$\}$, namely in $\hat{r}(\{$*small, distance-near, moon-yes*$\}) = \{$*Earth, Mars*$\}$.

$\odot$

Note that the same problems arise for infomorphisms and thus for channels. Transporting back and forth some information might result in a modification of the original information. Examples 12.2.1 and 12.3.1 will only work out even, if closed sets are considered. Interestingly, Barwise and Seligman never look at information that flows back into the classification it stems from. It is most likely that they did not think that such a situation could emerge from their theory of information flow and this is why they did not treat it. Otherwise, they would have seen without much doubt that an extension of the mappings *typ* and *tok* sets up a Galois connection, which entails a lot of very useful structure inside the classification. In our opinion, Barwise and Seligman were just caring about what happens between classifications and did not pay attention to its inner life. Presumably, this is also the reason why a formal definition of information cannot be found in the theory of information flow. We are missing all these useful consequences arising out of the Galois connection, such as the closure operators, the formal definition of pieces of information and their order.[5]

A further element we have been looking for in vain in both theories (Barwise and Seligman's theory of information flow and formal concept analysis) is something that

[5]See (Brunet, 2000) for a translation of Barwise and Seligman's theory into a lattice and Galois connections based formalism. It is shown that this extends the expressive power of their theory.

corresponds to the domain structure D of an information algebra. Such a domain structure is necessary when the focusing operation comes into play. Focusing is not considered in formal concept analysis. In Barwise and Seligman channels, something similar to focusing is hidden, as we will see later on.

Contexts and Infomorphisms

Until now, contexts have only been looked at separately. A single context $(\mathcal{M}, \mathcal{L}, \models)$ has been considered and its properties have been examined. However, the focusing operation (Definition 12.18) allows to take into account several, more specific contexts. Such a context relates to a certain $x \in D$ and is thus called x-context.

Definition 12.22 (x-Context)

An x-context $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ is a context defined for $x \in D$, where

$$\begin{array}{rcl} \mathcal{M}_x & = & \{v^{\Rightarrow x} : v \in \mathcal{M}\} \\ \mathcal{L}_x & = & \{s \in \mathcal{L} : \hat{r}(\{s\}) = (\hat{r}(\{s\}))^{\Rightarrow x}\} \\ \models_x \subseteq \mathcal{M}_x \times \mathcal{L}_x & : & v^{\Rightarrow x} \models_x s \textit{ if } v \models s \textit{ for all } v \in v^{\Rightarrow x} \end{array}$$

The mappings $\hat{r}_x$ and $\check{r}_x$ are defined for $B \subseteq \mathcal{L}_x$ and $A \subseteq \mathcal{M}_x$ by means of $\models_x$:

$$\hat{r}_x(B) = \{a \in \mathcal{M}_x : a \models_x b, \forall b \in B\}. \quad (12.13)$$

$$\check{r}_x(A) = \{b \in \mathcal{L}_x : a \models_x b, \forall a \in A\}. \quad (12.14)$$

Obviously, the x-context $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ can be deduced from the "domain-free" context $(\mathcal{M}, \mathcal{L}, \models)$. The first component of an x-context is the set $\mathcal{M}_x$ of valuations consisting of elements $v^{\Rightarrow x}$ which are valuations of $\mathcal{M}$, focused on x. As pieces of information are composed of closed sets, we are interested in the subsets of $\mathcal{M}_x$ which are $\models_x$-closed. For that purpose, note that the definition of focusing a set $V \in \mathcal{M}$ of valuations on $x \in D$ may also be written as $V^{\Rightarrow x} = \{v^{\Rightarrow x} : v \in V\}$. $V^{\Rightarrow x}$ is thus a subset of $\mathcal{M}_x$. By the closure property (Property 12.20), we know that $V^{\Rightarrow x}$ is $\models$-closed, if V itself is $\models$-closed. But $V^{\Rightarrow x}$ being $\models$-closed means that $V^{\Rightarrow x} = C_{\models}(V^{\Rightarrow x})$, which defines $\models_x$-closeness. So if V is $\models$-closed in $(\mathcal{M}, \mathcal{L}, \models)$, then $V^{\Rightarrow x}$ is $\models_x$-closed in $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$. A $\models_x$-closed set of valuations can be used to generate a piece of information in an x-context, as it determines a $\models_x$-closed set of sentences, by property (P3) of Lemma 12.8. The second component of an x-context is the set $\mathcal{L}_x$ of sentences. It contains the sentences of $\mathcal{L}$ satisfied by a set of valuations which is cylindric over x, so $\mathcal{L}_x \subseteq \mathcal{L}$. Finally, the third component, the relation $\models_x$ between $\mathcal{M}_x$ and $\mathcal{L}_x$, is defined as follows: $v^{\Rightarrow x}$, which is a set of valuations $v \in \mathcal{M}$, satisfies some sentence $s \in \mathcal{L}_x$, which is at the same time a sentence of $\mathcal{L}$, if s is satisfied by every v which is in $v^{\Rightarrow x}$.

But why is it interesting to introduce x-contexts and what is the difference between an x-context $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ and a general, "domain-free" context $(\mathcal{M}, \mathcal{L}, \models)$? These questions may be best answered by looking at the valuations. Contexts are a framework for specifying pieces of information. A piece of information in a context is either given by a set of sentences (syntax) or by a set of valuations, representing the meaning of the information. As the elements of $\mathcal{M}_x$ are valuations of $\mathcal{M}$, focused to onto x, the idea behind an x-context is obvious. $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ provides only information relative to $x \in D$. This may be interesting if only certain aspects of the available information matter. These aspects are captured by $x \in D$, so the information is focused onto x. An x-context is therefore used to provide customized information for some specific purpose. The "domain-free" context $(\mathcal{M}, \mathcal{L}, \models)$ is not restricted to any specific field of interest and supplies the available information in all aspects.

Having introduced contexts relative to some element of the lattice D allows us to establish context morphisms between them, which are very similar to infomorphisms (Definition 12.3). Consider two elements $x, y \in D$, such that $x \leq y$. The contexts relative to these elements of D are $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ and $(\mathcal{M}_y, \mathcal{L}_y, \models_y)$, which will be called x-context and y-context, respectively. Every valuation $v^{\Rightarrow x}$ of $\mathcal{M}_x$ is also in $\mathcal{M}_y$, as $x \subseteq y$ and $(v^{\Rightarrow x})^{\Rightarrow y} = v^{\Rightarrow x \cap y} = v^{\Rightarrow x}$. Thus, $\mathcal{M}_x \supseteq \mathcal{M}_y$. It follows from (P2) of Lemma 12.8 that $\mathcal{L}_x \subseteq \mathcal{L}_y$. We define now a contravariant pair of mappings:

$$\begin{array}{rccccl} f: & \mathcal{L}_x & \rightarrow & \mathcal{L}_y & & \\ & \mathcal{M}_x & \leftarrow & \mathcal{M}_y & : g \end{array}$$

As $x \leq y$, every sentence of $\mathcal{L}_x$ is also a sentence of $\mathcal{L}_y$. Therefore, the mapping f is simply defined as

$$f(s) = s. \tag{12.15}$$

The mapping g is the projection from $\mathcal{M}_y$ onto $\mathcal{M}_x$. Only valuations relative to $x \leq y$ are taken into account:

$$g(v^{\Rightarrow y}) = v^{\Rightarrow x}. \tag{12.16}$$

With exactly the same reasoning as the one given in Example 12.2.1, it can be shown that the pair of mappings $\langle f, g\rangle$ satisfies the following condition:

$$g(v^{\Rightarrow y}) \models_x s \quad \Leftrightarrow \quad v^{\Rightarrow y} \models_y f(s). \tag{12.17}$$

Definition 12.23 (Context Morphism)

A contravariant pair of mappings $c = \langle f, g\rangle$ between two contexts $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ and $(\mathcal{M}_y, \mathcal{L}_y, \models_y)$, where

- $x \leq y$,
- $f : \mathcal{L}_x \rightarrow \mathcal{L}_y$, *defined as* $f(s) = s$ *and*
- $g : \mathcal{M}_y \rightarrow \mathcal{M}_x$, *defined as* $g(v^{\Rightarrow y}) = v^{\Rightarrow x}$,

satisfying Equation 12.17, is called a context morphism.

$$
\begin{tikzcd}
\mathcal{L}_x \arrow[r, "f"] \arrow[d, dash, "\models_x"] & \mathcal{L}_y \arrow[d, dash, "\models_y"] \\
\mathcal{M}_x & \mathcal{M}_y \arrow[l, "g"]
\end{tikzcd}
$$

Figure 12.10: Context morphism between $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$ and $(\mathcal{M}_y, \mathcal{L}_y, \models_y)$

A context morphism between an x-context and a y-context allows to express information, which is originally relative to x, by means of information relative to y, or vice versa. This is illustrated by Figure 12.10.

Definition 12.12 establishes a partial order between pieces of information. As $\mathcal{M}_x \supseteq \mathcal{M}_y$ and $\mathcal{L}_x \subseteq \mathcal{L}_y$, the pieces of information in $(\mathcal{M}_y, \mathcal{L}_y, \models_y)$ are finer than those in $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$. So we may say that context $(\mathcal{M}_y, \mathcal{L}_y, \models_y)$ is finer than context $(\mathcal{M}_x, \mathcal{L}_x, \models_x)$, or that the context generated by $x \in D$ is coarser than the one generated by $y \in D$. This is reflected in the fact that $x \leq y$.

Context morphisms correspond in essence to infomorphisms, as given in Section 12.2. The difference between the definitions is that context morphisms are established between contexts and that the sets x, y generating the contexts have to obey an underlying order relation; x has to be a subset of y. Infomorphisms, as introduced by Barwise and Seligman, act between classifications. The classifications may be arbitrary, there is no such order condition as for context morphisms. Context morphisms are thus a little bit more restrictive than infomorphisms, but more structure is supplied, due to the underlying order.

From Lemma 12.4 it is known that the composition of two infomorphisms results again in an infomorphism. The same holds for context morphisms, with the restriction that the first context morphism in the chain has to be established between an x-context and a y-context, where $x \leq y$. The second context morphism linked to the first one has to be between the former y-context and a z-context, where $y \leq z$. This results in a context morphism between the x-context and the z-context. For the proof, consider the one of Lemma 12.4 of the chaining of the more general infomorphisms and the fact that $x \leq z$.

Channel and Transportation

The idea behind context morphisms is already known as an operation applied to a piece of information in an information algebra. If context morphisms are used in a situation as the one described by Barwise and Seligman as a channel, it corresponds to the transport operation of labeled information algebras, in the sense of Section 4.4.

A channel $\mathcal{C}$, as given by Definition 12.5, consist of a set of classifications C_j which are all connected by infomorphisms $i_j = \langle f_j, g_j \rangle$ to a special classification, called the core of $\mathcal{C}$. Barwise and Seligman say about this core classification that it is "the common codomain" of the infomorphisms i_j. Actually, the set of types of the core classification is the common codomain of all f_j.

Figure 12.5 illustrates a channel situation. How can such a channel be established using context morphisms? According to Definition 12.23, context morphisms require that one of the contexts is finer than the other one, in the sense discussed above. So consider an indexing set J and some elements x_j of the lattice D, $j \in J$, giving rise to x_j-contexts $C_{x_j} = (\mathcal{M}_{x_j}, \mathcal{L}_{x_j}, \models_{x_j})$. If a set of such contexts $(C_{x_j})_{x_j \in D}$ is given, the core context has to be found. Its pieces of information are finer than all the pieces of information provided by the contexts in $(C_{x_j})_{x_j \in D}$. The coarsest context, which is finer than all C_{x_j}, is the context induced by $\bigvee\{x_j : x_j \in D, j \in J\}$, the join of all x_j under consideration. Note that only if D is a complete lattice, $\bigvee X$ exists for all possible sets X of elements of the lattice D.

Example 12.5.5 (Channel using Context Morphisms) Consider two contexts $C_x = (\mathcal{M}_x, \mathcal{L}_x, \models_x)$ and $C_y = (\mathcal{M}_y, \mathcal{L}_y, \models_y)$, generated by two arbitrary elements $x, y \in D$. In order to establish a channel, their core context has to be found. It is $C_{x \vee y} = (\mathcal{M}_{x \vee y}, \mathcal{L}_{x \vee y}, \models_{x \vee y})$, as $x, y \leq x \vee y$. Then, two context morphisms $c_1 = \langle f_1, g_1 \rangle$, between C_x and $C_{x \vee y}$, and $c_2 = \langle f_2, g_2 \rangle$, between C_y and $C_{x \vee y}$ are added. The context morphism c_1 is given by $f_1(s) = s$ and $g_1(v^{\Rightarrow x \vee y}) = v^{\Rightarrow x}$. In the same manner, context morphism c_2 consists of $f_2(s) = s$ and $g_2(v^{\Rightarrow x \vee y}) = v^{\Rightarrow y}$. This sets up a channel connecting C_x and C_y, as shown in Figure 12.11.

$$\begin{tikzcd}
 & C_{x \vee y} \arrow[dl, shift left, "g_1"] \arrow[dr, shift right, "g_2"'] & \\
C_x \arrow[ur, shift left, "f_1"] & & C_y \arrow[ul, shift right, "f_2"']
\end{tikzcd}$$

Figure 12.11: Channel connecting C_x and C_y using context morphisms

$\odot$

Even if the the notion of coarser or finer pieces of information does not occur in Barwise and Seligman's theory, nor that of an order between classifications, the Definition 12.6 goes into this direction. It says that one channel is a refinement of another, if the channels share the same component classifications and there is an infomorphism between the channels' cores. This implies that there is no unique core for a channel connecting a set of classifications. Several cores are possible and there are infomorphism between them, called "refinement infomorphisms" by Barwise and

Seligman. So if there exist such a refinement infomorphism (see Figure 12.7, where channel $\mathcal{C}'$ is finer than $\mathcal{C}$), one of the classifications must be finer than another, and therefore having finer pieces of information. As there are several possible cores for a channel, and the core is the common codomain of the involved infomorphisms, there may be a smallest one, even if Barwise and Seligman do not tell anything about it. In the case of context morphisms, the join of all x_j under consideration generates the coarsest core context which is finer then all the contexts involved, as an underlying lattice is considered.

There is a further interesting observation one can make when looking at channels. It might be best seen graphically, e. g. in Figure 12.6. Due to the fact that an infomorphism $\langle f, g\rangle$ is a contravariant pair of mappings, we have the following setting in a channel: f is a function from a component classification to the core, applying to types, and g is a function from the core to a component classification, applying to tokens. So moving a piece of information from one component classification to another is done by using the corresponding mapping f to send a set of types to the core. The only possibility to send this information to another component classification is to use the appropriate mapping g, which applies to sets of tokens! Thus one has to switch from types to tokens in the core of a channel. In a channel using context morphism for moving a piece of information from one context to another, this is no problem at all. Pieces of information are closed in such a setting and switching from sentences to valuations or vice versa does not change the information. However, in Barwise and Seligman's version of a channel, using classifications and infomorphisms, the switch in the core from types to tokens could change the information transmitted, see Example 12.5.4.

We have mentioned above that something similar to the operation of focusing is hidden in Barwise and Seligman's description of channels. In Section 12.3, the tokens of the core classification of a channel were called connections. A passage from (Barwise & Seligman, 1997) was cited saying that these connections make it possible that the components of the channel carry information about each other. Information can only flow in a channel if the information relative to one of the component classifications has been expressed by means of the core classification, using the f-mapping of the corresponding component classification. For obtaining the information it carries about some other component classification, the g-mapping of the target component classification is applied. Having information algebras and contexts in mind, it is obvious to link this flow of information in channels to the transport operation, as introduced in Section 4.4. The definition of the transport operation (Definition 4.12) indicates that f and g of the above description correspond to the vacuous extension (in the sense of Definition 4.11) and the marginalization operation (see Definition 4.3). The information flow in a channel corresponds to a piece of information with domain $x \in D$ (it is thus a piece of information in an x-context) which is to be transported to domain y. For that purpose, it is first vacuously extended to the domain $x \vee y$, in order to count as a piece of information in an $x \vee y$-context. This might be seen as the mapping f from the component classification to the core. The second step in transporting a piece of information from domain x to domain y is marginalizing the piece of information, which is relative $x \vee y$, to the domain y, so that it counts as information in an y-context. This corresponds to the mapping g from the core to the target component classification. So Barwise and Seligman implicitly use focusing and its advantages, without naming it and perhaps without knowing the power of this operation. Note, however, the difference concerning the core. In the case of the transport operation, the core is determined by the join of the origin and the target domain, giving rise the coarsest context which is finer than both contexts involved. The transport operation is the same as a channel established between two contexts, connected by context morphisms. Context morphisms have an additional requirement that sets them apart from infomorphisms, namely that the codomain of the mapping f has to be finer than its domain. The join guarantees this condition, it is furthermore the coarsest one, so no unnecessary computation has to be done. The core is termed "common codomain" in Barwise and Seligman's definition of a channel and is not uniquely defined, as seen above. No further details about the nature of this common codomain are given.

However, the main difference between the channel in Barwise and Seligman's theory and the transport operation in information algebras is conceptual, namely in the nature of the element that Barwise and Seligman call core. In their theory of information flow, there is one central entity (the core), generating a system in which information flows. The core is analyzed by determining the component classifications, which are somehow artificial or custom-made, as they stem from the partition of the the whole entity into parts, creating an information system. The parts may be spatial, temporal or something else. Barwise and Seligman literally say that

> "the conception of information flow we develop is very broad, [...] We place no restriction in what kind of thing may count as a part, only that the choice of parts determines the way in which we understand what it is for information to flow from one part to another."

Recall their example of the light bulb, already introduced in Section 12.3. The division into parts is depicted by Figure 12.12.

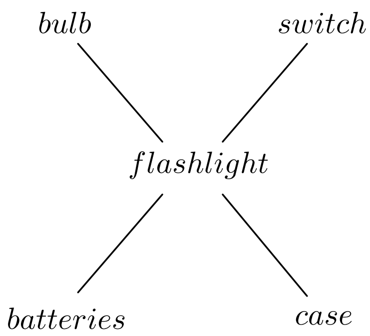


Figure 12.12: Example of a channel established by a flashlight

The flashlight (core) has been divided into four parts, the component classifications. An example of information flow in this system is: The light bulb carries the information that the switch is on and the batteries are charged. So the core is given, and depending on what information shall flow in the system, the component classifications are determined. Depending on the field of interest, there will be different or further component classifications. An engineer designing flashlights will perhaps need more scientific classifications as the ones given above. This is why the component classifications were called artificial or customized above.

The transport operation in information algebras gives also rise to a channel, but the respective roles are different: The core is determined by the component classifications, and not vice versa. Given some piece of information with domain $x \in D$, we use the transport operation for determining what kind of information relative to another domain $y \in D$ is provided by this piece of information. The component classifications of Barwise and Seligman's channel correspond in this setting to an x- and a y-context each providing a piece of information: the piece of information with domain x (given) and the same piece of information, but this time transported to y. These two domains x and y determine, by means of the underlying lattice, the context which is the core of the channel. Obviously, in the case of transport, the contexts corresponding to the component classifications are given and a core has to be found. This is the opposite situation of what happens when modeling a distributed system by a channel, in the sense of Barwise and Seligman. However, looking at the involved context morphisms or infomorphisms shows that the flow of information is the same in both situations.

If one looks at Equation 4.11, another possibility of transporting information is given, that time using the meet, and not the join of the involved domains. Transporting a piece of information in an x-context to a y-context may be equally done by going via the $x \wedge y$-context. This situation is depicted by Figure 12.13. A piece of information with domain x is marginalized to the domain $x \wedge y$, in order to count as a piece of information in an $x \wedge y$-context. Thereafter, it is vacuously extended to the domain y, so that its information relative to this domain becomes obvious. As before, marginalization is realized by the mapping g of the context morphism and vacuous extension by the mapping f. In such a setting, we switch from valuations to sentences in $C_{x \wedge y}$.

$$
\begin{tikzcd}
C_x \arrow[dr, shift left, "g_1"] & & C_y \arrow[dl, shift left, "g_2"'] \\
& C_{x \wedge y} \arrow[ul, shift left, "f_1"] \arrow[ur, shift left, "f_2"'] &
\end{tikzcd}
$$

Figure 12.13: Channel connecting C_x and C_y, going via $C_{x \wedge y}$

Such a scenario of a channel cannot be found in the theory of information flow. Barwise and Seligman do not admit information to flow from one component classification of the channel to another one by going via a classification which is a component itself and not the core, determined by the common codomain. We consider contexts to come along with a lattice structure, so that a context forms an information algebra. Context morphisms also act relative to this lattice structure. The correctness of the scenario depends also strongly on the information algebra axioms. There are thus two ways to establish a channel with context morphisms, either using the join or the meet.

Obviously, the properties of Ψ and D as well as the information algebra axioms have a great impact, as they allow information to be ordered and thereby bring a lot of structure into the theory. Furthermore, the axioms state or induce important properties of the combination and focusing operations. This is indispensable for doing automatic information processing with computers and permits at the same time to use local computation algorithms (see Section 4.5) for any information algebra instance. Barwise and Seligman do not dispose of such structuring elements in their theory, but we have shown above that their theory fits into our information algebra framework, as it is possible to obtain a context from every classification.

12.6 Conclusion

On the one hand, Barwise and Seligman present a theory describing how information flows in a distributed system. This system is modeled by classifications, which are connected by infomorphisms. The flow of information is formalized by a channel, constituted of a core classification and some component classifications which are connected to the core via infomorphisms. Barwise and Seligman show what happens *between* classifications in order that information can flow. On the other hand, formal concept analysts propose a formalism, called context, to represent information, which is quasi equivalent to classifications. In formal concept analysis, one looks at what happens *inside* a context, providing a formal definition of the nature of information. It turns out that the information algebra framework joins both approaches together, so that an overall image of information arises, including its properties and the operations one can apply to it.

Part IV

Conclusion

13
Making One out of Three? – A Unifying Approach

Theory of Information Transmission is not generally spoken of as dealing with the content of the messages transmitted, and hence it is regarded as a discipline having no connection with semantics.

Edward Colin Cherry (1914 - 1979)
A History of the Theory of Information

We presented in the third part of this thesis three semantic theories of information:

Chapter 10: Carnap and Bar-Hillel's theory of semantic information from the early 1950s is the oldest of the three theories. A very basic framework for semantic information and its measure is given. A semantic piece of information is perceived as a set of excluded possibilities. The theory is instantiated by a restricted monadic predicate logic language.

Chapter 11: Groenendijk and Stokhof's theory of the semantics of questions and the pragmatics of answers provides a framework for questions. It has been developed in the early 1980s and has been extended by van Rooij in the first decade of the 21st century. As questions are identified with their possible answers, a detailed description of the nature of answers (the information relative to a specific question) is also given. Two instances, propositional and predicate logic, are considered.

Chapter 12: Barwise and Seligman's theory of information flow came up in the late 1990s. It provides a framework for the representation and the transport of information in distributed systems. As to the representation, information is seen from a dual perspective, taking into account syntax and semantics. In this thesis, it has been identified with the context approach of formal concept analysis, which was introduced in the 1980s. Barwise and Seligman mainly exemplify their theory by predicate logic.

At the end of each of the three previous chapters, a comparison between the presented theory and the algebraic theory of semantic information was drawn. The elements of the respective theory were identified with their matches in the information algebra framework. The three theories presented above could be described by means of the information algebra vocabulary. None of the three theories entirely corresponds to the algebraic theory of semantic information, each of them covers only a subset of the elements provided by the information algebra framework.

This thesis could also have been written by starting with the presentation of the three theories of semantic information (part III of this thesis). Then the information algebra framework (part I of this thesis) could have been introduced as a generalization, followed by the propositional and predicate logic instances of part II. The algebraic theory of semantic information can be seen as an abstract theory, covering all three theories of Chapters 10, 11 and 12. However, each of the three presented theories puts emphasis on another aspect of the information algebra framework; but there are also information algebra characteristics, which are shared by all three of them.

In this chapter, we will summarize the main points of our algebraic theory of semantic information (Section 13.1) as a unifying approach of the three foregoing chapters. Then we will look at which properties of our theory can be found in all of the three others (Section 13.2). Thereafter, it is shown in which sense the information algebra framework is a generalization of each theory (Section 13.3). Finally, a concluding comparison, relating and confronting the theories with each other, is given in Section 13.4.

13.1 Algebraic Theory of Semantic Information

The algebraic theory of semantic information is based on the abstract, axiomatic information algebra framework from (Kohlas, 2003). An information algebra is a two-sorted algebra, so it has two main components: a set Ψ of pieces of information and a lattice D of questions. In the information algebra framework, there are two operations defined: combination, denoted by $\otimes$, and focusing, denoted by $\Rightarrow$. A set of five axioms is imposed on the structure (Φ, D) and the operations $\otimes$ and $\Rightarrow$. Here, we have chosen domain-free information algebras, as they are best used for theoretical concerns. The axioms imposed on such an information algebra can be found in Section 5.2.

There are many instances of this abstract information algebra framework. In order to count as an information algebra instance, a formalism has to provide definitions for $\Phi, D, \otimes$ and $\Rightarrow$. Furthermore, the five axioms have to be fulfilled. Two formalisms, propositional and predicate logic, have been proven to be information algebras in Chapters 8 and 9. But there are many others, like data base systems, systems of linear equations or systems of linear inequalities.

Our algebraic theory of semantic information gives answers to the following three questions:

1. What is information from the semantic point of view?
2. How is information processed?
3. How is information measured?

We will now briefly explain each of the three above points. In doing so, the content of the first part of this thesis is summarized.

13.1.1 What Is Information From the Semantic Perspective?

Ψ is the set of pieces of information.

- Information usually comes piecewise.

Furthermore, the lattice D, which is associated to Ψ, provides domains for the pieces of information $\psi \in \Psi$. A domain is seen as the question to which the piece of information ψ is an answer.

- Information is always relative to at least one question and is semantically perceived as an answer to that question.

Looking at the semantics of a piece of information as an answer means treating it as a set of possibilities. There are two possible interpretations for this set of possibilities: a disjunctive and a conjunctive way. In the first case, when a piece of information is interpreted disjunctively, it acts as a scheme of choice, describing a situation of uncertainty. One of the possibilities the piece of information makes available has initially been chosen, but it is not known which is the selected possibility. The piece of information answers a question by providing a set of possibilities, containing the selected element. In the second case, when a piece of information is interpreted conjunctively, it acts as an enumeration of possibilities. Such a piece of information provides everything that is placed at the disposal in a specific situation.

- A semantic piece of information is a set of possibilities, which may either be interpreted disjunctively or conjunctively.

Semantically speaking, every piece of information is a set. So it must be possible to express an information algebra as a system of subsets, which capture the meaning of the pieces of information in Ψ. This is done in (Kohlas, 2003, Chapter 6.3), where it is shown that every information algebra can be embedded in such a subset system.

- The semantic way of looking at information is universal.

13.1.2 How Is Information Processed?

Information does not only come piecewise, but usually also from different sources. Therefore, different pieces of information have to be aggregated in order to get the overall view and to make sure that the whole information available is taken into account. This is done by the information algebra operation of combination, denoted by $\otimes$.

- A basic operation used for information processing (aggregation) is the combination of pieces of information.

We have seen above that information should be seen as the answer to a question. Obviously, a piece of information can be rich of content relative to some question, but empty of content relative to another question. The information algebra operation of focusing, denoted by $\Rightarrow$, allows to extract the part of information relative to a specific question.

- Another basic operation used for information processing (extraction) is the focusing of a piece of information on a question.

Important properties of these two operations are given in the information algebra axioms. They also enable local computation and therefore, information algebras are generic structures for inference.

13.1.3 How Is Information Measured?

As already pointed out above, the content of a piece of information will vary from one question to another. So it is very essential to measure information always relative to a certain question. Furthermore, the content of a piece of information depends also on what is already known. It is thus measured relative to prior information, which may be vacuous.

- The content of a piece of information is to be measured relative to a question and relative to prior information.

There are two possible ways of comparing the content of a piece of information: qualitatively and quantitatively. Information is idempotent under combination, i. e., when a question is answered by a piece of information which was already known, nothing new is learned. This allows to establish a partial order of information content on the set of pieces of information.

- Information is quantitatively compared by means of a partial order, which arises from the idempotency property.

Information is quantitatively measured by the reduction of uncertainty, when getting an answer to a question. Without probabilistic considerations, the approach of (Hartley, 1928) is applied. But when attention is given to the probability of a piece of information to occur, information is measured by entropy, going back to (Shannon, 1948).

- Information is quantitatively measured by the reduction of uncertainty, going back to Hartley and Shannon.

13.2 Common Properties

The algebraic theory of semantic information and the three semantic information theories, proposed in Chapters 10, 11 and 12, have two basic properties in common. The first one concerns the nature of the theories (Section 13.2.1), the other one involves the perception of information (Section 13.2.2). As to Chapter 12, we are here only considering Barwise and Seligman's theory of information flow, as introduced in Sections 12.1 to 12.3, not taking into account our context extension of it (Section 12.4).

13.2.1 Theoretical Framework

In the previous section, we pointed out once again that the algebraic theory of semantic information is based on the abstract, axiomatic information algebra framework. Formalisms which fit into this framework are called instances. So there is a theoretical background, on the one hand, and formalisms which exemplify the theory, on the other hand.

Even if the three semantic information theories are mainly applied to logics, the difference between theoretical framework and instances is made in all three theories. This difference is most obvious in Carnap and Bar-Hillel's theory of semantic information. The authors really point out that there is a framework ("presystematic concept") and an instance ("explicatum"). On the contrary, Groenendijk and Stokhof originally only provided a theoretical framework, when coming up with their theory of the semantics of questions. In their later publications, they apply their theory to logics, from which the difference between the framework and the instances partially suffers. The theory is only applied to the examples, but not explained independent of them. But considering their whole work allows to make a distinction between the theoretical background and the formalisms, used as examples. In Barwise and Seligman's theory of information flow, the different parts of the theory are spontaneously exemplified by formalisms, which play the role of instances of the theory. They are simply named example, and a clear distinction between the theory and the examples is drawn.

13.2.2 Semantic Information

From Section 13.1.1 it is known that we perceive a semantic piece of information as a set of possibilities. This approach is shared by the authors of the three semantic information theories: Bar-Hillel and Carnap explicitly require in their framework that semantic information is to be treated as a set. Groenendijk and Stokhof postulate that a semantic answer (i. e. a piece of information) is a set of possible worlds. Barwise and Seligman, finally, talk of sets of types and sets of tokens.

13.3 The Information Algebra Framework as Unifying Approach

Based on the concise summary of the algebraic theory of semantic information, given in Section 13.1, we will now show in what sense the information algebra framework is a generalization of the three semantic information theories proposed in the third part of this thesis. Although none of the three theories is based on an algebraic structure with a set of axioms, it is possible to look at how information is handled. We know that the algebraic theory of semantic information provides answers to the three following questions from Section 13.1:

1. What is information from the semantic perspective?
2. How is information processed?
3. How is information measured?

It turns out that the three theories also give answers to (some of) these questions. The answers differ, depending on the selected theory, but every theory can be sketched by its answers to these questions.

13.3.1 Carnap and Bar-Hillel

In the case of Bar-Hillel and Carnap's theory of semantic information, it is probably most obvious that our algebraic theory of semantic information is a generalization of their theory. A detailed comparison of both approaches has already been provided in Sections 10.5 and 10.6. An information algebra counterpart could be found for every part of Carnap and Bar-Hillel's theory[1], but not vice versa. We will now have a look at how Bar-Hillel and Carnap's theory of semantic information answers the above three questions.

[1] Except for the information measure *cont*, which, however, never became widely accepted and is not of any importance in the later works of Bar-Hillel and Carnap.

What Is Information From the Semantic Perspective?

Two statements about the nature of semantic information can be found: First, information comes piecewise. Second, a semantic piece of information is a set. When Carnap and Bar-Hillel speak of the information carried by some statement or sentence i, they are dealing with a piece of information. As mentioned above, the authors require that semantic information is treated as a set. The most characteristic feature of Carnap and Bar-Hillel's framework for information is that a piece of information is defined by what it excludes. This is the dual point of view of our perception of information, but we have seen that it fits into the information algebra framework.

Questions do not occur in Carnap and Bar-Hillel's theory. Concerning the perception of semantic information, this is the main point where the algebraic theory of semantic information really goes beyond Bar-Hillel and Carnap's theory of semantic information.

How Is Information Processed?

Information processing is not of any importance in Carnap and Bar-Hillel's theory of semantic information. No operations for information processing are introduced. However, combination can be identified with logical conjunction, semantically interpreted by set intersection. This operation is taken for granted, which is certainly due to the authors' logical background, where such an operation is part of the basic equipment.

As questions are not part of Bar-Hillel and Carnap's theory, we are missing an operation similar to focusing.

How Is Information Measured?

Carnap and Bar-Hillel's theory of semantic information gives three answers to this question:

- The content of a piece of information is to be measured relative to prior information.
- Information can be measured qualitatively, which leads to a partial order.
- Information can be measured quantitatively, following Hartley's approach.

Here, again, the answers are very close to the ones of the algebraic theory of semantic information, which is more general in the third point, as Shannon's measure is an extension of Hartley's approach. As before, the concept of questions is missing, so the fact that the information content will vary with respect to different questions is not taken into account.

Conclusion

The theory of semantic information, proposed by Carnap and Bar-Hillel, is very close to the algebraic theory of semantic information. The latter is more general, as it provides

- an algebraic structure of information with two operations and an axiomatic framework,
- the concept of questions, as well as everything which is related to this concept, especially the focusing operation, and
- a wider range of instances (Bar-Hillel and Carnap's theory of semantic information only applies to logic).

13.3.2 Groenendijk, Stokhof and van Rooij

First of all, note that Groenendijk and Stokhof have developed a theory of question-answering, giving a semantic description of the act of asking a question and afterwards, obtaining an answer to that question. As the purpose of an answer is to convey information, we will now equate pieces of information with the answers of Groenendijk and Stokhof's theory of the semantics of questions. This allows us to find out how their theory answers the three questions on the nature of information, cited at the beginning of this section.

What Is Information From the Semantic Perspective?

Groenendijk and Stokhof's theory of the semantics of questions allows for the following statements:

- Information is always an answer to a question.
- Information is given by a set of possible worlds.

Obviously, Groenendijk and Stokhof's perception of a single semantic piece information is the same as in the algebraic theory of semantic information. But we are missing the idea of information coming piecewise. Furthermore, their theory does not provided a disjunctive and a conjunctive interpretation of information.

How Is Information Processed?

Information is not really processed in Groenendijk and Stokhof's theory. The authors' interest lies in the act of asking a question. It is nevertheless possible to identify two operations, which can be taken as combination and focusing. These operations are found in one of their further publications, which is, however, not clearly related to their work on the semantics of questions. The operations only apply to predicate logic and are not a fundamental component of Groenendijk and Stokhof's theory.

In the work of van Rooij, which concentrates on measuring information and questions, prior information is taken into account. This concept is put into practice by the intersection of sets. The idea of focusing a piece of information does not occur.

How Is Information Measured?

As a piece of information is an answer to a question, Groenendijk, Stokhof and van Rooij always measure the content of a piece of information relative to a question.

As a qualitative measure of information, Groenendijk and Stokhof propose to order pieces of information by the order relation $\ll_q$, of which our pre-order $\leq_q$ is a special case. As in Groenendijk and Stokhof's theory of the semantics of question, there is no need to compare all possible answers, regardless of their corresponding questions, there is no partial order between all pieces of information.

Van Rooij proposes to measure information quantitatively, following Hartley's approach. Furthermore, he applies Shannon's concept of entropy to measure the reduction of uncertainty, due to getting an answer to a question. In this point, the same approach as in the algebraic theory of semantic information is chosen.

The theory of the semantics of questions proposes a qualitative measure of information, which is, though, not a partial order as in the algebraic theory of semantic information. Based on the ideas of Hartley and Shannon, the information content of a piece of information (an answer to a question) is measured. Groenendijk, Stokhof and van Rooij measure information always relative to a question. When information is measured quantitatively, prior information is also taken into account. So their approach of measuring information is very similar to, but not equal to the one proposed in the algebraic theory of semantic information.

Conclusion

The theory of the semantics of questions, proposed by Groenendijk and Stokhof and extended by van Rooij, corresponds nearly one-to-one to that part of the algebraic theory of semantic information which concerns questions.[2] Our theory is however more general, as it

- deals not only with questions, but also with information in general, and
- provides an algebraic structure of information and questions, which comes along with operations involving both entities and with an axiomatic framework.

[2] This part is inadequately covered by our three questions on the nature of semantic information, so please refer to Section 11.7.

Furthermore, note that there is a wider range of instances, as Groenendijk, Stokhof and van Rooij's theory of the semantics of questions is only applied to logics.

13.3.3 Barwise, Seligman and Formal Concept Analysis

It is important to distinguish between Barwise and Seligman's theory of information flow, as introduced in Sections 12.1 to 12.3, and our context extension of it (Section 12.4). If simply the theory of information flow, without our extension, is taken into account, only the second question, "How is information processed?", can be answered. In order to answer also the first and the third question, formal concept analysis has to be brought in. Therefore, we will now look at how Barwise and Seligman's extended theory of information flow (Sections 12.1 to 12.4) answers the three questions on the nature of information.

What Is Information From the Semantic Perspective?

The following statements about semantic information can be made:

- Information is always considered in a context.
- A piece of information is a pair of sets. One of these sets (the set of valuations) expresses the semantics of the piece of information.

The context of a piece of information acts as the question it refers to. The pieces of information in a context are all possible answers to this question. As contexts correspond to Barwise and Seligman's classifications, one might say that the theory of information flow only provides the questions, but not the answers. In their theory of information flow Barwise and Seligman do not state what information is. Nevertheless, they talk of sets of tokens. These sets of tokens correspond to the sets of valuations which constitute a piece of information from the semantic point of view. However, it is not stated how to interpret this set (disjunctively or conjunctively), neither in the theory of information flow, nor in formal concept analysis.

How Is Information Processed?

Barwise and Seligman describe the flow of information by means of infomorphisms between classifications, which leads to the definition of a channel. From Section 12.5.2 it is known that the information algebra counterpart of an infomorphism is the transport operation. In Barwise and Seligman's theory of information flow, information is therefore processed by the transport operation, which is only defined for labeled information algebras. It corresponds to the focusing operation of domain-free information algebras. However, neither in formal concept analysis, nor in the theory of information flow, an operation similar to combination could be found, even if it is possible to define it, as seen in Section 12.5.1.

How Is Information Measured?

In Chapter 12, only a qualitative measure of information is proposed, which goes back to formal concept analysis. Barwise and Seligman do not measure information at all. The partial order of pieces of information in a context establishes a qualitative measure of information content. As only the pieces of information in a content are compared by this partial order, and not the pieces of information of different contexts, this qualitative measure is relative to a question.

Conclusion

In Section 12.5.1 it is shown how the information algebra framework and Barwise and Seligman's extended theory of information flow are related. Provided that certain properties (concerning the similarity model structure and the focusing operation) hold, an information algebra can be established from the extended theory of information flow. The algebraic theory of semantic information can be seen as a generalization of Barwise and Seligman's theory of information flow, extended with the ideas of formal concept analysis, since we are missing

- an interpretation of a semantic piece of information,
- the combination operation and
- a quantitative measure of information.

13.4 Concluding Comparison

The goal of this section is *not* to find out, which of the three semantic information theories, presented in the third part of this thesis, is the closest to the algebraic theory of semantic information. It is rather a question of getting a concluding, overall picture of what we have seen. For this purpose, we have established a table, shown in Figure 13.1, which confronts the three theories with the algebraic theory of semantic information, in a condensed form. The content of this table is based on Sections 13.1 and 13.3. The properties of the algebraic theory of semantic information are given in the column entitled "ATSI". Carnap and Bar-Hillel's theory of semantic information is listed in the column entitled "CBH". Groenendijk and Stokhof's theory of the semantics of questions, which has been extended by van Rooij, is depicted in the column entitled "GSvR". Finally, Barwise and Seligman's theory of information flow, extended with the ideas of formal concept analysis, is presented in the rightmost column, entitled "BS+". A cross $\times$ in a cell of the table means that the theory provides the characteristics listed on the left hand side. A tilde $\sim$ expresses that something similar to that property can be found in the theory.

		ATSI	CBH	GSvR	BS+
semantic information	piecewise	$\times$	$\times$		$\times$
	relativity to question	$\times$		$\times$	$\times$
	set	$\times$	$\times$	$\times$	$\times$
	disj. / conj. interpretation	$\times$	$\sim$		
processing	combination $\otimes$	$\times$	$\times$	$\times$	
	focusing $\Rightarrow$	$\times$			$\times$
measure	relative to question	$\times$		$\times$	$\times$
	relative to prior information	$\times$	$\times$	$\times$	
	qualitative (partial order)	$\times$	$\times$	$\sim$	$\times$
	quantitative	$\times$	$\sim$	$\times$	
	set of all pieces of information Ψ	$\times$			$\sim$
	lattice of questions D	$\times$		$\sim$	

Figure 13.1: Overall comparison of semantic information theories

It strikes that none of the three theories of the third part of this thesis globally looks at all possible pieces of information.[3] As there is no real counterpart for the set Ψ of all possible pieces of information, none of the three theories meets the requirements for an algebraic structure. Interestingly, none of the three theories considers the operation of focusing together with (a lattice of) questions. Barwise and Seligman transport information from one classification (context) to another, but they do not tell how the classifications are related. A side effect of Groenendijk and Stokhof's theory leads to the partition lattice of questions, but they do not consider sublattices of $Part(U)$, as it is the the case for D. They do neither provide an operation which allows to extract from an answer information relative to another question. In the algebraic theory of semantic information, a piece of information can be focused on different, but related questions. This approach is really new and cannot be found in other (semantic) theories of information. Even it cannot be seen from the table, it is worth mentioning that none of the three theories looks at the idempotent nature of information. The information algebra framework postulates that information is idempotent under combination, which is in our point of view one of the basic principles of information. Furthermore note that only the algebraic theory of semantic information provides so many different instances, whereas the three theories, presented in the third part of this thesis, mainly apply to logics.

So the algebraic theory of semantic information, which is based on the information algebra framework, may well be called a unifying approach to information expressed by logic.

[3]Barwise and Seligman's extended theory of information flow provides all possible pieces of information for one context, i. e. relative to one question, but all possible contexts (leading to all possible pieces of information) are not naturally considered.

14

Synopsis and Discussion

Finally, in conclusion, let me say just this...

Pink Panther

A synopsis and a discussion of this thesis are given in this chapter. In Section 14.1, the information theoretical aspects of this thesis are reconsidered. Thereafter, in Section 14.2, the results relating to logics are discussed. Finally, in Section 14.3, open questions for future research are briefly summarized.

14.1 Information Theoretical Aspects

The information algebra framework has been presented in this thesis for the first time as an information theory by itself, independent of its computational origins. In doing so, the information algebra framework has been considered from a semantic point of view. It turned out that information has two sides. On the one hand, the representation of information by a formal language can be looked at. On the other hand, the meaning of information can be examined, and this is what we did in this thesis. This leads to a semantic information theory, providing a definition of the concept of information.

Information usually comes piecewise and is always perceived as an answer to a question. A semantic piece of information is a set of possibilities, which may either be interpreted disjunctively or conjunctively.

All pieces of information, which are possible in some situation, usually count as answers to different questions. A question is given by its possible answers. Questions were shown to be related and to form a lattice.

Two operations of information processing are defined on the set of possible pieces of information and the lattice of questions. One is combination, used for information aggregation, the other one is focusing, applied in order to extract information relative to some question.

These properties of information and questions are formalized by an algebraic structure, consisting of the set of all possible pieces of information, the lattice of questions, the operations of combination and focusing and of a set of five axioms, which describe the properties of information.

The content of a piece of information is to be measured relative to a question and relative to prior information. Information is quantitatively measured by means of the partial order in the set of all pieces of information. Quantitatively, information is measured by the reduction of uncertainty, following Shannon's approach.

These results constitute the algebraic theory of semantic information, which applies to many instances, including logics.

Furthermore, it has been shown that the algebraic theory of semantic information is very general. It unifies three semantic theories of information and of questions, respectively. The three theories, which have been formally presented and linked to information algebras, came up in the 1950s (Carnap and Bar-Hillel), in the 1980s (Groenendijk and Stokhof) and the 1990s (Barwise and Seligman). They all relate to logics, since they have been developed by logicians, philosophers, mathematicians and linguists. As it could be shown that the algebraic theory of semantic information is a generalization of these three theories, we can justifiably speak of it as a unifying approach to semantic information theory.

Moreover, this thesis can be seen as a validation of the information algebra framework, as the information theory arising from it complies with already established theories. It is even a further development in the field of semantic information theories, as, for the first time, it is pointed out that a piece of information can be focused on different, but related questions.

14.2 Considerations Related to Logics

It has formally be shown in this thesis that propositional logic and predicate logic fit into the information algebra framework.

Propositional logic was already known to be an information algebra instance before, but our proof involves the semantic side of propositional logic, which has not been considered so far.

Predicate logic as an information algebra instance has been brought to discussion for the first time on the semantic and on the syntactic level, considering the links between them. Since we are interested in semantics, we have chosen many-sorted predicate logic, reflecting the characteristics of variables in a natural way. It turned out that in predicate logic, the information is not only localized in the set of models of a formula, but also in the structure.

14.3 Future Work

The following research areas were not covered in this thesis and could be analysed in future work:

- *Uncertain information:*
  The wide field of uncertain information has not been investigated in this thesis. Starting from (Kohlas & Eichenberger, 2009), it would be interesting to relate the algebraic theory of semantic information to uncertainty, which can be modeled by random variables taking values in information algebras. This would also give rise to further measures of information.

- *New instances:*
  The proof of Theorem 12.21 allows to find new information algebra instances, induced by contexts. So further research on the basis of (Mengin & Wilson, 2001) and related to the field of formal concept analysis may be worthwhile.

- *Relational data bases:*
  The process of query answering in relational data bases is also interesting. Based on (Schneuwly, 2007), it may be possible to give a precise, information theoretical analysis of the localization of information.

- *Deductive data bases:*
  Deductive data bases, which are not only dealing with pieces of information, but also with inference rules between them, could also be incorporated in the proposed theory.

- *Natural language:*
  Based on natural language interfaces to data bases, the links of the proposed theory to natural language could be examined.

References

Bar-Hillel, Y. 1952. Semantic Information and Its Measures. *Pages pp.33–48 of:* Foundation, Josiah Macy Jr. (ed), *Transactions of the Tenth Conference on Cybernetics.*

Bar-Hillel, Y. 1955. An Examination of Information Theory. *Philosophy of Science*, **22**(2), 86–105.

Bar-Hillel, Y. 1964. *Language and Information.* Addison-Wesley.

Bar-Hillel, Y., & Carnap, R. 1952. *An Outline of a Theory of Semantic Information.* Tech. rept. 247. Research Laboratory of Electronics, Massachusetts Institute of Technology.

Bar-Hillel, Y., & Carnap, R. 1953. Semantic Information. *The British Journal for the Philosophy of Science*, **4**(14), 147–157.

Barwise, J., & Seligman, J. 1997. *Information Flow: The Logic of Distributed Systems.* Cambridge Tracts in Theoretical Computer Science, no. 44. Cambridge U.K.: Cambridge University Press.

Brunet, O. 2000. Lattice Approach to Classifications. *Pages 34–44 of: In European summer school logic, language and information, Birmimgham (UK).*

Carnap, R. 1950. *Logical Foundations of Probability.* University of Chicago Press.

Carnap, R. 1952. *The Continuum of Inductive Methods.* University of Chicago Press.

Davey, B.A., & Priestley, H.A. 2002. *Introduction to Lattices and Order.* second edn. Cambridge University Press.

Fine, T. L. 1970. *Theories of probability: an examination of foundations.* New York, Academic Press.

Grätzer, G. 2003. *General Lattice Theory.* Birkhäuser Verlag.

Groenendijk, J. 1999. The Logic of Interrogation (Classical Version). *In:* Matthews, T., & Strolovitch, D.L. (eds), *Proceedings of the Ninth Conference on Semantic and Linguistic Theory.*

Groenendijk, J. 2003. Questions and answers: Semantics and logic. *Pages 12–23 of:* Bernardi, R., & Moortgat, M. (eds), *Proceedings of the 2nd CologNET-ElsET Symposium. Questions and Answers: Theoretical and Applied Perspectives.* OTS.

Groenendijk, J., & Stokhof, M. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers.* Ph.D. thesis, Universiteit van Amsterdam.

Groenendijk, J., & Stokhof, M. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy*, **14**(1), 39–100.

Groenendijk, J., & Stokhof, M. 1997. Questions. *Chap. 19, pages 1055–1124 of:* ter Meulen, Alice, & van Benthem, Johan (eds), *Handbook of Logic and Language.* Elsevier Science Publishers.

Hacking, I. 2001. *An Introduction to Probability and Inductive Logic.* Cambridge University Press.

Haenni, R. 2004. Ordered Valuation Algebras: a Generic Framework for Approximating Inference. *International Journal of Approximate Reasoning*, **37**(1), 1–41.

Haenni, R., Kohlas, J., & Lehmann, N. 2000. Probabilistic Argumentation Systems. *Pages 221–287 of:* Kohlas, J., & Moral, S. (eds), *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 5: Algorithms for Uncertainty and Defeasible Reasoning.* Kluwer, Dordrecht.

Hájek, A. Winter 2008. Interpretations of Probability. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/archives/win2008/entries/probability-interpret/.

Halmos, P. R. 1962. *Algebraic Logic.* New York: Chelsea.

Halmos, P. R. 1974. *Naive Set Theory.* Springer.

Hamblin, C. L. 1958. Questions. *Australasian Journal of Philosophy*, **36**(3), 159–168.

Hamblin, C. L. 1973. Questions in Montague English. *Foundations of Language*, **10**(1), 41–53.

Hartley, R.V.L. 1928. Transmission of Information. *Bell System Technical Journal*, July, 535–563.

Hawthorne, J. Winter 2008. Inductive Logic. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/archives/win2008/entries/logic-inductive/.

Henkin, L., Monk, J. D., & Tarski, A. 1971. *Cylindric Algebras.* Amsterdam: North-Holland.

Hintikka, J. 1974. Questions about Questions. *Pages 103–158 of:* Munitz, M.K., & Unger, P.K. (eds), *Semantics and Philosophy.* New York University Press.

Hintikka, J. 1976. The Semantics of Questions and the Questions of Semantics. *Acta Philosophica Fennica*, **28**(4).

Hintikka, J. 1978. Answers to Questions. *Pages 279–300 of:* Hiz, H. (ed), *Questions.* D. Reidel Publishing Company.

Jensen, F.V., Lauritzen, S.L., & Olesen, K.G. 1990. Bayesian Updating in Causal Probabilistic Networks by Local Computation. *Computational Statistics Quarterly*, **4**, 269–282.

Kohlas, J. 2002. *Information in Context.* Tech. rept. 02–15. Department of Informatics, University of Fribourg.

Kohlas, J. 2003. *Information Algebras: Generic Structures for Inference.* Discrete Mathematics and Theoretical Computer Science. London, Berlin, Heidelberg: Springer-Verlag.

Kohlas, J., & Eichenberger, Ch. 2009. Uncertain Information. *In:* Sommaruga, G. (ed), *Formal Theories of Information.* Springer Verlag.

Kohlas, J., & Monney, P.A. 1995. *A Mathematical Theory of Hints. An Approach to the Dempster-Shafer Theory of Evidence.* Lecture Notes in Economics and Mathematical Systems, vol. 425. Springer.

Kohlas, J., & Schneuwly, C. 2009. Information Algebra. *In:* Sommaruga, G. (ed), *Formal Theories of Information.* Springer Verlag.

Kohlas, J., & Stärk, R. F. 2007. Information Algebras and Consequence Operators. *Logica Universalis*, **1**(1), 139–165.

Kohlas, J., & Wilson, N. 2008. Semiring Induced Valuation Algebras: Exact and Approximate Local Computation Algorithms. *Artificial Intelligence*, **172**(11), 1360–1399. doi:10.1016/j.artint.2008.03.003.

Langel, J. 2004. *Local Computation with Models in Propositional Logic.* M.Phil. thesis, Department of Informatics, University of Fribourg.

Langel, J., & Kohlas, J. 2005. *A Semantic Theory of Propositional Information.* Tech. rept. 05-15. Department of Informatics, University of Fribourg.

Langel, J., & Kohlas, J. 2008. *Algebraic Structure of Semantic Information and Questions. Predicate Logic: an Information Algebra.* Tech. rept. 08-02. Department of Informatics, University of Fribourg.

Lauritzen, S. L., & Spiegelhalter, D. J. 1988. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statis. Soc. B*, **50**, 157–224.

Mellouli, K. 1988. *On the Propagation of Beliefs in Networks Using the Dempster-Shafer Theory of Evidence.* Ph.D. thesis, School of Business, University of Kansas.

Mengin, J., & Wilson, N. 1999. Logical Deduction Using the Local Computation Framework. *Pages 386–396 of:* Hunter, Anthony, & Parsons, Simon (eds), *Symbolic and Quantitative Approaches to Reasoning and Uncertainty, European Conference, ECSQARU'99, London, UK, July 5–9, 1999, Proceedings.* Lecture Notes in Computer Science, vol. 1638. Springer.

Mengin, J., & Wilson, N. 2001. Embedding Logics in the Local Computation Framework. *Journal of Applied Non-Classical Logics*, **11**(3-4), 239–261.

Plotkin, B. I. 1994. *Universal Algebra, Algebraic Logic, and Databases.* Mathematics and its applications ; vol. 272. Kluwer Academic Publishers.

Pouly, M. 2008. *A Generic Architecture for Local Computation.* Ph.D. thesis, Department of Informatics, University of Fribourg.

Pratt, V. 1999 (July). *Chu Spaces.* Course Notes. School in Category Theory and Applications, Coimbra, Portugal.

Schneuwly, C. 2007. *Computing in Valuation Algebras.* Ph.D. thesis, Department of Informatics, University of Fribourg.

Shafer, G. 1976. *A Mathematical Theory of Evidence.* Princeton University Press.

Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell Systems Technical Journal*, **27**, 379–423, 623–656.

Shenoy, P. P., & Shafer, G. 1990a. Axioms for probability and belief-function propagation. *Pages 169–198 of:* Shachter, Ross D., Levitt, Tod S., Kanal, Laveen N., & Lemmer, John F. (eds), *Uncertainty in Artificial Intelligence 4.* Machine intelligence and pattern recognition, vol. 9. Amsterdam: Elsevier.

Shenoy, P.P. 1989. A Valuation-Based Language For Expert Systems. *Int. J. of Approximate Reasoning*, **3**, 383–411.

Shenoy, P.P., & Shafer, G. 1990b. Axioms for Probability and Belief Function Propagation. *Pages 169–198 of:* R.D. Shachter, T.S. Levitt, J.F. Lemmer, & Kanal, L.N. (eds), *Uncertainty in Artif. Intell. 4.* North Holland.

Stokhof, M. 2002. Meaning, interpretation, and semantics. *Pages 217–240 of:* van Benthem, J., Scotto di Luzio, P., & Barker-Plummer, D. (eds), *Words, proofs, and diagrams.* CSLI Publications.

van Rooij, Robert. 2009. Comparing Questions and Answers: A bit of Logic, a bit of Language and some bits of Information. *In:* Sommaruga, G. (ed), *Formal Theories of Information.* Springer Verlag.

Index

Curriculum Vitae

Personal Data

Name:	Jutta Langel
Address:	Rue de la Carrière 20, CH - 1700 Fribourg
Date of birth:	31 May 1979
Place of birth:	Freiburg i. Brsg. (Germany)
Telephone:	+41 264 000 234
E-mail:	jutta.langel@unifr.ch
Nationality:	German
Marital status:	single

Education

2004 - 2009: Ph.D. studies in theoretical computer science
University of Fribourg (Switzerland)

2001 - 2004: studies in computer science, philosophy of language and French linguistics
University of Fribourg (Switzerland)
degree: Master of Science in Computer Science

2000 - 2001: studies in computer science and French linguistics
Université d'Orléans (France)

1998 - 2000: studies in computer science, French linguistics and French literature
Universität Leipzig (Germany)

1989 - 1998: high school
St. Ursula Gymnasium, Freiburg i. Brsg. (Germany)
degree: Abitur (A–level, specialization: mathematics and Latin)

1985 - 1989: primary school, Eschbach (Germany)

Work Experience

12 2004 – 12 2009:	research assistant in theoretical computer science, teaching assistant for courses at Bachelor's and Master's level, development and realization of courses in computer science and programming for children and young adults University of Fribourg (Switzerland) Departement of Informatics Theoretical Computer Science research group
08 2003 – 11 2004:	undergraduate assistant for courses at Bachelor's / Master's level University of Fribourg (Switzerland) Departement of Informatics
06 – 08 2001:	software developer for automated test software Thales Air Defense, Fleury-les-Aubrais (France)
03 – 04 2000:	trainee system integration data integral GmbH, Freiburg i. Brsg. (Germany)

Languages

German:	mother tongue
French:	near-native, bilingual studies
English:	fluent
Latin:	9 years at high school